Howe, K. (2004). A Critique of Experimentalism. *Qualitative Inquiry*, 10 (4), 42-61. Copyright © 1998 by Sage Publications, Inc. Reprinted by permission of Sage Publications, Inc.

A Critique of Experimentalism

Kenneth R. Howe
University of Colorado at Boulder

The concept of scientifically based research occupies a central place in the thinking of the newly formed Institute of Education Sciences and seems well on its way to becoming the dominant paradigm in educational research more generally. What interpretation becomes recognized as the correct one thus has important implications. This article identifies two versions of experimentalism that have emerged: neoclassical and mixed methods. Both versions of experimentalism are judged to be methodologically retrograde. Neoclassical experimentalism is little more than a throwback to the Campbell-Stanley era and its dogmatic adherence to an exclusive reliance on quantitative methods. Mixed-methods experimentalism, although incorporating an auxiliary role for qualitative methods, fails to understand the deeper epistemological roots of qualitative methods. The article briefly sketches the alternative of mixed-methods interpretivism, which elevates the voice of research participants to a primary position and thereby reverses the epistemological ordering of quantitative-experimental and qualitative interpretivist methods.

Keywords: *scientifically based research; quantitative methods; qualitative methods; experimentalism; interpretivism*

The concept of scientific research[1] occupies a central place in the No Child Left Behind Act of 2001, as well as in the newly formed Institute of Education Sciences. It appears to be well on its way to becoming the dominant standard for designing and evaluating educational research more generally. Thus, how it is interpreted has potentially huge implications for the future course of educational research.

In this article I critically examine the meaning of scientific research for the two methodological frameworks that are currently ascendant: *neoclassical experimentalism* and *mixed-methods experimentalism*. Both place experimental-quantitative research methods first among scientific methods and relegate qualitative methods to an auxiliary role. They differ with regard to how significant this auxiliary role is. I then briefly describe an alternative: *mixed-methods interpretivism*, a view that reverses the epistemological primacy of quantitative-experimental and qualitative-interpretivist methods. I argue

Qualitative Inquiry, Volume 10 Number 1, 2004 42-61
DOI: 10.1177/1077800403259491
© 2004 Sage Publications

that both forms of experimentalism are retrograde, whereas mixed-methods interpretivism has moved forward with the evolution of social science methodology during the past quarter century.

NEOCLASSICAL EXPERIMENTALISM

Classical experimentalism is a term that may be applied to the approach to educational research articulated by Campbell and Stanley (1963) in their seminal monograph *Experimental and Quasi-Experimental Designs for Research*. There they lauded the experiment as

> the *only* means for settling disputes regarding educational practice, as the *only* way of verifying educational improvements, and as the *only* way of establishing a cumulative tradition. (p. 2, italics added)

When conceived as the comprehensive, one best methodology for educational research, classical experimentalism had all but disappeared. But it has recently made a comeback in the context of the debate about what constitutes scientific research. This *neo*classical experimentalism, as I shall call it, fits squarely within the Campbell-Stanley (1963) framework. Both emphasize investigating causal relationships as the means by which to build a repertoire of "what works," and both rely almost exclusively on quantitative methods.

Neoclassical experimentalism differs from classical experimentalism in two minor but nonetheless noteworthy ways. First, neoclassical experimentalism is even more restrictive in terms of the designs and analysis techniques it endorses than is classical experimentalism. Campbell and Stanley (1963) generally construed the "true," randomized experiment as the methodological gold standard, like the neoclassical experimentalists. However, although Campbell and Stanley believed that "quasi experiments" are often quite defensible, neoclassical experimentalists are highly critical of these designs, to the point where it is not clear if they endorse using them at all (e.g., Coalition for Evidence-Based Policy, 2002; Whitehurst, 2003). Second, neoclassical experimentalism exploits the perceived role of randomized experiments in medical research to support the contention that educational research should recommit itself to an emphasis on randomized experiments (e.g., Boruch, 2002; Coalition for Evidence-Based Policy, 2002).

Neoclassical experimentalism is retrograde. It ignores the evolution in social science research methodology regarding the value of qualitative methods and their status with respect to quantitative methods. And it provides no fresh answers to long-standing criticisms of classical experimentalism's penchant to trade external for internal validity, to oversell randomization, and to oversell the ability of randomized experiments to provide causal explanations.

Ignoring the Evolution of Thought on Social Research Methodology

Neoclassical experimentalists would do well to pay close attention to how Campbell changed his views in the decade following publication of the Campbell-Stanley (1963) monograph. For he "recanted" his dismissive attitude toward qualitative methods, partly in response to growing disappointment with experimentalist research and partly in response to developments in the philosophy of science. Here is what he had to say in "Qualitative Knowing in Action Research":

> The polarity of quantitative-experimental versus qualitative approaches to research on social action remains unresolved, if resolution were to mean a predominant justification of one over the other. . . . Each pole is at its best in its criticisms of the other, not in invulnerability of its own claims to descriptive knowledge. . . . If we are to be truly scientific, we must reestablish the qualitative grounding of the quantitative. (Campbell, 1974, pp. 29-30)

In this, his more mature view, Campbell eschewed the single-minded pursuit of quantitative-experimental methods as the road to scientific research. In this view, quantitative and qualitative methods do not exclude one another and neither occupies a position of ultimate authority. Instead, quantitative and qualitative methods of knowing cross-check one another.

