



Breaching the Conditions for Success for a National Advisory Panel

Jere Confrey, Alan P. Maloney, and Kenny H. Nguyen

The authors identify six conditions for success for the work of high-level national panels and identify breaches in these conditions in the recent *Foundations for Success: The Final Report of the National Mathematics Advisory Panel* (2008). They question the trustworthiness, validity, and intellectual integrity of its findings and advice to the nation because of (a) an inappropriate composition of Panel expertise and biased selection of literature (Condition 1); (b) failure to appropriately and consistently apply methodological standards (Condition 2); and (c) inconsistencies between the task group and subcommittee reports and the final report (Condition 6). In asking what difference these breaches make, the authors recount recent events suggesting that these breaches have already contributed to degrading the national discussion of curriculum standards for K–12 mathematics education.

Keywords: conditions for success; mathematics education; methodological standards; National Mathematics Advisory Panel; policy

Only rarely is a field of intellectual endeavor the subject of a presidential executive order. The National Mathematics Advisory Panel (NMAP) was charged by President George W. Bush with advising him and the U.S. secretary of education on how to “foster greater knowledge of and improved performance in mathematics among American students . . . with respect to the conduct, evaluation, and effective use of the results of research relating to proven-effective and evidence-based mathematics instruction” (Executive Order 13398, cited in NMAP, 2008, p. 7). The responsibility for “appointment of members and oversight of the Panel” in the executive order (April 18, 2006) was assigned to the secretary of education. On March 13, 2008, the NMAP released its report, titled *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*.

Preparation of valuable and credible policy recommendations must be based on the exercise of the highest standards of scientific and technical quality among a panel of experts. To accomplish this, we propose six *conditions for success* that must underlie the process of panel composition and deliberation:

1. Appropriate composition of an expert panel representing the relevant fields;
2. Articulation of methodological standards that are appropriate for the questions under investigation and are applied consistently and fairly for the duration of the panel’s work;
3. Solicitation of appropriate testimony and documented responsiveness to it;
4. Provision of adequate time and opportunity for deliberation, and support to enable the panel to complete its charge;
5. Adequate independent external review for quality, impartiality, and objectivity; and
6. Adequate internal review for internal consistency among preliminary reports and final products in terms of accuracy, conciseness, readability, and intellectual integrity.

A breach of these conditions will compromise the trustworthiness of such a report. In this article, we document ways in which Conditions 1, 2, and 6, above, were breached in the NMAP report, and we discuss how the report has had detrimental effects on policy-related activity in the field.¹

Condition 1. Composition of the Panel

The composition of the NMAP was unusual, considering the charge. The chair of the Panel was a chemist and distinguished emeritus university president. The other Panel members included six psychologists, four mathematics educators (one of these added during the final year), four mathematicians, one special educator, one middle school mathematics teacher, one policy researcher, and one reading researcher (U.S. Department of Education, 2008). This means that, of 19 members, only 5 (the mathematics educators and the mathematics teacher) regularly had sustained interactions with mathematics instruction at the K–12 level; fewer than half the Panel members had documented academic preparation in mathematics. In the area of research, more of the experience on the Panel was based in psychology (cognitive and developmental) than in any other area. Mathematics education as a field includes experts in sociology, anthropology, and critical theory, but the Panel was deficient in expertise in those areas. Because of the documented and sizable achievement gaps in education related to race, socioeconomic status, and second-language learners, representation of those fields of expertise would ordinarily be expected for a panel of this kind. Moreover, because of the scope of the task demanded of the Panel in a limited time frame, much of the work was carried out by five task

groups and three subcommittees; and the absence of relevant expertise was even more marked in the composition of the subcommittees and task groups.

