Educational Researcher

http://er.aera.net

he Consequences of Experimentalism in Formulating Recommendations for Policy and Practice in Mathematics Education

Paul Cobb and Kara Jackson

EDUCATIONAL RESEARCHER 2008; 37; 573
DOI: 10.3102/0013189X08327826

The Consequences of Experimentalism in Formulating Recommendations for Policy and Practice in Mathematics Education

Paul Cobb and Kara Jackson

In this response to *Foundations for Success: The Final Report of the National Mathematics Advisory Panel* (2008), the authors argue that the Panel's assumption that only experimental research studies can produce scientific evidence limits the power of the Panel's recommendations to improve mathematics teaching and learning. The authors first discuss the theoretical underpinnings, potential contributions, and limitations of experimental studies. Against this background, they focus on three issues that are central to improving mathematics learning and teaching, those of equity, the nature and content of textbooks, and graduate education. In doing so, the authors illustrate the limitations of developing implications for policy and practice by relying exclusively on research conducted using a single methodology.

Keywords: equity; mathematics education; research methodology

The National Mathematics Advisory Panel's (2008) recent report sought to synthesize the "'best available scientific evidence' to [recommend] ways '. . . to foster greater knowledge of and improved performance in mathematics among American students'" (p. xiii). We support the effort to develop recommendations for policy makers and teachers that are based on high-quality research. The Panel produced a comprehensive report with an impressive array of supporting documents. Unfortunately, the Panel took an overly narrow view of what counts as scientific evidence, thereby failing to capitalize on much of what is known about mathematics learning and teaching. As a consequence, the Panel's report is less effective than it would otherwise have been in supporting policy makers and teachers to make substantial improvements. In this response to the report, we first discuss the theoretical underpinnings and the potential contributions and limitations of experimental research studies. We go on to argue that other methodologies produce different forms of knowledge that complement the findings of the Panel and would have increased the usefulness of the Panel's report for policy and practice. Against this background, we then focus on three issues that are central to improving mathematics learning and teaching, those of equity, the content of textbooks, and graduate education.

The Approach the Panel Used to Produce Its Recommendations

The trustworthiness of research findings depends on the soundness of the method used to produce those findings (Bernstein, 1983; Lakatos, 1970; Popper, 1972). Similarly, the value of the Panel's recommendations depend on the soundness of the method the Panel used to (a) discriminate between trustworthy and suspect research findings and (b) synthesize findings judged to be trustworthy. In our view, there is good reason to be concerned about both of these steps in the Panel's approach.

The Panel used three categories to discriminate between trustworthy and suspect studies. The first category of high-quality scientific evidence is reserved for "studies that test hypotheses, meet the highest methodological standards (internal validity), and have been replicated with diverse samples of students under conditions that warrant generalization (external validity)" (p. 7-4). The Panel assumed that only one methodology can produce high-quality scientific evidence: experimental and quasi-experimental studies. The studies in the Panel's second category of promising or suggestive findings "represent sound, scientific research that needs to be further investigated or extended" (p. 7-4) by conducting experimental studies of instructional practices or existing curricula. "The final category corresponds to statements based on values, impressions, or weak evidence; these are essentially opinions as opposed to scientifically justified conclusions" (p. 7-4). Logic dictates that this third category includes all studies conducted using any methodology other than the experimental methodology, regardless of their quality. Having differentiated between research studies in this manner, the Panel grouped studies that met its standards for scientific evidence according to the independent variables investigated, synthesized the findings of each group of studies, and where possible, derived recommendations.

In critiquing the Panel's approach, we differentiate between experimental research as a methodology and *experimentalism* as an ideology that holds that only studies conducted using this methodology constitute a trustworthy basis for making recommendations to policy makers and practitioners. As we will clarify, we consider experimental research to be a valuable methodology that has the potential to make important contributions to the improvement of mathematics teaching and learning.[1] However, we will argue that experimentalism is an unhelpful position. The Panel interpreted its charge of making policy recommendations

Educational Researcher, Vol. 37, No. 9, pp. 573–581
DOI: 10.3102/0013189X08327826

on the basis of available scientific evidence to mean that it should consider only experimental and quasi-experimental studies. With the exception of the report of the Instructional Practices Task Group, the materials the Panel produced contain few indications that it viewed its charge as overly restrictive. Instead, the Panel argued that although the research it reviewed did not allow it to make policy recommendations in many areas central to improving mathematics instruction, this problem can be solved by developing the capacity of the research community to conduct experimental studies (see, e.g., pp. xxvi, 63). As the Panel did not justify the thoroughgoing experimentalist stance apparent in its report, we prepare the ground for our response to the Panel's report by first explicating the underlying suppositions and assumptions of experimental research. Next, we focus on the issue of making causal claims and then consider the potential contributions and limitations of experimental research. In doing so, we adopt the pragmatic perspective that different methodologies make different assumptions about the phenomena under investigation, ask different questions, and produce different forms of knowledge that are useful for different purposes and that are frequently complementary.