Campbell (1974) unpacked his general conception of the relationship between quantitative and qualitative methods of knowing in terms of what he called the "presumptive" nature of knowledge. The basic idea here is that the only way to test a given claim within a given theory or conceptual scheme is to presume the truth of the vast majority of other claims within that theory or scheme. For Campbell, a large number, perhaps most, of the claims that are presumed are qualitative and, thus, so are many of the judgments that have to be made in the course of conducting research. When quantitative-experimental research is detached from its qualitative grounding, the result is an "unhealthy division of labor" (Campbell, 1974, p. 13).

Campbell was not alone among prominent quantitative researchers in jettisoning classical experimentalism and the rigid quantitative/qualitative divide along with it. Lee Cronbach became one of experimentalism's most comprehensive and effective critics. Cronbach (1975) began his assault with his celebrated "Beyond the Two Disciplines of Scientific Psychology."[2] He cast considerable doubt on the idea that social science could be modeled on natural science, particularly regarding the possibility of accumulating robust generalizations about human behavior. Cronbach (1980, 1982) subsequently elaborated his criticisms of classical experimentalism in two volumes on evaluation research. Several of these criticisms are taken up below.

Trading External Validity for Internal Validity

Among the major drawbacks of randomized experiments are problems with external validity, including inconsistency in implementing interventions across contexts ("dispensing a curriculum" is not quite the same as "dispensing a pill"). There is a trade-off between internal and external validity: The more investigators restrict the population and the treatment to achieve internal validity, the less external validity the study will have.

External validity should take greater priority than internal validity, at least in practical fields such as education. Ironically, despite their emphasis on "what works," this is not the ordering that experimentalists embrace. Thomas Cook (2002), for example, who adopts randomized experiments as the generally preferred methodology for education research, said, "randomized experiments are best when a causal question is simple, sharply focused and easily justified" (p. 179). Cook goes on to say that when applied to schools, among the things that make random assignment "most feasible" are treatments that are short and that require no teacher training (p. 184).

This is the research-methodology tail wagging the educational-practice dog. Putting a premium on internal validity encourages educational researchers to focus on easy-to-manipulate, simplistic interventions and to avoid questions about existing policy and practice that for one reason or another, are not suited to being investigated via randomized experiments. Consider important policy questions of the day such as the effects of standards-based accountability and public school choice, both features of the No Child Left Behind Act of 2001. Randomized experiments can be conducted only at the margins of these policies. Thus, however internally valid such experiments may be, they are only marginally relevant to determining the effectiveness of the educational policies and practices currently of greatest concern.

More concretely, consider the celebrated experiment on New York's school voucher program (Myers, Peterson, Mayer, Chou, & Howell, 2000). Students were randomly assigned to receive vouchers for private schools from a pool of applicants larger than the number of vouchers available. The study thus controlled for an important source of selection bias, namely, that those applying for vouchers are likely to be more motivated, and so no, than the general population. The subsequent analysis compared those who had applied and received the vouchers with those who applied but did not receive vouchers. The researchers concluded that the vouchers resulted in some modest gains in achievement for African American students.

Whatever the internal validity of this research (which is threatened by the differential drop-out problem discussed below), its external validity is quite limited. Even if there were evidence of modest achievement gains (the find-

ing of any achievement gains whatever is disputed[3]), it would only show that voucher students outperform would-be voucher students. This would not come close to establishing that vouchers are an educational reform that will improve achievement overall (as its proponents so often claim). Students left behind would also need to improve (or at least not be hurt) by the departure of voucher students, a question the study did not address. Furthermore, the experiment failed to rule out several competing hypotheses for why voucher students might show improved achievement. One alternative hypothesis is that improved achievement (were it to be documented) could be the general effect of school choice, which could be just as powerful in public school choice system as in a voucher system. Another alternative is that, independent of choice, improved achievement could be produced by providing students with better *public* schools!

Overselling Randomization

Randomization is touted as if it were some magic bullet for what allegedly ails educational research (e.g., Boruch, 2002; Coalition for Evidence-Based Policy, 2002; Cook, 2002). To be sure, all other things being equal, randomization effectively reduces bias in estimating effects compared to other methods of control. But all other things rarely are equal.

First, social researchers typically must forgo random *selection* and make do with random *assignment*. The resulting estimates, however unbiased, are thus restricted to a population of volunteers. This does not always create a significant problem. But where the population of volunteers is likely to be substantially different from the target population, whatever reduction in bias is achieved by randomization may be outweighed by the need to generalize to that target population. Imagine using only volunteers to investigate the effectiveness of a program designed to reduce school violence.

Second, randomization provides no defense against the bias that often results from differential drop-out rates between treatment and control groups. In the case of school vouchers, for example, those who drop out of the treatment are likely to be unhappy with it, be seeing no benefit, and so on. No similar kind of systematic reasons apply to dropouts from the control group(s). Statistical adjustments may be employed to help equalize treatment and control groups, but this is one of the very things that neoclassical experimentalists want to avoid by employing true rather than quasi experiments.

Third, more often than not, researchers cannot employ random assignment; at least they cannot and also investigate the most important questions in educational policy and practice. Random assignment is often ruled out on political/legal grounds. For example, in states with charter schools on the books, neither students nor larger units of analysis may be randomly assigned to participate. Yet, the effects charter schools are producing is an

extremely important policy question that clearly ought to be, and is being, investigated. Credible nonexperimental results are being produced (e.g., as summarized by Gill, Timpane, Ross, & Brewer, 2001) that only the most hidebound experimentalist could ignore.

Overselling Experiments as Establishing Cause

The randomized experiment is frequently touted as the surest, if not *only*, way to make a causal inference in social research.[4] This is sheer dogma. The technique of the randomized experiment is neither sufficient nor necessary for establishing causation.