Condition 2. Articulation and Consistent Application of Appropriate Methodological Standards

Methodologically, the Panel was charged with marshaling “the best available scientific evidence and offer[ing] advice on the effective use of the results of research related to proven, effective and evidence-based mathematics instruction” (NMAP, 2008, p. 81). Examining the methodological standards for rigor in the NMAP report is challenging, because the standards are presented in multiple locations, and because they vary across those locations. To write this article, we had to repeatedly review the Panel’s statements on methodology. The global standards of evidence were reported by the Subcommittee on Standards of Evidence (Reyna, Benbow, Boykin, & Whitehurst, 2008), then summarized in Appendix C of the report (NMAP, 2008, pp. 81–85) and in the Executive Summary (NMAP, 2008, pp. xv–xvi). In addition, each task group reported its own methodological decisions, based on the subcommittee’s statement that

to ensure identification of the best available evidence in the research literature, each task group has developed guidelines for the literature search that identify the relevant topics and the screening criteria to be used to select the studies the task group will consider for review. (p. 4)

Thus, to discuss the methodological quality of the NMAP report overall, we begin by examining the global standards of evidence proposed by the Subcommittee on Standards of Evidence and then describe how these standards varied across different task group reports.

The Subcommittee on Standards of Evidence described three study categories representing relative levels of confidence. It stated the highest standard (Category 1) as follows: “The Panel’s strongest confidence will be reserved for studies that test hypotheses, that meet the highest methodological standards (internal validity), and that have been replicated with diverse samples of students under conditions that warrant generalization (external validity)” (Reyna et al., 2008, p. 1). Category 2 included “promising or suggestive studies that do not meet the highest standards of scientific evidence, but . . . represent sound, scientific research that needs to be further investigated or extended” (p. 1). The two examples cited in this category were laboratory studies, a stance that reveals the Panel’s assumption that the route to insight into classroom practice is through the laboratory rather than through the use of design studies in which the complexity of practice is studied in situ (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Confrey, 2006; Simon, 2000). Category 3 studies, as described by the Panel, included those “based on values or weak evidence; these are essentially unfounded claims and will be designated as opinions as opposed to scientifically justified conclusions” (Reyna et al., 2008, p. 1). It seems clear that, by the choice and description of three categories, the subcommittee conveyed disdain for non-“experimental” research. Its disregard for nonexperimental studies is carried forward in its recommendations for future research support.

Although the subcommittee did permit task groups to “select their own studies” and “develop their own screening criteria,” the Executive Summary states that the task groups used methodological standards consistent with the global ones: “One of the subcommittee reports covers global considerations relating to standards of evidence, while individual task group reports *amplify* [italics added] the standards in the particular context of each task group’s work” (NMAP, 2008, p. xvi). The choice of the word *amplify* implies an intensification of the global standards, but based on our review, the verb would more correctly be *modify*. Task groups modified the global standards significantly, which, we will argue, was appropriate to their tasks. However, this resulted in considerable variability among the task groups’ methodological standards of evidence. As a result, the portrayal of the standards of evidence in the Executive Summary is not only inaccurate but also misleading. Thus we question whether the methodological conditions for success—those of consistency and fairness—have been met in the overall presentation of the report and Executive Summary.

For example, the methodology section of the report by the Task Group on Instructional Practices refers to the global standards but then properly asserts:

However, it is particularly germane to this topic, in that before requiring widespread implementation of a particular instructional practice or intervention, or committing significant resources toward such implementation, it seems critical to know that it will in all likelihood lead to higher levels of mathematics proficiency than alternatives. (Gersten, Ferrini-Mundy, Benbow, Clements, Loveless, Williams, et al., 2008, p. 199)

Furthermore, the task group quotes *Scientific Research in Education* (National Research Council, 2002), saying that “[the authors] believe that attention to the development and systematic testing of theories and conjectures across multiple studies and using multiple methods—a key scientific principle . . . is currently undervalued in education relative to other scientific fields” (as quoted in Gersten et al., 2008, p. 199). The task group’s reference to the National Research Council statement seems to suggest that the task group is casting a broader methodological net. Yet when explicating its own standards, the task group returns to a distinction, similar to that found in the global standards, between “Category 1, experimental and quasi-experimental studies based on What Works Clearinghouse (WWC) requirements with modest modifications,” and “Category 2, weak group comparison studies and other quantitative designs that attempt to infer causality.” The task group states that “Category 2 studies are used only when there is an insufficient body of information from the evidence provided by Category 1–level studies. Flawed studies can never compensate for high-quality experimental or quasi-experimental studies” (Gersten et al., 2008, p. 202).