The Theoretical Grounding of Experimental Research

Slavin (2004) succinctly described the forms of knowledge experimental studies produce under the best of circumstances when he clarified that well-designed studies of this type are not limited to *x* versus *y* comparisons but can "also characterize the conditions under which *x* works better or worse than *y,* the identity of the students for whom *x* works better or worse than *y,* and often produce rich qualitative information to supplement the quantitative comparisons" (p. 27). The key point to note for our purposes is that knowledge claims associated with experimental studies reflect a particular conception of the individual. The knowledge claims refer to an abstract, collective individual or statistical aggregate that is constructed by combining measures of psychological attributes of the participating students (e.g., measures of mathematics achievement). This statistically constructed individual is abstract in the sense that it does not correspond to any particular student.

This methodological approach of investigating the performance of the abstract, collective subject rests on two underlying assumptions. First, students possess larger or smaller measurable amounts of discrete, isolatable psychological attributes (Danziger, 1990). This assumption makes it legitimate to combine measures of student performance. Second, the learning environments in which students acquire these attributes are composed of independent features that the investigator can manipulate and control directly. The implicit ontology is that of designed environmental settings made up of separate independent variables and students composed of collections of dependent psychological attributes. Together, these two theoretical suppositions ground experimental studies that seek to discern causal relationships between the manipulation of instructional conditions and the subsequent performance of the collective, abstract individual.

Causal Relations

We infer that the Panel's commitment to review only experimental studies is based on the assumption that no other methodology can contribute to the establishment of causal claims about the effectiveness of instructional interventions. We follow Maxwell (2004) in arguing that this assumption is unwarranted. Maxwell distinguishes between two complementary treatments of causal explanation. The first of these two treatments, which Maxwell terms the *regularity type of causal description,* is central to the experimental methodology and is based on observed regularities across a number of cases. Maxwell calls the second treatment *process-oriented explanation* and clarifies that it "sees causality as fundamentally referring to the actual causal mechanisms and processes that are involved in particular events and situations" (p. 4). Process-oriented explanations are therefore concerned with "the mechanisms through which and the conditions under which that causal relationship holds" (Shadish, Cook, & Campbell, 2002, p. 9, cited in Maxwell, 2004, p. 4). In contrast to the regularity conception of causality, viable explanations of this type can be developed based on a relatively small number of purposefully selected cases (Maxwell, 2004). For example, studies employing the design research methodology have been conducted to develop process-oriented causal explanations of the relations between teachers' instructional practices, instructional tasks as they are actually enacted in the classroom, the learning opportunities that arise for students as they engage in the tasks, and students' resulting learning in particular mathematical domains (P. Cobb, McClain, & Gravemeijer, 2003; Confrey & Smith, 1995; Lehrer & Schauble, 2004). Rigorous, systematic studies conducted using the design research and observational methodologies can produce knowledge of the actual processes by which improvements in learning outcomes are produced that are sufficiently robust to support causal claims (Brown, 1992; Yin, 2003). Furthermore, these claims are potentially generalizable, provided that the necessary aspects of these processes are differentiated from the contingent aspects (P. Cobb & Gravemeijer, in press; Kelly, 2004).

In the Panel's defense, the assumption that only experimental research can produce evidence of causal relations is easily conflated with a second issue, namely, the extent to which a methodology is underpinned by a clearly articulated scheme of argumentation that links data to analysis and to final claims and assertions. As Kelly notes, the experimental methodology is underpinned by an explicit scheme of argumentation that makes it relatively straightforward to differentiate between strong and weak experimental studies. Kelly goes on to observe that other methodologies differ in the extent to which a corresponding scheme of argumentation has been articulated. As a consequence, the process of identifying high-quality studies conducted using other methodologies is not always a simple matter. The fact that making these discriminations can be demanding does not, in our view, constitute an adequate rationale for failing to consider entire bodies of relevant research. The Panel would have done the research community a service if it had followed Kelly's lead and made the development of explicit schemes of argumentation for other methodologies one of its major research recommendations.[2] In doing so, the Panel might have noted that the six defining characteristics of scientific education research proposed by the National Research Council (2002) provide useful guidance because they cut across a range of methodologies. As Eisenhart and Towne (2003) clarify, these characteristics or principles were identified by reviewing actual research programs. This empirical grounding serves to differentiate the National Research Council's view of high-quality research from approaches that favor one methodology to the exclusion of all others. It also provides an antidote to the all-too-common practice of formulating recommendations for policy and practice by picking and choosing studies conducted using a range of methodologies to suit a favored argument.

The Usefulness of Experimental Research

Potential contributions. Our observation that experimental research reflects a particular set of assumptions about students and the environments in which they learn does not in any way threaten the legitimacy of the claims associated with the methodology, provided that the methodology was appropriate given the research question under investigation (Kelly & Yin, 2007). The research practices established by any research community necessarily entail theoretical commitments. We take a pragmatic approach that focuses squarely on the usefulness of experimental research in contributing to the improvement of mathematics teaching and learning. To address the question of usefulness adequately, we have to consider both for whom the forms of knowledge produced by experimental research are useful and for what purposes. This qualification is important because a particular form of knowledge might be especially useful for school district administrators but not for classroom teachers, and vice versa.