Randomized experiments are not sufficient to establish causal relationships because the inferences drawn from them so often involve only a very "gappy," black box account of relationships such that the precise cause(s) of the effect(s) cannot be identified (Cronbach, 1982). Acquiring a better understanding of causal mechanisms requires substantive knowledge of the contents and workings of the black box, something that cannot be obtained merely by employing the formal device of the randomized experiment.

Randomized experiments are not necessary to establish causal relationships because in cases where substantive background knowledge is available, robust causal relationships (at least in the statistical sense) can be established without randomized experiments. "Cigarette smoking causes cancer" is a good example. The gappy, black box correlation has been filled in with animal studies on the effects of exposure to tar, examinations of smoker's lungs, analyses of the chemical contents of cigarette smoke, studies that correlate the duration of cigarette smoking with the likelihood of developing lung cancer, and so on. There probably are not any causal relationships in education that are as firmly established as the link between cigarette smoking and lung cancer, but the principle is the same. For example, there have been no randomized experiments to show that race and parental income are causally related to academic performance, but the persistence of the associations makes it hard to deny that such causal relationships exist. Furthermore, various gap-filling explanations for low performance have been proffered, including lack of resources and experiences in the home, students being called on to care for siblings, peer pressure to avoid "acting White," and so on.

Misappropriating Medical Research

In addition to the general failure to acknowledge and entertain the methodological pitfalls described above, neoclassical experimentalism misappropriates medical research. Neoclassical experimentalists (e.g., the Coalition for Evidence-Based Policy) heavily tout medical research as the model for educa-

tional research, particularly the random clinical trial. But the analogy is seriously flawed: It submerges important differences between medical and educational research as well as important similarities.

An important difference between the fields of medical research and educational research is that it is typically much easier to zero in on the treatment and to maintain its consistent administration in clinical medicine than in education. Compare the treatment defined as "x mgs. of compound y each morning" to the treatment defined as "instruction in connected math 5 hours per week." The context of the administration of the treatment is much less of a complicating factor in clinical medicine than in education. As indicated above, "dispensing a curriculum" is quite different from "dispensing a pill." Finally, the precision with which outcomes can be measured varies considerably. Compare "a 10-point reduction in diastolic blood pressure" with "a 10-month growth in mathematics understanding."

An important similarity between the fields of medical research and educational research is that nonrandomized clinical trials are quite common in medical research, particularly outside of pharmaceutical research.[5] (Randomized clinical trials are most common in pharmaceutical research because, as the above illustration suggests, it is relatively easy to enforce the experimental conditions, including maintaining the standardization of the treatment.) Clinical medical research is divided into four phases, ranging from Phase 1, exploratory research on safety and side effects, to Phase 4, tracking the effects of a treatment after it has been put into general use. (Phase 4 pertains to medical research's "external validity" question.) Randomized clinical trials typically are employed in Phase 3 trials but they are not required in any phase.

Another important similarity between the fields of medical research and educational research is that the knowledge that has been accumulated through randomized clinical trials is not the only or even the most important factor in improved public health. Improved hygienic conditions and better nutrition, followed next by the (nonexperimental) development of immunizations and antibiotics, are credited with being, by far, the most important measures historically (e.g., Schneiderman & Speers, 2001). In a related vein, the effectiveness of medical care for given groups interacts with their socioeconomic status. Just as socioeconomic status is highly correlated with school performance, it is also highly correlated with health status. (According to one commentator, "low socioeconomic position is as strong a risk factor for poor health outcomes as smoking," Lynch, 2001, p. 52). And just as there is a persistent gap in school performance associated with socioeconomic status, there is a persistent gap in health status. Thus, if contributing to closing the achievement gap is one of the overriding goals of educational research, then medical research is not a model to be emulated.[6]

Conclusion

Boruch (2002)—in the vanguard of neoclassical experimentalism—warned that the educational research community may well resist embracing randomized experiments as the methodological ideal because of the "ideological posturing that so often substitutes for evidence in the education world" (p. 11). This remark is not only gratuitous; it gets the shoe on the wrong foot. Neoclassical experimentalism's attempted appropriation of medical research is patently selective. And it otherwise ignores the criticisms initiated by Campbell and Cronbach in the mid-1970s that were subsequently taken up and extended by a number of thinkers from the mid-1980s on.[7]

MIXED-METHODS EXPERIMENTALISM

Mixed-methods experimentalism is the view exemplified in the National Research Council (2002) report *Scientific Research in Education* (SRE). SRE assigns a significant role to qualitative methods. And they may be employed either singly or in combination with quantitative methods. In this respect, mixed-methods experimentalism is an advance over neoclassical experimentalism.

On the other hand, like neoclassical experimentalism, mixed-methods experimentalism places quantitative-experimental research methods and determining "what works" at the center of education science. Also like neoclassical experimentalism, mixed-methods experimentalism calls for a greater emphasis on randomized experiments without providing any real defense of why and without responding to the criticisms regarding the overselling of experimentalism.[8] In this and several other ways to be discussed below, mixed-methods experimentalism is not that distant from neoclassical experimentalism. It, too, is retrograde.

An Essentialist Conception of Science

The strategy in SRE was to set down the general features of science and then to determine what forms of educational scholarship fit. The committee seemed to assume that various sciences have a shared essence that it was their task to *discover*. SRE says things such as "At its core, scientific inquiry is the same in all fields" (National Research Council, 2002, p. 2) and the "accumulation of knowledge is the ultimate goal of [all] research" (National Research Council, 2002, p. 24). Although the committee did some hand waiving about how scientific progress is disjointed and is characterized by uncertainty and by fits and starts, they eventually identified the general aims of scientific edu-

cational research as "theory building" and "rigorous studies of interventions" (National Research Council, 2002, p. 126).