Then, based on the paucity of “high-quality experimental research” and the earlier acknowledgment of a role for broader methodologies, the task group permits a brief discussion of Category 2 studies if “a pattern emerges that might be worthy of mentioning” and as long as such studies are “not used to make claims of causality or effectiveness” (Gersten et al., 2008, p. 202). Ultimately, the task group states, “Panelists were free to use any

type of research (descriptive, correlational, qualitative) to set the context for their meta-analysis” (p. 202).

Given the breadth of important and controversial studies reviewed by the Task Group on Instructional Practices, we would characterize these variations from the global standards *not* as amplifying those standards but as modifying them. The modifications appropriately permitted the task group to consider how alternative methods frame or constrain its results, a necessary means to nuance, contextualize, and limit the generalizability of the overall results and conclusions. These modifications, however appropriate and necessary, are not reflected in the description of the methodological standards in the final report and Executive Summary, with the result that they are inaccurate and misleading relative to the work actually conducted by the task groups. For example, the report of the Task Group on Instructional Practices contains nuanced discussions concerning contexts, samples, approaches, and possible differences with regard to students identified as “low achieving” and “students with disabilities.” It concludes that the following are defining features of effective instructional approaches:

1. Concrete and visual representations (mathematical drawings)
2. Explanations by teachers
3. Explanations and math talk by students in whole-class discussion
4. Students working together
5. Carefully orchestrated practice activities with feedback
6. High but reasonable expectations (Gersten et al., 2008, p. 76)

This list, by the Panel’s own categorization, includes features of both explicit and primarily implicit instructional approaches; yet the final report discusses only its “review of 26 high-quality studies,” stating simply that they indicate that “explicit methods of instruction are effective with LD [learning disabled] and LA [low-achieving] students.” It omits the promising results concerning implicit instruction. Thus we see how misleading conclusions result from the complicated methodological shifts between the individual task group reports and the final report.

In its report, the Task Group on Learning Processes chose to apply standards of evidence that can in no way be cast as “amplifications” of the global standards. The task group selected studies that tested explicit hypotheses about the mechanisms promoting the learning of declarative knowledge (arithmetic facts), procedural knowledge, and conceptual knowledge:

The multiple approaches, procedures, and study types reviewed and assessed with regard to convergent results include the following:

- Verbal report (e.g., of problem solving approaches).
- Reaction time and error patterns.
- Priming and implicit measures.
- Experimental manipulation of process mechanisms (e.g., random assignment to dual task, or practice conditions).
- Computer simulations of learning and cognition.
- Studies using brain imaging and related technologies.
- Large-scale longitudinal studies.
- International comparisons of math achievement.
- Process-oriented intervention studies. (Geary et al., 2008, p. 1)

We would argue that presenting these methods as an amplification of the global standards is not only inaccurate but misleading. Furthermore, modifying the standards to include these types of studies should have obligated the task group to include *all* studies on learning that fit these descriptions.

With support from the National Science Foundation, our research team is currently synthesizing the research on rational number reasoning, a topic that includes multiplication and division, fraction, decimal and percent, ratio, rate and proportion, area and volume, and similarity and scaling. All of these topics are covered in the report of the Task Group on Learning Processes (Geary et al., 2008). However, when we compared our database with the task group report’s citations on multiplication, division, and fractions, we found that between our database of more than 500 studies and the task group’s 159 citations, there were only 15 sources in common.² In seeking to explain the differences, we found that, despite the NMAP report’s focus on the learning of mathematics, of the total of 466 journal articles referenced in the task group’s report, mathematics education articles constituted only slightly more than 10%, whereas psychological accounts (i.e., those published in journals on psychology, child development, and educational psychology) constituted approximately 70% of the references. Of the 10 journals with 10 citations or more each, 9 (90%) were psychology-oriented.³ Only one leading research journal in mathematics education had more than 10 citations.