Danziger's (1990) historical analysis is directly relevant to the issue of usefulness. He reported that, in the early years of the 20th century, educational administrators drew on the findings of experimental studies as they rationalized the rapidly expanding U.S. public education system. Knowledge about the responses of the statistically constructed collective individual to different instructional conditions served the needs of administrators who managed institutions in which instruction was carried out in groups. Furthermore, the conception of learning environments as composed of manipulable independent variables was directly relevant to administrators who sought to manage classroom instruction. Danziger illustrated that the experimental methodology was subsequently refined to produce forms of knowledge that fulfill administrators' needs. The resulting forms of knowledge enable administrators who are removed from the classroom to make informed decisions about the curricula and other types of instructional intervention, especially if they are supplemented by the findings of evaluation studies conducted using other methodologies (Confrey, 2006; National Research Council, 2004). In our view, this is a potentially important contribution to the improvement of the quality of mathematics instruction. Commercial and nonprofit vendors currently market a wide array of instructional programs and services to schools and districts, many of which are experiencing severe financial stress. In our view, it is an ethical imperative that groups and organizations marketing programs designed to influences what Delpit (1995) termed the education of other people's children seek independent evaluations of their products, using evaluation methods that include experiments.

Limitations. We have indicated that the findings of experimental studies can be valuable to educational administrators. However, it is important to acknowledge that the process of selecting from the array of programs and services offered to schools and districts is not necessarily straightforward even when experimental evidence is available. School and district instructional leadership that includes ambitious goals for students' mathematical learning is a complex practice. In bare-bones form, this type of leadership practice involves developing a vision of high-quality mathematics instruction, formulating a coherent plan for how to make this vision a reality in school or district classrooms within the constraints of available resources, and enacting (and adjusting) this plan by supporting the learning of personnel at all levels of the school or district (Coburn, 2005b; Elmore, 2006; Glennan & Resnick, 2004; McLaughlin, 2006; Nelson & Sassi, 2005).

This type of leadership practice involves developing supports for instructional improvement by drawing on both external programs and internal resources. It is critical that the various components of the improvement effort are well aligned and mutually reinforce each other (P. Cobb & Smith, in press; Coburn, 2005a; Cohen, Moffitt, & Goldin, 2007; Resnick, Besterfield-Sacre, Mehalik, Sherer, & Halverson, 2007). Consequently, in selecting external programs, school and district leaders have to go beyond experimental evidence by considering whether a particular program meets or can be adjusted to local needs and whether it contributes to the overall coherence of the improvement effort. This highly specialized reasoning is an aspect of what Spillane (2005) terms the *how* of leadership practice. The experimental methodology is not well suited to the challenge of gaining insights into what instructional leaders actually *do* to organize and support instructional improvement. As a consequence, the Panel's report offers little empirically grounded advice to school and district leaders about the how of their work. In our view, it is unfortunate that the Panel chose not to consider studies conducted using other methodologies that have investigated what instructional leaders do and why they do it.

In our reading of its report, the Panel takes the position that the findings of experimental studies are as useful to classroom teachers as they are to school and district leaders. Danziger's (1990) analysis of the development of the experimental methodology in education provides an initial reason for treating this assumption with caution. Although the forms of knowledge produced by experimental studies can, on occasion, provide broad guidance to teachers, the methodology is not well suited to the how of teaching or, in other words, to the contingencies of classroom practice. This is because the knowledge produced of individual students is, as Danziger put it, "a knowledge of strangers" (p. 165) who are known only through their standing in the group. As a consequence, this form of knowledge does not touch on the challenges, dilemmas, and uncertainties that arise in the classroom as teachers attempt to use a curriculum to achieve a mathematical agenda while simultaneously taking account of their students' proficiencies, interests, and needs (Ball, 1993; Fennema, Franke, & Carpenter, 1993; Lampert, 2001; McClain, 2002).

To its credit, the Panel's Task Group on Instructional Practices acknowledges that "mathematics teaching is an extraordinarily complex activity involving interactions among teachers, students, and the mathematics to be learned in real classrooms" (p. 4-10). In addition, this task group goes on to clarify that "not all of the questions that teachers, policymakers, and the public wish to have answered are easily studied or lend themselves to experimental and quasi-experimental research" (p. 4-10). This acknowledgement of the limitations of the experimental methodology is a welcome exception in the materials that the Panel produced. We are therefore left wondering why the Panel continued to adhere to the canon that the experimentalist methodology has a monopoly on the production of scientific evidence in education. Because the Panel chose not to consider high-quality studies conducted using a range of methodologies, it could offer teachers only the following three global aphorisms:

- monitor what students understand and are able to do mathematically;
- design instruction that responds to students' strengths and weaknesses based on research when it is available; and
- employ instructional approaches and tools that are best suited to the mathematical goals, recognizing that a deliberate and conscious mix of strategies will be needed. (p. 4-21)

It is unlikely that most teachers will derive useful guidance from these recommendations, given their notable lack of specificity. In our view, the failure of the Panel's recommendations to make contact with the how of teaching and of policy formulation and enactment is a direct consequence of its decision to review only experimental studies.