The project of articulating a general conception of science was largely abandoned in the wake of positivism, particularly a general conception that would include both the natural and social sciences (e.g., Chalmers, 1999). This does not mean that nothing hangs on the question of what qualifies as science, such that reasoned argument about it would be out of place. (Consider, for example, the question of whether "creation science" really qualifies as science.) The approach of SRE just is not very helpful.

The answer to the question of whether a certain inquiry approach is to be called "science" or "scientific" cannot be provided by *inspection* to determine whether that approach fits a set of predetermined categories. In addition to the inherent vagueness found at the edges of any concept, the terms "science" and "scientific" have a clear evaluative dimension (e.g., Chalmers, 1999). When the question of whether to apply these terms to a given activity is contested, we have to *decide* whether the activity counts as science or as scientific. Such a decision unavoidably turns on what values will be promoted or blunted.

An Outmoded Philosophy of Social Science

Related to its essentialism, SRE implicitly assumes some version of the principle of the "unity of science," in which social science (if it is to be science) must mimic the natural sciences. In this view, social science and natural science exhibit only a *difference of degree* such that social science is simply more complex than natural science in virtue of involving many more relevant variables over which investigators have little or no control.

Against this kind of view, Anthony Giddens (1976) has remarked, "Those who still wait for a Newton of social science are not only waiting for a train that won't arrive, they're in the wrong station altogether" (p. 13). The kind of *interpretivist* philosophy of social science[9] Giddens's remark exemplifies embraces a *difference of kind* between social science and natural science such that human behavior, unlike atoms and molecules, can be fully understood only from the insiders' perspective, in terms of the interpretation of meanings that actors employ. SRE does little more than waive at this alternative.[10] Human agency—a fundamental feature of social life bound up with the fact that humans are pervasive interpreters of others as well as themselves (Taylor, 1987)—is treated as little more than a factor that complicates social research by making humans less well behaved than billiard balls.[11]

The failure of SRE to give more serious attention to the interpretivist perspective is a significant lacuna. Interpretivism's influence is strongly felt in the philosophy of social science, and its various variants are embraced by a significant proportion of educational researchers (see Howe, 2003). More-

over, the interpretivist perspective challenges several prominent features of SRE's characterization of scientific educational research.

The question of demarcation. SRE asserts that scientific educational research is "empirical," to be cleanly distinguished from other kinds of education scholarship, the humanities in particular.[12] But the line between empirical social research and the humanities cannot be drawn very distinctly, if at all, given interpretivist methodology, in which the aims, requisite skills, and vocabularies of social science and the humanities significantly overlap (e.g., Cronbach, 1975; Taylor, 1987). It may be worth drawing such a line nonetheless, but SRE makes no convincing case for doing so. The committee seemed to just go with their intuitions and customary usage and to presuppose that the matter is unproblematic. But the matter is not unproblematic to the extent that it encourages the belief that empirical questions can be emptied of conceptual and value content and once emptied, may serve as the pristine foundation of *truly* empirical science. The project of demarcating science in this way—and thereby rendering it uncontaminated by metaphysics and values—was (is) the pipe dream of positivist social science.[13]

Education science as cumulative. As indicated previously, SRE holds that "the accumulation of knowledge is the ultimate goal of [all] research" (National Research Council, 2002, p. 24). But the idea that social science proceeds by piling up more and more on truths that survive testing is by no means unproblematic.[14] As Cronbach (1975) observed, "generalizations decay"; what at one time describes the social situation well might later be "valid only as history" (pp. 122-123). The point is not that generalizations are fallible and that mistaken ones may need to be subsequently corrected. (This would apply to the belief that bleeding patients is an effective medical treatment, for example. Bleeding is not now effective and never was.) The point is that what may be correctly generalized about human institutions and practices changes over time. A generalization that is now false could have been true at an earlier time and place. (Women compose a distinct minority of students enrolled in medical schools, for example.)

Giddens (1976) added a twist to Cronbach's (1975) observation: The practice of social research itself can be a factor in hastening the decay of generalizations, as a consequence of the "double hermeneutic." Social researchers engage in various interpretive (hermeneutical) acts in the process of coming to an understanding of the group they are studying. When researchers subsequently disseminate their findings to a public audience, members of this audience engage in (or at least may engage in) their own interpretive (hermeneutical) acts. This constitutes the "double" part of the double hermeneutic, and it has the potential to stimulate behavior on the part of the public that results in the decay of generalizations about social life. For "critical" researchers, *making* generalizations decay—generalizations documenting

oppressive relationships, in particular—is an explicit goal of social research (e.g., Fay, 1975, 1987).

My aim in this section has not been to jettison the idea of the accumulation of knowledge in education science wholesale, but to point to limitations in SRE's characterization. There are two. First, research does not have to be cumulative in the sense of building on what research has shown is true of social life; it may aim to demolish such truths. Second, even when the aim is accumulating knowledge in the straightforward sense of building on to what is true, such truths are subject to decay.

Causal relationships and causal mechanisms. SRE endorses a variety of research questions and a variety of research methods. In chapter 5, "Designs for the Conduct of Scientific Research in Education," it specifies how the two should be fitted together. At the most rudimentary level are descriptive questions: "What is happening?" These fit with quantitative methods such as surveys and qualitative methods such as ethnographies. At the next level are causal questions: "Is there a systematic effect?" These fit with quantitative methods, particularly the "ideal" of the randomized experiment. Qualitative methods are a source of causal hypotheses and may be used to "strengthen" causal inferences by helping eliminate alternative hypotheses. At the final level are causal mechanism questions: "Why or how is it happening?" These fit with a wide variety of quantitative and qualitative methods.