We sought an explanation for the paucity of mathematics education citations from the task group chair. We learned that, methodologically, while admitting a variety of study types—including verbal reports (e.g., of problem-solving approaches) and process-oriented intervention studies, as described in cognitive science—the task group had rejected the use of clinical interviews and design experiments. We informed the chair of the task group that we were conducting an analysis of its report for inclusion in this article. We inquired as to the differences between verbal reports and clinical interviews, and among process-oriented intervention studies, teaching experiments, and design studies, from the task group’s point of view. The chair responded as follows:

By verbal reports we mean things like children’s description of how they solved a problem—often included with experimenter observation, reaction times, or video coding—to their justification as to why something like $5 + 7 = 7 + 5$. Clinical interviews were the initial bases for some of these techniques, such as how kids solve arithmetic problems. The studies we used used more standardized approaches: a set of problems designed to result in a range of problem solving strategies, simultaneous use of reaction times to verify, etc.

Process oriented means we were interested in children’s learning or how they processed the information (e.g., Siegler & Stein). Many of these studies were lab based and not part of a formal curriculum. These latter studies were reviewed by the instructional practices group. (David Geary, personal e-mail, July 13, 2008)

When asked to confirm that these comments meant that clinical interviews, teaching experiments, and design experiments were excluded from the report of the Task Group on Learning Processes, the chair responded in the affirmative and said, “We did not consider teaching experiments to be part of our group’s charge unless one of the dependent measures assessed some aspect

of learning.” Based on this response it seems likely that Geary was unfamiliar with the methodology of “teaching experiments,” which are interactive studies of learning, not intervention programs that assess effects on student outcomes. However, those who regularly use these approaches know that design studies and teaching experiments are fundamentally about learning. Piaget (1976) wrote years ago about why he chose to eschew standardized interviews and instead pursue students’ thoughts and their spontaneous behaviors and interactions through clinical interviews; subsequently, both teaching experiments and design studies evolved in mathematics education as means to study learning in situ, where it is examined in relation to interactions among teachers and students, with access to tools and resources, and studied over longer periods of time through a process of conjecturing and testing (Brown, 1992; Cobb et al., 2003; Confrey, 2006; Steffe & Thompson, 2000).

In reporting this exchange, we do not question the task group chair’s expertise in his own discipline. Rather, we report it to show how the synthesis in this key area of mathematics education was deleteriously affected by the composition of the task group and by choices regarding the portions of the research base to draw upon. Reporting the exchange also shows how the integrity of the final report is undermined by assertions about standards of evidence that are contrary to the (appropriate) variations from those standards by task groups.

This analysis demonstrates that the variations in the standards of evidence in the report of the Task Group on Learning Processes should have permitted the inclusion of a number of studies that were not included, which originated with the mathematics education community. Doubtless, part of the reason, as predicted earlier, was the composition of the task group, which consisted entirely of psychologists (David C. Geary, A. Wade Boykin, Susan Embretson, Valerie Reyna, and Robert Siegler). Nonetheless, the resulting bias in the methodologies and research produced a task group report that overlooks a key body of research and discredits mathematics educators by omission.

Notwithstanding this blatant omission, the task group understood quite properly that the selection of methodology must be driven by the questions under investigation (National Research Council, 2002), not by a “global” set of standards specified a priori as a litmus test for evidence inclusion. The result, however, is that, judging by the methodologies actually used in the synthesis on learning processes, one cannot describe the NMAP report overall as conforming to the global standards. Nonetheless, the NMAP report and the Executive Summary imply that the global standards for a so-called “rigorous methodology” were applied across the board, effectively sweeping under the rug the matter of selection of investigation-appropriate methodology by task groups. Doing so in the report and Executive Summary both conceals the omission of a substantial volume of scholarly work in mathematics education and gives an unfair and misleading representation of the methodological standards applied by the Panel overall.

To summarize: The Panel initially imposed a rigorous “global standard” on the task groups. The task groups could not uphold the standard and, at the same time, respond appropriately and professionally to their assignments; therefore, they had to modify the standards. However, the final report and Executive

Summary leave the impression that the task groups not only upheld the global standards but, indeed, amplified them. We claim that behavior like this is indicative of ideological attachment to a “degenerating paradigm” (Lakatos, 1970), one that is constantly generating anomalies. The degenerating paradigm in this case is based on the assumption that the only source of “rigorous,” high-quality studies is “experimentation.” Had the Panel recognized that rigor accrues from careful consideration of the results of studies based in multiple methodologies (National Research Council, 2002, 2004), and that synthesis requires one to draw conclusions across such studies, the report could have provided stronger and more comprehensive advice.