The Classroom as an Instructional System

Thus far, we have focused on the first step in the process that the Panel followed to produce its findings: that of identifying research studies that met its standards for scientific evidence. The second step was to group the identified studies according to the independent variables investigated, synthesize the findings of each group of studies, and where possible, derive recommendations. In carrying out this second step, the Panel attempted to answer several questions that cannot be addressed in a useful way by research. These include the impact on student achievement of using real-world contexts to introduce mathematical ideas, using instructional software, and using calculators. Good (2008) indicated why these are poor research questions in his response to the Panel when he observed that "any single variable has meaning only as part of an instructional *system*" (p. 4). A large number of studies indicate the value of viewing the classroom as a system of interrelated elements rather than as a set of isolated variables, the most well known of which is the TIMSS Video Study (Stigler & Hiebert, 1999). Elements of the classroom instructional system include the instructional tasks, the organization of classroom activities (including the norms of participation in each phase of lessons), the tools available for students to use, and the nature of classroom discourse (including standards of mathematical argumentation; P. Cobb, 2001). These four elements are interrelated in that, for example, instructional tasks involving real-world contexts as they are actually enacted in the classroom depend on each of the three other elements. Similarly, the actual activities in which students engage as they use tools such as instructional software or calculators depend on the other three elements of the system.

From this perspective on the classroom, the results of meta-analyses conducted to evaluate the impact of using real-world contexts, instructional software, or calculators constitute a poor basis for developing recommendations for policy and practice. The critical issue is not the mere presence or absence of particular types of tasks or tools but how those tasks and tools are actually used in the classroom and what students are learning as they engage in those activities. In this regard, studies conducted to understand the independencies between various elements of the classroom system can inform all phases of experimental studies, including the formulation of good research questions and the delineation of potentially meaningful independent variables. In addition, observational and design research studies that treat classrooms as instructional systems can explain variance in implementation or, in other words, why different enactments of the same intervention lead to different student outcomes. Furthermore, studies conducted using these methodologies can address questions of the following type: "Can calculators be used to enhance students' mathematical learning and, if so, how and under what conditions?" Answers to such questions that are useful for both policy and practice situate particular types of tasks and tools firmly in the context of the classroom by specifying their role with respect to the other elements of the classroom system. In this way, we suggest that alternative methods are more appropriate for such education research questions than are experimental studies (National Research Council, 2002).

Our view that studies conducted using a range of methodologies can contribute to the improvement of mathematics teaching and learning should not be confused with arguments that oppose quantitative and qualitative approaches. Instead, our basic position is that different types of research questions lend themselves to different methodologies and that "research questions, not method, ought to drive educational research" (Ercikan & Roth, 2006, p. 21). In this regard, it is worth noting that studies conducted using other methodologies are often framed as nothing more than precursors to experimental studies. For example, it is sometimes assumed that the sole purpose of design research studies is to refine an intervention than can be evaluated experimentally. We have sought to clarify that although design research studies can serve this purpose, they and studies conducted using other methodologies can also produce forms of knowledge that have value in their own right and that can inform experimental studies and complement the forms of knowledge they produce.

In considering the approach the Panel used to develop its findings, we have questioned both the criteria the Panel used to identify trustworthy research findings and its focus on single variables in isolation. With regard to the second step of the Panel's approach, we concur with Good's (2008) observation that prior reform efforts have "failed because single variables (and even single themes) do not have an independent effect on student learning" (p. 3). It is unfortunate that the historical record from the early part of the last century and from the 1960s and 1970s did not give the Panel pause for thought, as the Panel does not support the assumption that experimental studies by themselves constitute adequate basis for a process of continual instructional improvement (Danziger, 1990).

Issues of Equity

Our discussion of methodological issues serves as the backdrop against which we consider the Panel's recommendations for improving the quality of mathematics instruction for all students. It is well documented that the aggregate groups of Native American, African American, and Hispanic students, and students whose families have low incomes, tend to achieve at lower levels in mathematics than do White and Asian students and students whose families have higher incomes (Darling-Hammond, 2007). The Panel recognized this pattern and offered five findings and recommendations that focus on addressing disparities in achievement across groups of students. We examine these recommendations in relation to the assumptions of the experimental methodology as outlined above and the nature of the concomitant explanations given for achievement disparities.

In its report, the Panel attributed achievement disparities between groups of students to the following differences: "[mathematical] knowledge that kindergartners bring to school" (Findings 8 and 9); "social, affective, and motivational factors" (Finding 13); and "children's goals and beliefs about learning" and their "engagement and sense of efficacy" (Finding 14). In doing so, the Panel accounted for achievement disparities in terms of whether particular groups of students tend to have more or less of these psychological attributes and traced some of these group differences to background variables such as socioeconomic status, race, or ethnicity (see, e.g., pp. 2-63–70).[3] The key point to note for our purposes is that explanations of this type locate the source of achievement disparities exclusively in particular groups of students, their families, and their communities. In our view, the Panel's formulation of such explanations reflects the underlying assumptions of the experimental methodology. The ontology of the experimental methodology appears to assume the sources of achievement disparities are differences between groups of students that are present at the outset of formal mathematics instruction or that manifest themselves during instruction. Crucially, mainstream mathematics instruction is not seen as a possible source of achievement disparities. As a consequence, it is not surprising that none of the Panel's recommendations for addressing achievement disparities focus on improving mainstream instructional practices.