Quantitative and qualitative methods, then, cut across the three types of questions (i.e., both kinds of methods can be appropriately employed with respect to each type of question). Beyond this, SRE is rather vague on how things fit together, particularly regarding causation. Randomized experiments are singled out as the "ideal" for investigating the "causal relationships" type questions. As indicated above, qualitative methods are a source of causal hypotheses and may play the auxiliary role of helping "strengthen" inferences about causal relationships by eliminating alternative hypotheses. Qualitative methods do not otherwise play a role in inferring causal relationships.[15] The status of qualitative methods is less clear when it comes to investigating "causal mechanisms." Here it seems that qualitative methods may play a central (as opposed to auxiliary) role in the logic of causal inference. In particular, they may be used to help get beyond the black box, "gappy" understanding of causal mechanisms to which randomized experiments are often limited.

There is a tension, if not incoherence, in SRE's position on the role of qualitative methods in making causal inferences. In particular, it is difficult to see any real difference between "causal mechanisms" and "causal relationships." Questions about the causal mechanism that results in some effect, *E*, are more refined and precise than the questions about the causal relationship that results in *E*. This is a relative and rather arbitrary difference. It is not at all clear why, given SRE's position on investigating causal relationships, investigating causal mechanisms should not proceed in the same experimentalist way.

Causal mechanisms seem to be nothing other than more fine-grained causal relationships within the black box.

There is an important distinction to be made within social science regarding causation, but it is different from SRE's distinction between causal relationships and causal mechanisms. In particular, it is the distinction between the *regularity* and *intentional* conceptions of causation. The regularity conception construes causation in terms of relationships among *descriptive* variables grounded in the *outsider's perspective*. The intentional conception construes causation in terms of relationships among *intentional states* and *actions* grounded in the *insider's perspective*.[16]

According to John Searle (1984, 1995), human behavior must be understood against a complex background of "intentionality" that defines norm-regulated practices. Documented regularities among descriptive variables do not constitute causal explanations of human behavior; they call for them (Searle, 1984). Searle turned the typical, experimentalist construal on its head: Quantitative findings documenting regularities constitute the auxiliary, discovery work; filling in the black box requires investigating matters best handled with qualitative methods. Take the following example. We begin with the observed regularity that African American students living in "Trackton" exhibit low academic achievement.[17] This regularity in and of itself is not a causal explanation of anything. To provide such an explanation, we conduct an ethnographic study that gets at the perceptions and practices of the actors involved. We conclude that the differences among the linguistic practices of African American students and their White teachers cause distorted communication between them that in turn, causes lower academic performance on the part of the students.

Intentional causation is a central element of the interpretivist perspective, but there are no good reasons not to also employ the regularity sense of causation in social and educational research as appropriate (e.g., Fay 1975, 1996; Giddens, 1976; Howe, 2003). For example, "Poverty causes low performance in school" is "gappy" to be sure, and fails to speak to intentions, but it is an informative and coherent claim that does no violence to the concept of causation. Social and educational research can make fruitful use of such causal claims about complex social mechanisms over which people may have little control or awareness.

Conclusion

Mixed-methods experimentalism is a direct descendant of classical experimentalism and is less congenial to qualitative methods than it might first appear. It elevates quantitative-experimental methods to the top of the methodological hierarchy and constrains qualitative methods to a largely auxiliary role in pursuit of the *technocratic* aim of accumulating knowledge of "what

works." It is not that qualitative methods can never be fruitfully and appropriately used in this way, but their natural home is within an interpretivist framework with the *democratic* aim of seeking to understand and give voice to the insider's perspective regarding various educational policies and practices. This feature of qualitative-interpretivist methodology—associated with fundamental developments in the philosophy of social science during approximately the past quarter century—is scarcely even acknowledged in SRE.

Mixed-methods experimentalism gives primacy to quantitative-experimental methods over qualitative-interpretive methods to determine "what works," as if determining "what works" were somehow the self-certifying aim of educational research. "What works" is not an innocent notion. Concrete instances of the claim "Intervention *I* works" are elliptical for instances of the claim "Intervention *I* works to accomplish outcome *O*." The desired outcomes are embraced (if only tacitly) as more valuable than other possible outcomes, and the question of their value is off the table for anyone except policy makers and researchers. In this way, the aim of determining "what works" is *technocratic*: it focuses on the question of whether interventions are effective in achieving *given* outcomes.

AN ALTERNATIVE: MIXED-METHODS INTERPRETIVISM

What I call "mixed-methods interpretivism" reverses the primacy of quantitative-experimental and qualitative-interpretive methods such that quantitative methods play an auxiliary role in an overarching interpretivist-qualitative framework (Howe, 2003). The question of "what works" is construed much more expansively: The question of the value of the desired outcomes remains on the table to be assessed by various stakeholders.

Mixed-methods interpretivism actively engages stakeholder participation through the principles of *inclusion* and *dialogue*.[18] Inclusion is a general methodological principle that serves to control bias by ensuring the representativeness of samples. But it also has (or can have) a democratic dimension: ensuring that all relevant voices are heard.

The principle of dialogue adds an interpretivist dimension to inclusion and thickens its democratic dimension. Interpretivism emphasizes understanding people in their own terms, in their own social settings. Engaging them in dialogue is the most effective means of achieving this aim. As in the case of inclusion, there are both methodological and democratic justifications for employing dialogue: The deeper and more genuine expressions of beliefs and values that emerge through dialogue both foster a more accurate description of views held and *un*distort democratic deliberation.