We claim that our analysis demonstrates that the conduct of the syntheses in the NMAP report represents a breach of the second condition for success: that methodological standards appropriate for the questions under investigation be articulated and applied consistently and fairly for the duration of a panel’s work.

Condition 6. Adequate Internal Review for Internal Consistency, Accuracy, Conciseness, Readability, and Intellectual Integrity

We also consider Conditions 3–5 to be essential for successful preparation of a synthesis report of this kind; however, we have no access to information on how fully they were carried out in the NMAP process. Thus we turn to Condition 6.

The Executive Summary and the report are often the only parts of a panel’s work that are broadly read. It is essential that these components exhibit consistency with the work of the subcommittees and task groups and that the necessary condensation of those more extensive and detailed reports accurately reflect their content and findings. In the case of the Executive Summary and report, a number of concerns merit examination.

Inadequate Referencing in the Report

The referencing in the report is woefully inadequate and unscholarly. Very few citations are included (43 for a 65-page text, including an unpublished manuscript). Most of the references cited are broad descriptions (books and reports) of national and international trends in performance and do not pertain to the evidence base for the report’s recommendations or findings, but instead merely set the context for the report.

New Sources of Authority Introduced

New sources of authority are introduced in the report that countermand or trump the supposed standards of rigor. The report states:

The Panel also took into consideration the structure of mathematics itself, which *requires* [italics added] teaching a sequence of major topics (from whole numbers to fractions, from positive numbers to negative numbers, and from the arithmetic of the rational numbers to algebra) and an increasingly complex progression from specific number computations to symbolic computations. The structural reasons for this sequence and its increasing complexity *dictate* [italics added] what must be taught and learned before students take coursework in Algebra. (NMAP, 2008, p. 17)

Such a statement may appear harmless and even logical to many, but the view that there is a single legitimate logical

progression to mathematics, and further, that by itself the logic of mathematics dictates instructional approaches and learning, has been subjected to sustained challenges by years of research in mathematics education.

One such challenge comes from within the mathematics community itself. Lynn Steen (2007), former president of the Mathematics Association of America, directly challenged the secondary school practice of always sequencing curricular decisions toward preparation for calculus. Similar questions have been raised by other applied mathematicians and members of the mathematical client disciplines, especially in light of the introduction of new technologies. Some mathematics educators, statisticians, and statistical educators focus on quantitative reasoning (Thompson, 1993), modeling (Lehrer & Schauble, 2006), integrated mathematics (Hirsch, Fey, Hart, Schoen, & Watkins, 1996), realistic mathematics (Gravemeijer, 1995; Streefland, 1991), measurement (Davydov, 1990), and data and statistics (Shaughnessy, 2007); their research demonstrates that students very effectively learn mathematics and reason mathematically by different trajectories. Changing the orientation of the mathematics enterprise—and, in doing so, acknowledging a role for judgment, preference, and values in one's purpose for teaching of mathematics—challenges the assumption that a singular portrayal of mathematical logic can ever “dictate” educational practice.

A second challenge concerns the question of the relationship between the logical structure and sequence of mathematics and the pathways for learning. Since the time of Piaget and the development of constructivism—a widely held theoretical view expunged by the Panel in its discussion of theory (NMAP, 2008, p. 58)—scholars in mathematics education have empirically documented that the path of development of mathematical knowledge does *not* mirror the structure of formal mathematics in any simple way. This view is often accompanied by the observation that “children are not miniature or incomplete scientists or mathematicians.” With the onset of sociocultural and situated approaches, researchers have challenged assumptions about the order and development of fundamental cognitive processes and have recognized the key role of tasks, materials, and tools in influencing the sequencing of learning. For example, Moss and Case (1999) introduced percentage in advance of fractions and decimals, discovering that this approach, drawing on the children's exposure to percentages in society, produced impressive learning results with children. Likewise, Confrey (1988), using ratio to introduce fractions, also saw strong student outcomes (Confrey & Scarano, 1995). These researchers argued that although the logical structure of mathematics should always be taken into careful consideration in designing school mathematics instruction, structure can neither “require” nor “dictate” issues of teaching and learning. Careful theoretical and empirical study of children's learning of mathematics should be the guide for instructional decisions (Cobb & Steffe, 1983; Confrey, 1990; Piaget, 1970).