The line of reasoning that resulted in the Panel's identifying individual psychological attributes as the source of achievement disparities can be questioned on two counts. First, the assumption that the different groups of students receive similar instruction ignores the well-documented fact that, in the aggregate, nondominant groups of students (e.g., non-White students, students who come from families with low incomes, students who do not speak English as a first language) do not receive the same opportunities to learn in school as do students from dominant backgrounds (Darling-Hammond, 2007). None of the report's recommendations acknowledge the disparities in access to structural resources (e.g., types of curriculum, access to higher level mathematics courses, class size) that exist across schools and districts in the United States (Darling-Hammond, 2007). Although the Panel recognized that there are differences in the quality of mathematics teachers across schools and districts, it did not discuss this issue in relation to achievement disparities. The report of the Task Group on Learning Processes discusses inequities in access to structural resources for groups of students who tend to achieve at different levels. For example, this task group acknowledged differences in both "tangible resources such as availability of quality teachers and textbooks" (p. 2-57) and "school-based factors such as features of teaching and learning contexts" (p. 2-52). However, the Panel did not pick up these leads when accounting for achievement disparities and proposing solutions.

The second reason for challenging the Panel's reasoning about the source of achievement disparities is the assumption that the nature of mainstream U.S. mathematics instruction plays no role in the differential manifestation of characteristics such as motivation and efficacy across groups of students during instruction. An extensive body of research has investigated how mathematics classrooms (viewed as instructional systems) can be organized to support all students' engagement (see, e.g., Boaler, 1997, 2002; P. Cobb, Gresalfi, & Hodge, in press; Enyedy & Mukhopadhyay, 2007; Martin, 2000; Nasir, 2002). This work indicates that the differential manifestation of motivation and efficacy is related to how mathematics classrooms are structured. It therefore supports the view that motivation and efficacy are relations between students and particular ways of organizing mathematics instruction rather than relatively stable attributes of students. From this perspective, it is reasonable to treat the documented disparities in motivation and efficacy as outcomes of mainstream instruction (Nicholls, 1989). This conclusion then orients us to draw on high-quality research conducted using a range of methodologies, including experimental studies, to consider how mathematics instruction might be organized to support all students' learning. In our view, it is unfortunate that the Panel's dogged adherence to experimentalism restricted its explanation of disparities in mathematics achievement and thus the solutions it proposed.

The Panel's specific recommendations regarding achievement disparities are far removed from the how of practice. For example, the Panel concluded that "social and intellectual support from peers and teachers is associated with higher mathematics performance for all students, and that such support is especially important for many African-American and Hispanic students" and called for experimental evaluations of such interventions (p. xix; see also pp. 2-22, 2-52–53, 2-75–76). It is unclear how these conclusions might be useful to school and district leaders and to teachers. We are left wondering about the specific ways in which teachers might modify their classroom practices to provide social and intellectual support to particular groups of students.

As a final observation about the Panel's treatment of issues of equity, many school districts across the country are serving increasing numbers of students whose first language is not English, including districts that previously only served English-dominant students (Capps et al., 2005). We were surprised to find only one instance across all the materials the Panel produced that explicitly addressed English language learners—the Task Group on Assessment noted that it is important to keep issues of language in mind when creating items, particularly for English language learners (pp. 6-10, 6-44; see also p. 61). The research on how to support students' language development in the context of teaching mathematics is thin. The Panel could have provided important leadership by noting the pressing need for research in this area.

Textbooks and Mathematical Content

Having discussed the Panel's recommendations for addressing disparities in achievement, we now turn our attention to its recommendations for textbooks and mathematics content for mainstream instruction. In doing so, we first critique the limited depth of the Panel's analysis of textbooks and then consider its recommendations for streamlining the K–8 mathematics curriculum.

Relatively few experimental studies of the effectiveness of mathematics curricula met the Panel's criteria for its view of what constitutes scientific evidence. The Panel also indicated a problem with many of the curriculum studies it reviewed when it recommended that work be undertaken to develop appropriate assessment instruments of students' mathematical learning. Schoenfeld's (2006) recent critique of such studies indicates that the quality of the assessment instruments used in such studies is an endemic problem. We therefore wonder why the Panel nonetheless included studies that are flawed in this way in the synthesis with which it attempted to develop policy recommendations. The Panel bemoaned the poor design of many of the experimental studies it reviewed. This is surely a legitimate concern. However, sophistication in experimental design cannot compensate for the use of inappropriate or inadequate measures. To cite the well-known adage: garbage in, garbage out.