Qualitative research methods such as participant observation, interviews, focus groups, and the like are well suited for promoting dialogue. Each technique involves some form of interaction between researchers and research participants that permits researchers to get below surface appearances to obtain a richer and more nuanced understanding of social life. There is a tension here, of course. Research participants can be mistaken or misinformed about the harms and benefits of various educational policies and practices, including to themselves. They typically also lack background knowledge and technical expertise. Thus, when researchers enter into dialogue with research participants, simply elucidating how participants think things work, and ought to work, can be no more than one element of full-blown—or *critical*—dialogue. Critical dialogue includes bringing expert knowledge to bear and subjecting the views and self-understandings of research participants to rational scrutiny.

Because mixed-methods interpretivism is, indeed, a mixed-methods approach, it cannot be distinguished from mixed-methods experimentalism simply in terms of the methods employed. How the role of participants is construed—democratic versus technocratic—is the determining factor. Below are three examples that help illuminate the differences between mixed-methods experimentalism and mixed-methods interpretivism.[19]

Example 1: School choice policy. The question of whether a school choice policy is a good one is not the same as the question of whether it works in the sense of attaining the goals policy makers and theoreticians have for it. Judgments about the worth of program goals themselves, about competing goals, about unintended consequences, and about how to balance these are all relevant to the question of whether a school choice policy is a good one. These are the kinds of things many stakeholders know about and that in the name of democracy, ought to have an effective say about. Having an effective say requires that stakeholders be included in genuine dialogue.

Qualitative methods are best suited for fostering dialogue, though as indicated above, they also may be used instrumentally to determine "what works," with no real commitment to the democratic dimension of dialogue. On the other hand, although quantitative methods are best suited for situations in which the variable and outcomes of interest are settled on ahead of time, they may be employed within an overarching democratic framework in which the effects of given polices are crucial to deliberation. For example, whether school choice policies cause increased segregation is critical in the current policy debate. Once again, the crucial difference between mixed-methods experimentalism and mixed-methods interpretivism is how the role of participants is construed.

Example 2: Research on the "hidden curriculum." Research on the "hidden curriculum" and related features of schools fit naturally with the dialogical

approach associated with mixed-methods interpretivism. Lois Weis and Michelle Fine (1993) wrote in the introduction to *Beyond Silenced Voices*, there is "a *discursive underground* of students and adults that flourish within the margins of our public schools. These voices need to be heard . . . if we are serious about schools as a democratic public sphere" (p. 2). How femininity, masculinity, race, and beliefs and attitudes about sexual orientation are shaped and reinforced by school cultures and curricula are among the issues with which students and teachers are engaged in dialogue by educational researchers. Among the more specific questions addressed are how certain voices are missing from the culture and curricula, what the consequences might be, and what measures might be taken in response.

Perhaps there is a randomized experiment lurking here somewhere, but it is difficult to see how an emphasis on experimentation would significantly advance this general area of research. On the contrary, such an emphasis would hamper it.

Example 3: Research on teaching and learning. Mixed-methods interpretivism applies least obviously and straightforwardly to research on teaching and learning, and this is where experimentalism applies best. Because there is often considerable agreement on explicit goals and outcomes in this arena, dialogue is rendered less necessary and researchers are relatively safe to proceed with ascertaining "what works."

But they are not *completely* safe, for educational research can never be free of value commitments. Here, it will be helpful to distinguish between *value neutrality* and *value freedom* (Howe, 2003). Educational research can be value neutral in the sense of being *neutral among* different moral-political stances. This is not to say that educational research can ever be value free, however. For example, research on the acquisition of basic computation skills is a good candidate for the kind of investigation that can quite plausibly be characterized as value neutral. But it is important to note that to the extent that the acquisition of computation skills may be correctly characterized as value neutral, *neutrality does not go all the way down*. Broader questions about values are lurking, even in a relatively uncontroversial area such as math education. For example, what approaches to math curricula and instruction best prepare students to become competent democratic citizens? What approaches to math curricula and instruction are least likely to be exclusionary of certain kinds of students? What approaches to math curricula and instruction are most likely to make students critical mathematical thinkers and to foster a healthy skepticism of mathematics as an all-purpose intellectual tool? What trade-offs are to be made among mathematics and other subjects? And so on. These are not technical "what works" questions.[20]

CONCLUDING REMARKS

The contrast between the technocratic thrust of experimentalism and the democratic thrust of interpretivism gets at the political dimensions of educational research. Although it is clearly an important part of the academic debate, there is a political dimension that is considerably less high flown. In these concluding remarks, I speculate about what besides academic debate is driving the demand to define standards for scientific educational research and about what some of the consequences are likely to be.

First, the imposition of external standards is something that K-12 has been enduring for some time. Standards have recently arrived at the academy: first in teacher education and now in research. Arguably, the standards-setting movement in education is inherently about power and control, under the banner of protecting the public from incompetent or misguided practitioners.[21] Groups like the committee that worked diligently to prepare SRE have scrambled to show how what they do qualifies as science to regain some of their power to influence policy makers. But they may very well have been diverted onto the trail of a red herring. Organizations such as the Fordham Foundation, the Manhattan Institute, and the Heritage Foundation, for example, have considerable influence on policy makers, and it is doubtful that this is because the product they produce measures up to high scientific standards. More likely, it is because their product is sensitive to the ideological predilections and agendas of the current powers that be and that it is conceived and marketed accordingly (Howe, 2002).