The introduction of this “new source of authority” in the report resulted in another breach of the standards of evidence. It permitted assertions by mathematicians to be automatically included in the report, without even requiring peer review. For example, in multiple places an unpublished manuscript and presentations by an NMAP mathematician are used as warrants for claims, contrary to the stated standards of rigor.

Inadequate Distinction Among Types of Evidence

In the report—the document that policy makers will see—it is impossible to distinguish among summary statements, opinions, and evidence-backed statements (or to know that different standards underlie the various evidence-backed statements, making some of them more trustworthy than others). Even the serious reader is confused upon seeing the following statement:

A small number of questions have been deemed to have such currency as to require comment from the Panel, even if the scientific evidence was not sufficient to justify research-based findings. In those instances, the Panel has spoken on the basis of collective professional judgment, but it has also endeavored to minimize the number and scope of such comments. (NMAP, 2008, p. 12)

Although such statements are not unusual in policy documents, collective judgment must be carefully distinguished from evidence-based claims.

Summary of the Breaches of Conditions for Success

The breaches we have discussed, which are related to Conditions 1, 2, and 6 and which occur throughout the report and Executive Summary—nonrepresentative Panel composition, dogmatic claims of adherence to a single standard for evidence, inconsistencies in methodological standards, restricted and poor referencing of research, unjustified new sources of authority, and failure to distinguish among the sources of evidence—produce a very disturbing result. Most of the public, and especially policy makers and decision makers, will read the report and take the findings and recommendations at face value. Therefore, the report taken as a whole—and despite the considerable contributions and good-faith efforts of many Panel members—lacks intellectual integrity. A key place for holding the line against such breaches is in the report and Executive Summary.

Having participated in and chaired reports from the National Research Council (2002, 2004), the first author of this article knows firsthand the quality of reports that can be produced from a well-balanced, diverse, representative, and independent panel, and knows as well that it is the responsibility of each panel member to refuse to sign off and permit such a report to proceed to publication if it contains breaches in the conditions for success. Such breaches represent compromises that exceed or deviate from proper professional standards.

The NMAP report's release was anticipated to have major implications for educational policy, instructional practice, and research support and directions. We contend that because of the breaches in the conditions for success, the conclusions and recommendations of the report are suspect. We are unable in the space of this article to specify how the breaches of conditions for success of the NMAP's work affect those conclusions and recommendations.

Readers should consider that the points raised in this critique are not unique to mathematics education. However, if the political and policy climate is permitted to compromise the scholarly standards that are the bulwark of our profession, such that the conditions for success of such an important report are neglected or circumvented, we are all at risk.

Does It Matter?

William James, in *Pragmatism* (1907), wisely argued that an expressed difference (or distinction) should make a difference: “If no practical difference whatever can be traced [between the different notions in a dispute], then the alternatives mean practically the same thing, and all dispute is idle” (p. 45). In the remainder of this article, we discuss how a report that incorporates breaches in the conditions for success can be detrimental to its charge, in this case the charge to “foster greater knowledge of and improved performance in mathematics among American students . . . with respect to the conduct, evaluation, and effective use of the results of research relating to proven-effective and evidence-based mathematics instruction” (NMAP, 2008, p. 7). Such a report can be detrimental in relation to subsequent activities that take place outside or beyond the Panel.

