Mathematicians and the National Eighth-Grade Test

Hyman Bass

The committee to develop specifications for the Voluntary National Mathematics Test included three mathematicians: Don Kreider, Jim Lewis, and me. I offer here some reflections on our work with that committee and on some larger educational issues encountered in that work. I thank Deborah Ball for helpful discussion of these issues.

First Reception of the Idea

In his 1997 State of the Union address President Clinton proposed an audacious education agenda. One of its highlights was an unprecedented Voluntary National Mathematics Test (VNMT), to be administered to eighth-grade students, as well as a fourth-grade test in reading. The VNMT was to be designed to provide individual students, families, and teachers with an indicator of what each student knows and is able to do with respect to challenging but appropriate tasks in mathematics at the eighth-grade level. Just as the fourth grade marks a passage from "learning to read" to "reading to learn", the eighth grade is seen as a critical gateway toward future mathematical and scientific study and achievement.

I think it is fair to say that first reactions to this proposal were mixed, often hesitant, in all of the current camps of educational debate. My first disposition was that this was not the best place to put top political priority. Professional preparation and development of teachers and improvement of curricula seemed to be much more urgent and central concerns. Moreover, given the inequity in opportunities to learn in this country, a national test might aggravate rather than support correction of these disparities by highlighting poor performance of students in underserved schools without providing the resources needed to change this inequity. At the same time, the existence of such a test seemed at first like a political given, and it would come with the highest presidential backing. On that premise, discussion shifted to how one might make this high profile VNMT an instrument to help advance these other agendas. Before describing that train of thought, let me say something about the debates on National Standards, debates which impinged on the VNMT.

There has been widespread and often confused and nonproductive debate about the NCTM Standards. The generality of their language, partly imposed by the current political inadmissibility of prescribing anything resembling a national curriculum in this country, leaves them open to diverse (often misguided) interpretations and implementations. Moreover, even persons most philosophically committed to the Standards recognize that they must inherently be viewed as a document subject to ongoing review and improvement. (In fact, a very substantial revision process, enlisting the views of many communities, including mathematicians, is currently under way.) The most virulent foes of the Standards often cite misguided implementations or misreadings of them and take these to be authoritative representations of what the Standards are or were intended to be. Adding confusion to this discourse is the lack of consensus about what constitutes appropriate and well-grounded evidence for claims. For example, among the unfortunate outcomes of these debates is a false antagonism proposed between the nature and acquisition of basic skills on the one hand and conceptual understanding and problem-solving skills

Hyman Bass is Adrain Professor of Mathematics at Columbia University. His e-mail address is `hb@cpw.math.columbia.edu`.

on the other. I know of no one who has reflected seriously on the improvement of mathematical education who does not believe that all of these things are to be desired. Indeed, most mathematicians believe that they are mutually dependent. To elevate this discussion and make it more grounded and constructive, how can we, in a national framework, give a more concrete, unambiguous, and consensual indicator of high curricular expectations for all students?

This was a possibility that some of us saw for the VNMT, provided that it was well designed and developed. If established, then the collection of mathematical tasks presented on this test would, de facto, represent over time a concrete and ample sampling of what informed professional thinking deemed to be challenging and appropriate mathematical expectations at the eighth-grade level. Moreover, the publication of the tests and supporting materials and the prompt feedback to students, families, and teachers could afford a rich source of diagnostic information and guidance for curriculum design by teachers. If the test were subscribed to by a large number of states and perhaps later extended to more grade levels, then, some argued, it could become a powerful resource for curricular and textbook improvement, as well as a resource for professional development of teachers focused more on curriculum, children's understandings, content, and the pedagogy of that content.

It was with these admittedly very speculative and optimistic hopes that some of us agreed to work in support of the VNMT, being vigilant about the quality controls in its development. The mathematics community felt that it would be crucial, in order to help assure the mathematical quality of the test items, to have professional mathematicians involved throughout the development and oversight processes, much more so than has been the pattern in development of other national educational policy documents and programs. A letter addressed to U.S. Secretary of Education Riley from the AMS Committees on Education and Science Policy expressed this concern. When Don Kreider and I were appointed to the Mathematics Committee (to develop specs for the mathematics test, a parallel committee being named for the reading test), I reiterated this concern to the Committee itself, and so Jim Lewis was added to our ranks.