Second, the strong endorsement of randomized trials as the "gold standard" (by the neoclassical experimentalists) and the "ideal" (by the mixed-methods experimentalists) creates the imperative to anticipate and vigorously control all the variables that are predetermined to be relevant vis-à-vis "what works." This puts blinders on researchers and drives educational research in a certain political direction: away from raising critical questions about the social and institutional context of schooling and toward various interventions that "work" given the status quo. Researchers who persist in the social critique of educational practice are liable to be written off for their alleged "ideological posturing."

Finally, it is not just the "methodological fundamentalists"[22] who have bought into the "what works" approach. A sizable number of rather influential and otherwise sensible educational researchers—several on the National Research Council committee, for example—have also signed on. This might be a compromise in response to the current political climate; it might be a backlash against the perceived excesses of postmodernism; it might be both. It is an ominous development, whatever the explanation.

NOTES

1. I include cognates such as "scientifically based" and "evidence based" as part of the general territory.

2. Interestingly, Cronbach presented this paper at the same 1974 meeting of the American Psychological Association as the Campbell paper described above. Cronbach (1982) noted the similar turns their thinking had taken in these papers.

3. See Winerip (2003) for a summary.

4. The concept of cause is particularly nettlesome in social research. It can assume two different meanings. One is the "intentional" meaning of cause, in which causal explanations are *teleological* (i.e., appeal to agents' purposes and interpretations of meanings). This meaning is rarely (if ever) considered in the context of experimentalism. (Intentional causation is discussed in greater detail in the section herein on mixed-methods experimentalism.) The second is the more familiar "regularity" meaning of cause, where causal explanations are *descriptive* (i.e., appeal only to observed regularities). The regularity sense is seen as particularly problematic in social research because the putative causal relationships are statistical, not deterministic ones (Cronbach, 1982).

5. I recently did a study, "unscientific" to be sure, but revealing nonetheless. I went to the Medline Web site and looked under "coronary disease"—the treatment of which has got to be considered one of modern medicine's success stories. As I expected, I found that the majority of the clinical studies described there were not randomized experiments. Indeed, I found one study comparing medical and surgical treatments of coronary ischemia in which the researchers claimed that their nonrandomized study was actually methodologically *superior* to a randomized clinical trial because it included a much more representative sample of the population of interest than would be possible with a sample of patients willing to volunteer for a randomized study.

6. Of course, it is highly dubious that better research methodology will solve the problem in either case. We already have pretty good understandings of the direction of causation between socioeconomic status on one hand and school performance and health status on the other hand, however "gappy" they may be.

7. A lot more could be said about the quantitative/qualitative debate—a debate that neoclassical experimentalists have conveniently avoided participating in. But that is a discussion for another place. See Howe (2003) for a comprehensive treatment of the issues.

8. Here is what the National Research Council (2002) had to say on the issue:

> In estimating the effects of programs, we urge the expanded use of random assignment. Randomized experiments are not perfect. Indeed, the merits of their use in education have been seriously questioned. . . . For instance, they typically cannot test complex causal hypotheses, they may lack generalizability to other settings, and can be expensive. However, we believe that these and other issues do not generate a compelling rationale against their use in education research and that issues related to ethical concerns, political obstacles, and other potential barriers often can be resolved. . . . Establishing cause is often exceedingly important—for example, in large-scale deployment of interventions—and the ambiguity correlational of quasi-experiments can be undesirable for practical purposes. (p. 125)

9. I use "interpretivism" in an expansive way to include a variety of postpositivist views that insist social research must include a special "intentionalist" vocabulary and an associated array of "qualitative" methods. Many such interpretivist views, including Giddens's (1976) and my own pragmatic view (Howe, 2003), advocate mixed-methods, not an exclusive reliance on an intentionalist vocabulary and qualitative methods.

10. This is about all the National Research Council (2002) had to say on the topic:

> Differences in the phenomena typically under investigation do distinguish the research conducted by physical and social scientists. . . . Unlike atoms or molecules, people grow up and change over time. The social, cultural, and economic conditions they experience evolve with history. The abstract concepts and ideas that are meaningful to them vary across time, space, and cultural tradition. *These circumstances have led some social science and education researchers to investigate approaches that look distinctly different from those of physical researchers* [italics added], while still aligning with the guiding principles outlined [in this volume]. (p. 81)

11. See especially *Scientific Research in Education* subsection "Human Volition," (National Research Council, 2002, beginning on p. 86).

12. Interestingly, philosophy and history, both humanities disciplines, fit the general characterization of science provided early in *Scientific Research in Education*: "a continuous process of rigorous reasoning supported by a dynamic interplay among methods, theories, and findings" (National Research Council, 2002, p. 2).

13. One of positivism's basic tenets was that science is the paragon of knowledge, the only *real* kind of knowledge (save logic and math). One of its ambitions was to *demarcate* science from pretenders to knowledge, such as metaphysics, ethics, and aesthetics. The project failed, and one of the primary reasons was that Quine, Kuhn, and others convincingly undermined the notion that you could cleanly isolate the empirical contents of science from its conceptual (or theoretical) content. So, there are relatively fundamental philosophical reasons to be suspicious of the assumption that "scientifically based" research can be isolated on the basis of its "empirical" content.

14. This conception of scientific progress has been seriously challenged even in physical sciences, most famously by Thomas Kuhn (1962).

15. This division of labor between qualitative and quantitative methods was typical in the early days of the quantitative/qualitative debate. It maps, à la positivism, qualitative methods on to the "context of discovery" and quantitative methods on to the "context of justification."