Does it matter that a report is published containing serious breaches of the conditions of success? The answer is yes. The Missouri Department of Elementary and Secondary Education (DESE), in cooperation with the Mathematics, Engineering, Technology and Science (METS) Alliance, recently invited public comment on a proposed new set of K–12 mathematics curriculum standards, the product of many months of interdisciplinary and collaborative effort informed by the work of other states and several national organizations and agencies. A set of 39 faculty members, mostly mathematicians, including a former University of Missouri System president, decided that these proposed standards were misguided. In a letter sent to DESE and METS administrators (and shared with local news media), they cited the NMAP report multiple times. For example, they wrote:

The proposed Missouri K–12 document is based narrowly and almost exclusively on the (National Council of Teachers of Mathematics) standards that were the motivation of much of the mathematics curriculum work of the LAST decade rather than the work of the (National Mathematics Advisory Panel) for the NEXT decade. (Catching, 2008)

Commenting on the letter, an NMAP member stated, “I’d hate to see any state, especially my own, not take advantage of the NMAP report” (Heavin, 2008). The newspaper article also stated, “The math professors question whether those who write the state curriculum are fully qualified to do so.”

Barbara Reys, a distinguished professor in mathematics education who co-chaired the Missouri standards-writing group, reported that (a) the proposed standards were based on a review of *multiple* sets of standards (the National Council of Teachers of Mathematics standards, the Focal Points, the College Board Standards, the ACHIEVE standards, and the GAISE report, as well as the NMAP report); (b) the proposed standards were available for public comment for more than two months; and (c) not a single one of the letter’s signers had posted any feedback, constructive or otherwise, nor had any of them contacted the committee members with suggestions or comments. Reys wrote:

It is curious that those who signed the letter did not feel it necessary, productive, or professional to approach any of the members of the writing group or the commissioning bodies . . . prior to sending their

concerns to the Commissioners of the DESE and DHE [Department of Higher Education]. . . . It appears they are more interested in media attention than in improving the quality of the DRAFT document. (B. Reys, personal communication, July 22, 2008)

This is just one example, like many others across the country, of how the flawed NMAP report is contributing to a marginalization of mathematics educators and to the neglect of decades of research on children’s learning of mathematics. In a larger sense, the report contributes to the deterioration of constructive, open deliberation among key parties. Education is a widely distributed practice; progress depends heavily on how the NMAP report is interpreted by policy makers and practitioners across the nation, in state departments of education, and in school boards. Therefore, a flawed report can have widespread deleterious ramifications for policy and practice.

Perhaps even more fundamental, however, are the ways in which the breaches of the conditions for success may have distorted the particular recommendations in the report. It will take much more time and a variety of critical reviews to discern the extent of the damage in this regard. Because the U.S. Department of Education was responsible for the composition and the oversight of the NMAP, it should be held responsible for breaching the conditions for success of the Panel and, in effect, undermining the potential benefits to the country of this endeavor. The result was a loss of opportunity for improving mathematics education in our nation.

NOTES

¹The other three conditions for success are not discussed here, as we are not in a position to evaluate the internal deliberations of the Panel and do not have access to any external review documents, if they exist.

²We have subsequently modified our database to include relevant references from the psychology literature.

³*Journal of Educational Psychology* (47), *Journal of Experimental Psychology* (31), *Child Development* (24), *Journal of Experimental Child Psychology* (23), *Cognition and Instruction* (20), *Developmental Psychology* (14), *Cognitive Psychology* (11), *Psychological Review* (10), and *Psychological Science* (10). The single mathematics education journal was *Journal for Research in Mathematics Education* (28).

REFERENCES

- 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.
- Catching, E. (2008, June 10). Math professors seek change in state’s K–12 math curriculum. *Columbia Missourian*. Retrieved June 12, 2008, from <http://www.columbiamissourian.com/stories/2008/06/10/missouri-math-professors-look-change-state-k-12-ma/>
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Cobb, P., & Steffe, L. P. (1983). The constructivist researcher as teacher and model builder. *Journal for Research in Mathematics Education*, 14(2), 83–94.
- Confrey, J. (1988). *Multiplication and splitting: Their role in understanding exponential functions*. Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, DeKalb, IL.