In addition to reviewing relevant experimental studies, the Panel conducted what might be termed a conceptual analysis of textbooks. This analysis brings two substantive issues to the fore, namely, the frequent occurrence of mathematical errors in textbooks and the lack of coherence of some textbook series (i.e., prior topics do not cover the prerequisites for subsequent topics). These weaknesses are indeed reason for disquiet. However, the Panel also devoted much of its analysis to issues of form rather than content. The concerns the Panel raised include the length of mathematics textbooks, the number of pictures, and the titles of lessons and instructional units. The Panel was preoccupied with the form of textbooks and overlooked the fact that different textbook series that have been evaluated experimentally were designed to support students' achievement of different types of instructional goals. For example, the intent of some elementary textbook series is to support students in learning symbolic procedures for adding, subtracting, multiplying, and dividing fractions, whereas the intent of other series is to support students in learning to reason symbolically about fractions as quantities. This is a nontrivial distinction. It is disappointing that the Panel failed to take account of this critical difference when conducting its conceptual analysis, given that the delineation of worthwhile instructional goals serves to specify what counts as instructional improvement and thus what constitutes adequate assessment measures. This is particularly the case given that specification of worthwhile instructional goals could have oriented a subsequent analysis that sought to determine whether the content of particular textbook series constitutes a coherent trajectory for students' mathematical learning that aims toward those goals. Analyses of this type complement the findings of experimental studies and are especially relevant when the experimental research base is inadequate (Confrey, 2006). The Panel could have provided a service to the research and practitioner communities by taking its analysis of the coherence of textbook series one step further. In particular, the Panel might have focused not merely on whether prerequisites are covered but also on the quality of sequences of instructional tasks as supports for student learning. In doing so, the Panel would have modeled the processes of delineating worthwhile instructional goals and of assessing the design of textbook series as supports for student learning trajectories that aim toward those goals.

In addition to reviewing textbooks, the Panel made far-reaching recommendations for the reorganization of the mathematics curriculum. The Panel clarified in the opening pages of its report that "the mathematics curriculum in Grades K–8 should be streamlined and should emphasize a well-defined set of the most critical topics in the early grades" (p. xiii). The Panel acted on this proposal by arguing that success in high school algebra, specifically Algebra II, is critical. The Panel then mapped backward from algebra to identify topics that are critical prerequisites at earlier grade levels. These critical topics include whole-number arithmetic, fractions, and selected aspects of geometry. The Panel is to be commended for taking seriously the role of algebra as a gatekeeper to future educational and economic opportunities (cf. Moses, Kamii, Swap, & Howard, 1989).

The Panel took care to state that the almost exclusive focus of its content recommendations on arithmetic and algebra should not be taken as specifying a complete mathematics curriculum for Grades K–8. However, we fear that this caveat will be overlooked by many policy makers, especially given the listing of algebra and arithmetic topics (not learning goals) in Tables 1 and 2 of the report. In particular, we are concerned that the report quite literally says nothing about the place of statistics in the K–8 curriculum. In developing the rationale for its content recommendations, the Panel observed that "during most of the 20th century, the United States possessed peerless mathematical prowess" (p. xi). The Panel then raised the concern that "without substantial and sustained changes to its educational system, the United States will relinquish its leadership in the 21st century" (p. xi). However, the Panel seemed to assume that in identifying critical high school mathematics courses, it will necessarily delineate the key aspects of mathematical literacy. This assumption is open to question.

The ever increasing use of computers elevates the importance of quantitative reasoning in general and statistical reasoning in particular (G. W. Cobb & Moore, 1997). In addition to being a critical workforce demand, statistical literacy is also important from the perspective of equity. The Panel observed that citizens and policy leaders who deal with the public interest "will surely have to handle quantitative concepts more fully and more deftly than at present" (p. xii). As the Panel's report demonstrates, public policy proposals typically involve statistical argumentation. In many respects, statistical discourse has become the language of power in the public policy arena. As a consequence, inability to participate in this discourse results in de facto disenfranchisement. Seen in this light, statistical literacy that involves developing and critiquing relatively sophisticated data-based arguments relates directly to both equity and participatory democracy. In our view, this constitutes a compelling reason for considering seriously the role of statistics, as well as algebra, when making recommendations for the K–8 mathematics curriculum.

Graduate Education

Our primary goal to this point has been to highlight the consequences of relying on research conducted using a single methodology when developing recommendations for policy and practice, especially when there appears to be little recognition of the limitations of that methodology. In making this argument, we have stressed that we consider the experimental methodology to be a valuable research tool and have spoken of its potential contributions to instructional improvement. We now focus on the Panel's recommendations for graduate education in order to consider what might be necessary for this potential to be realized.

The Panel's primary recommendation for graduate education is that "the rigor and amount of coursework in statistics and experimental design be increased in graduate training education" (p. 7-10). There are two observations that we want to make about this recommendation. First and most obviously, this recommendation sits uncomfortably with the Panel's recommendations for streamlining the K–8 mathematics curriculum and reinforces our contention that statistics is a core aspect of quantitative literacy.