16. The idea of "intentional" has a much broader meaning in philosophy than in ordinary usage. It has to do with the "aboutness" that characterizes concepts crucial to understanding human behavior, such as knowledge, belief, doubt, and so on. Knowledge, belief, and doubt are each about the planets, algebra, politics, and so on. Concepts in the physical sciences and "behaviorese" do not have this "aboutness" feature and are thus not intentional.

17. The example is based on Heath (1983).

18. Here I collapse three principles first articulated in House and Howe (1999)—inclusion, dialogue, and deliberation—into two.

19. These examples are adapted from Howe (2003).

20. Examples like this make me less than sanguine about following the authors of *Scientific Research in Education* (National Research Council, 2002) in drawing a sharp line between science and philosophy—as if we can avoid doing (or presupposing) something *philosophical* as we engage in empirical research.

21. The American Educational Research Association membership got a bit of a scolding from Russ Whitehurst in this vein at the 2003 convention (Whitehurst, 2003).

22. This term is borrowed from House (2003).

REFERENCES

Boruch, R. (2002). The virtues of randomness. *Education Next*, *2*(3), 36-42. Retrieved from http://www.educationnext.org/20023/36.html

Campbell, D. (1974, September). *Qualitative knowing in action research*. Kurt Lewin Award Address, Society for the Psychological Study of Social Issues, presented at the meeting of the American Psychological Association, New Orleans, LA.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Chalmers, A. (1999). *What is this thing called science?* (3rd ed.). Indianapolis, IN: Hackett.

Coalition for Evidence-Based Policy. (2002). *Bringing evidence-driven progress to education: A recommended strategy for the U.S. Department of Education*. Author. Available from http://www.excelgov.org/displayContent.asp?NewsItemID=4541&Keyword=prppcEvidence

Cook, T. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Education Evaluation and Policy Analysis*, *24*(3), 175-199.

Cronbach, L. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116-127.

Cronbach, L. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Cronbach, L., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., et al. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.

Fay, B. (1975). *Social theory and political practice*. Birkenhead, UK: Allen and Unwin.

Fay, B. (1987). *Critical social science*. Ithaca, NY: Cornell University Press.

Fay, B. (1996). *Contemporary philosophy of social science*. Malden, MA: Blackwell.

Giddens, A. (1976). *New rules of sociological method*. New York: Basic Books.

Gill, B., Timpane, M., Ross, K., & Brewer, D. (2001). *Rhetoric versus reality: What we know and what we need to know about vouchers and charter schools*. Santa Monica, CA: RAND.

Heath, S. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.

House, E. (2003). Bush's neo-fundamentalism and the new politics of evaluation. *Studies in Education Policy and Educational Philosophy, 2*. Available from www.upi.artisan.se

House, E., & Howe, K. (1999). *Values in evaluation and social research*. Thousand Oaks, CA: Sage.

Howe, K. (2002, April 10). Free market free for all. *Education Week*, *21*(30), 32, 34-35.

Howe, K. (2003). *Closing methodological divides: Toward democratic educational research*. Dordrecht, the Netherlands: Kluwer.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lynch, J. (2001). Socioeconomic factors in the behavioral and psychosocial epidemiology of cardiovascular disease. In N. Schneiderman, M. Spears, J. Silva, H. Tames, & J. Gentry (Eds.), *Integrating behavioral and social sciences with public health* (pp. 51-72). Washington, DC: American Psychological Association.

Myers, D., Peterson, P., Mayer, D., Chou, J., & Howell, W. (2000). *School choice in New York City after two years: An evaluation of the school choice scholarship program*. Washington, DC: Mathematica Policy Research.

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Schneiderman, N., & Speers, N. (2001). Behavioral science, social science, and public health in the 21st century. In N. Schneiderman, M. Spears, J. Silva, H. Tames, & J. Gentry (Eds.), *Integrating behavioral and social sciences with public health* (pp. 3-30). Washington, DC: American Psychological Association.

Searle, J. (1984). *Minds, brains, and science*. Cambridge, MA: Harvard University Press.

Searle, J. (1995). *The construction of social reality*. New York: Free Press.

Taylor, C. (1987). Interpretation and the sciences of man. In P. Rabinow & W. Sullivan (Eds.), *Interpretive social science: A second look* (pp. 33-81). Los Angeles: University of California Press.

Weis, L., & Fine, M. (1993). Introduction. In L. Weis & M. Fine (Eds.), *Beyond silenced voices: Class, race, and gender in United States schools* (pp. 1-8). Albany, NY: State University of New York Press.

Whitehurst, G. (2003, April). *The Institute of Education Sciences: New wine and new bottles.* Presentation at the annual meeting of the American Educational Research Association, Chicago, IL.

Winerip, M. (2003, May 7). What some much-noted data really showed about vouchers. *New York Times.* Available from http://www.nytimes.com/2003/05/07/education/07EDUC.html?pagewanted=print&position=

Kenneth R. Howe is professor in the Educational Foundations, Policy, and Practice Program area, director of the Education and the Public Interest Center, and associate dean of Graduate Studies, University of Colorado at Boulder. He specializes in education policy, professional ethics, and philosophy of education. He has conducted research on a variety of topics, ranging from the quantitative/qualitative debate to a philosophical examination of constructivism to a defense of multicultural education. His books include the Ethics of Special Education *(with Ofelia Miramontes),* Understanding Equal Education Opportunity: Social Justice, Democracy and Schooling, Values in Evaluation and Social Research *(with Earnest House), and* Closing Methodological Divides: Toward Democratic Educational Research. *He teaches courses in the social foundations of education, the philosophy of education, and philosophical issues in educational research.*