- Confrey, J. (1990). What constructivism implies for teaching. In C. Maher, R. Davis, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 107–124). Reston, VA: National Council of Teachers of Mathematics.
- Confrey, J. (2006). The evolution of design studies as methodology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 135–152). New York: Cambridge University Press.
- Confrey, J., & Scarano, G. H. (1995, October). *Splitting reexamined: Results from a three-year longitudinal study of children in grades three to five*. Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.
- Davydov, V. V. (Ed.). (1990). *Soviet studies in mathematics education: Vol. 2. Types of generalization in instruction: Logical and psychological problems in the structuring of school curricula*. Reston, VA: National Council of Teachers of Mathematics.
- Geary, D. C., Boykin, A. W., Embretson, S., Reyna, V., Siegler, R., Berch, D. B., et al. (2008). *Report of the Task Group on Learning Processes*. Washington, DC: U.S. Department of Education.
- Gersten, R., Ferrini-Mundy, J., Benbow, C., Clements, D. H., Loveless, T., Williams, V., et al. (2008). *Report of the Task Group on Instructional Practices*. Washington, DC: U.S. Department of Education.
- Gravemeijer, K. P. E. (1995). *Developing realistic mathematics instruction*. Utrecht, the Netherlands: Freudenthal Institute.
- Heavin, J. (2008, June 11). Professors' petition critical of state math guidelines. *Columbia Daily Tribune*. Retrieved June 12, 2008, from <http://www.columbiatribune.com/2008/Jun/20080611News006.asp>
- Hirsch, C. R., Fey, J. T., Hart, E. W., Schoen, H., L., & Watkins, A. E. (2008). *Core-plus mathematics: Contemporary mathematics in context: Course 2*. New York: McGraw-Hill.
- James, W. (1907). *Pragmatism: A new name for some old ways of thinking*. New York: Longman Green.
- 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.
- Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 371–388). New York: Cambridge University Press.
- Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education*, 30, 122–147.
- 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 Academies Press.
- National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations*. Washington, DC: National Academies Press.
- Piaget, J. (1970). *Genetic epistemology*. New York: Columbia University Press.
- Piaget, J. (1976). *The child's conception of the world*. Totowa, NJ: Littlefield, Adams.
- Reyna, V. F., Benbow, C. P., Boykin, A. W., & Whitehurst, G. R. (2008). *Report of the Subcommittee on Standards of Evidence*. Washington, DC: U.S. Department of Education.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1010). Charlotte, NC: Information Age.
- Simon, M. A. (2000). Research on the development of mathematics teachers: The teacher development experiment. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research in mathematics and science education*. Mahwah, NJ: Lawrence Erlbaum.
- Steen, L. A. (2007). Facing facts: Achieving balance in high school mathematics. *Mathematics Teacher*, 100, 86–95.
- Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In R. Lesh & A. E. Kelly (Eds.), *Research designs in mathematics and science education* (pp. 267–307). Hillsdale, NJ: Lawrence Erlbaum.
- Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm of developmental research*. Dordrecht, the Netherlands: Kluwer.
- Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *Educational Studies in Mathematics*, 25, 165–208.
- U.S. Department of Education. (2008). *Biographies of panel members—National Mathematics Advisory Panel*. Retrieved July 16, 2008, from <http://www.ed.gov/about/bdscomm/list/mathpanel/bios/index.html#panel>

AUTHORS

JERE CONFREY is the Joseph D. Moore Distinguished University Professor of Mathematics Education at North Carolina State University, Friday Institute for Educational Innovation, College of Education, 1890 Main Campus Drive, Raleigh, NC 27606; jere_confrey@ncsu.edu. Her research interests include analyzing national policy, designing diagnostic assessments in mathematics, and synthesizing research on rational number learning and reasoning.

ALAN P. MALONEY is a senior research fellow at North Carolina State University, Friday Institute for Educational Innovation, College of Education, 1890 Main Campus Drive, Raleigh, NC 27606; alan_maloney@ncsu.edu. His research interests include designing diagnostic assessments in mathematics, synthesizing research on rational number learning, and the use of technology on children's mathematics learning.

KENNY H. NGUYEN is a graduate student in mathematics education at North Carolina State University, College of Education, 1890 Main Campus Drive, Raleigh, NC 27606; khnguye2@ncsu.edu. His research interests include designing diagnostic assessments in mathematics, synthesizing research on rational number learning, and the learning sciences.

Manuscript received July 18, 2008
Revisions received October 21, 2008
Accepted October 21, 2008