Second and more substantially, we concur with the Panel's judgment that the current capacity of the education research community to conduct high-quality experimental studies is inadequate, but we question whether increased coursework in statistics and experimental design will, by itself, solve the problem. This becomes apparent when we note Moore's (1996) observation that data are numbers plus context. In other words, data are measures of attributes of a situation that are judged to be relevant given the question under investigation. Moore's intent in making this observation was to emphasize that a relatively deep understanding of the phenomenon under investigation is central to proficient statistical practice. This understanding is implicated in the delineation of relevant attributes of the situation under investigation, in the construction of adequate measures of those attributes, and in the interpretation of the results of statistical analysis (Sloane & Gorard, 2003).

The Panel states, "It is self-evident that teachers cannot teach what they do not know" (p. xxi). In our view, it is equally self-evident that individual researchers and research teams cannot adequately investigate phenomena that they do not understand. At an absolute minimum, it would seem critical that researchers conducting experimental studies have a deep understanding of mathematical knowledge for teaching at the relevant grade levels. In the case of studies of classroom interventions, it also seems critical that researchers have a firm grasp of how classrooms can function as instructional systems to support all students' mathematical learning. Clearly, statistics and experimental design are important elements of a solid graduate education. It is unfortunate that the Panel did not consider the additional prerequisite competencies that are necessary if we are to increase the capacity of the research community to conduct high-quality experimental studies. The experimental methodology will not realize its potential to contribute to instructional improvement unless we broaden our view of the critical elements of graduate programs that aim to prepare future researchers.

Conclusion

In responding to the National Mathematics Advisory Panel's (2008) report, we have made it clear that experimental studies can make a critical contribution to the improvement of mathematics teaching and learning. However, we have questioned the Panel's unflagging adherence to experimentalism, arguing that it adversely affects the quality and usefulness of the recommendations the Panel has produced for both policy and practice. We have contrasted the Panel's position with a more moderate approach that seeks to capitalize on the findings of high-quality studies that use a range of methodologies. Toulmin (1963) noted long ago that explanation and understanding are the driving forces of scientific inquiry and that prediction and control are by-products of the resulting foresight. Consistent with Toulmin's observation, we contend that a national agenda for mathematics teaching and learning should be informed both by studies that have sought to identify causal relations across a number of cases and by studies that treat classrooms as instructional systems and have sought to understand the processes through which and the conditions under which causal relations hold.

NOTES

We are grateful to Dale Farran and three anonymous reviewers for constructive comments on a previous draft of this article. We are, of course, solely accountable for the views expressed in the article.

[1]As actions speak louder then words, it is worth acknowledging that the first author is the principal investigator of a randomized field trial of a first-grade mathematics program and is also the director of a postdoctoral fellowship program designed to build capacity to conduct experimental studies.

[2]P. Cobb and Gravemeijer (in press) made an initial attempt in the case of design research by proposing key elements of a scheme of argumentation for this methodology.

[3]To its credit, the Task Group on Learning Processes recognized that although socioeconomic status (SES) is often cited as at the root of the achievement gap, "the mechanisms linking these broad constructs [e.g., parental education, poverty level, parental income, or a composite index] to mathematics learning and achievement are not well understood, nor are the relationships among SES, ethnicity, and mathematics learning" (p. 2-144). Unfortunately, this caveat was not integrated into the Panel's report.

REFERENCES

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *Elementary School Journal, 93,* 373–397.

Bernstein, R. J. (1983). *Beyond objectivism and relativism: Science, hermeneutics, and praxis.* Philadelphia: University of Pennsylvania Press.

Boaler, J. (1997). Reclaiming school mathematics: The girls fight back. *Gender and Education, 9*, 285-305.

Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education, 33*, 239-258.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*, 141-178.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104,* 801–823.

Cobb, P. (2001). Supporting the improvement of learning and teaching in social and institutional context. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 455–478). Mahwah, NJ: Lawrence Erlbaum.

Cobb, P., & Gravemeijer, K. (in press). Experimenting to support and understand learning processes. In A. E. Kelly (Ed.), *Handbook of design research.* Mahwah, NJ: Lawrence Erlbaum Associates.

Cobb, P., Gresalfi, M., & Hodge, L. L. (in press). An interpretive scheme for analyzing the identities that students develop in mathematics classrooms. *Journal for Research in Mathematics Education.*

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21,* 1–78.

Cobb, P., & Smith, T. (in press). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In T. Wood, B. Jaworski, K. Krainer, P. Sullivan, & D. Tirosh (Eds.), *International handbook of mathematics teacher education.* Rotterdam, Netherlands: Sense.

Coburn, C. E. (2005a). The role of non-system actors in the relationship between policy and practice: The case of reading instruction in California. *Educational Evaluation and Policy Analysis, 27,* 23–52.

Coburn, C. E. (2005b). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy, 19,* 476–509.

Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education, 113,* 515–548.

Confrey, J. (2006). Comparing and contrasting the National Research Council report on Evaluating Curricula Effectiveness with the What Works Clearinghouse approach. *Educational Evaluation and Policy Analysis, 28*, 195-213.

Confrey, J., & Smith, E. (1995). Splitting, covariation, and their role in the development of exponential functions. *Journal for Research in Mathematics Education, 26,* 66–86.

Danziger, K. (1990). *Constructing the subject.* New York: Cambridge University Press.

Darling-Hammond, L. (2007). The flat earth and education: How America's commitment to equity will determine our future. *Educational Researcher, 36*(6), 318–334.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: New Press.

Eisenhart, M. A., & Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher, 32*(8), 31–38.

Elmore, R. F. (2006, June). *Leadership as the practice of improvement.* Paper presented at the OECD International Conference on Perspectives on Leadership for Systemic Improvement, London.

Enyedy, N., & Mukhopadhyay, S. (2007). They don't show nothing I didn't know: Emergent tensions between culturally relevant pedagogy and mathematics pedagogy. *Journal of the Learning Sciences, 16,* 139-174.

Ercikan, K., & Roth, W.-M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher, 35*(5), 14–23.

Fennema, E., Franke, M. L., & Carpenter, T. P. (1993). Using children's mathematical knowledge in instruction. *American Educational Research Journal, 30,* 555–583.

Glennan, T. K., Jr., & Resnick, L. B. (2004). School districts as learning organizations: A strategy for scaling educational reform. In J. T. K. Glennan, S. J. Bodilly, J. Galegher, & K. Kerr (Eds.), *Expanding the reach of education reforms: Collected essays by leaders in the scale-up of educational interventions* (pp. 517–564). Santa Monica, CA: RAND.

Good, T. L. (in press). Taching mathematics in grades 3-5 classrooms. *Teachers College Record.*

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences, 13,* 115–128.

Kelly, A. E., & Yin, R. K. (2007). Strengthening structured abstracts for education research: The need for claim-based structured abstracts. *Educational Researcher, 36,* 133–138.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-195). Cambridge, UK: Cambridge University Press.

Lampert, M. (2001). *Teaching problems and the problems of teaching.* New Haven, CT: Yale University Press.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41,* 635–679.

Martin, D. B. (2000). *Mathematics success and failure among African-American youth*. Mahwah, NJ: Erlbaum.

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3–11.

McClain, K. (2002). Teacher's and students' understanding: The role of tool use in communication. *Journal of the Learning Sciences, 11,* 217–249.

McLaughlin, M. W. (2006). Implementation research in education: Lessons learned, lingering questions, and new opportunities. In M. I. Honig (Ed.), *New directions in educational policy implementation* (pp. 209–228). Albany: State University of New York Press.

Moore, D. S. (1996). New pedagogy and new content: The case of statistics. In B. Phillips (Ed.), *Papers on statistics education* (pp. 74-96). Hawthorne, Australia: Swinburne.

Moses, R. P., Kamii, M., Swap, S. M., & Howard, J. (1989). The Algebra Project: Organizing in the spirit of Ella. *Harvard Educational Review, 59,* 423–443.

Nasir, N. S. (2002). Identity, goals, and learning: Mathematics in cultural practice. *Mathematical Thinking and Learning, 4,* 213-248.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

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

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K-12 mathematics evaluations.* Washington, DC: The National Academies Press.

Nelson, B. S., & Sassi, A. (2005). *The effective principal: Instructional leadership for high-quality learning*. New York: Teachers College Press.

Nicholls, J. (1989). *The competitive ethos and democratic education.* Cambridge, MA: Harvard University Press.

Popper, K. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press.

Resnick, L., Besterfield-Sacre, M., Mehalik, M., Sherer, J. Z., & Halverson, E. (2007). A framework for effective management of school systems performance. In P. A. Moss (Ed.), *Evidence and decision making: 106th yearbook of the Society for the Study of Education* (pp. 155–185). Malden, MA: Blackwell.

Schoenfeld, A. H. (2006). What doesn't work: The challenges and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35*(2), 13–21.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Slavin, R. E. (2004). Educational research can and must address "what works" questions. *Educational Researcher, 33*(1), 27–28.

Sloane, F. C., & Gorard, S. (2003). Exploring modeling aspects of design experiments. *Educational Researcher, 32*(1), 29–31.

Spillane, J. P. (2005). *Distributed leadership.* San Francisco: Jossey Bass.

Stigler, J. W., & Hiebert, J. I. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Toulmin, S. (1963). *Foresight and understanding.* New York: Harper Torchbook.

Yin, R. K. (2003). *Case study research* (3rd ed.). Thousand Oaks, CA: Sage.

AUTHORS

PAUL COBB is a professor of mathematics education at Vanderbilt University, Peabody College, Box 330, Nashville, TN 37203; *paul.cobb@vanderbilt.edu*. His interests are in classroom instructional processes and learning policy.

KARA JACKSON is a postdoctoral research fellow in the Department of Teaching and Learning at Vanderbilt University, Peabody College, Box 330, Nashville, TN 37203; *kara.j.jackson@vanderbilt.edu*. Her interests are mathematics learning, socialization, and identity development across contexts and how to support teachers to provide access to equitable learning opportunities in mathematics for all students.

Manuscript received July 18, 2008
Revision received October 10, 2008
Accepted October 13, 2008