The Composition and Guidelines of the Mathematics Committee

The mathematics and reading tests were under the jurisdiction of the Department of Education (DOE). To develop item and test specifications, DOE contracted MPR Associates (to handle administrative and logistical matters) and the Council of Chief State School Officers (CCSSO), a natural choice, since these would be key agents in decisions by the states on whether to adopt the voluntary national tests. A National Test Panel provided overall policy guidance and review for the project. Two content committees reporting to the test panel—one for math (to which this article refers) and one for reading—were constituted to draft specifications for the assessment, make recommendations regarding scoring and reporting, and contribute to the development of an ongoing research agenda. A Technical Advisory Group consisting of experts in educational measurement supported the work of both the panel and the content committees. The recommendations of the content committees, once approved by the panel, would guide the contractor named to develop, administer, evaluate, and update the actual tests.

The work of the content committees was bounded by some fundamental policy constraints, the principal one being that the tests be based on the framework of the DOE's National Assessment of Educational Progress (NAEP). For twenty-five years NAEP has been measuring national performance in mathematics and reading. It has a well-developed educational measurement methodology on which the Voluntary National Test development could rely. Also, the alignment with the NAEP framework affords the possibility to predict NAEP performance from VNT results. In the case of mathematics, correlation can also be made with TIMSS, because TIMSS was already designed to be linked to NAEP for U.S. students.

At the same time, NAEP and the VNMT are fundamentally different, in that NAEP is based on a modest sample of students and measures system performance, whereas the VNMT is designed to be taken by every student, with test results promptly shared with individual students, parents, and teachers. The VNMT will be not only an evaluation tool but, more importantly, a resource meant to support improved teaching and learning by promptly returning student work, publishing the tests, and providing booklets with sample test items accompanied by hypothetical student work and scoring rubrics.

Further initial conditions given to the Mathematics Committee required that the tests:

- assess the same high standards for all students;
- take place every spring;
- be 90 minutes in length;
- include approximately 80% machine-scored items and 20% non-machine-scored, including one extended written response item;
- call for students to spend approximately equal amounts of time on machine-scored and non-machine-scored items;
- allow for reliable and valid linking of individual student performance to the NAEP and TIMSS achievement levels;

- be scorable within three weeks;
- allow the results to be returned to students, families, and teachers within two months of test administration; and
- be released in its entirety following each administration.

Since concerns were expressed about the membership of the Mathematics Committee, I list here its members: John Dossey (chair) (Distinguished Univ. Prof. of Math., Illinois State U.), Hyman Bass (prof. math., Columbia U.), Gail Burrill (vice chair) (pres., NCTM, secondary teacher and associate researcher, U. Wisconsin), Bettye Forte (program dir. math., Ft. Worth Indepen. Sch. Dist.), Ana Maria Golan (regional dir., Renaissance Proj., Fountain Valley, CA), Joan Grampp (math/science resource teacher, Milwaukee USI), Ann Kahn (past pres., National PTA), Edward Kifer (prof. and chair, Educ. Policy Studies and Evaluation, U. Kentucky), Donald Kreider (prof. math. and CS, Dartmouth Coll.), Steven Leinwand (math. consultant, Conn. State DOE), William James Lewis (chair, Dept. Math. and Stat., U. Nebraska), Debra Paulson (math. chair, Dr. Hornedo Middle Sch., El Paso, TX), Marjorie Petit (deputy commissioner, Vermont DOE), Edward Silver (prof., Cognitive Studies and Math. Educ., Learning R&D Center, U. Pittsburgh), Kris Warloe (teacher, Cheldelin Middle Sch., Corvallis, OR), Patricia Wilson (assoc. prof., Math. Educ., U. Georgia). Several support staff from MPR Associates and CCSSO also attended the meetings.

The Committee first met in Washington in late May, and after an intense series of summer meetings and public hearings around the country delivered the final draft of its report on schedule in early September.

The Working Culture of the Committee and Some Sources of Tension

This was not the first time since working in the world of education that I found myself at a table of mostly unfamiliar faces from professional communities with which I had very limited acquaintance. I came with concerns about the mathematical level and quality of the test items and quickly learned that this was but one, albeit very important, aspect of what goes into the construction of a large-scale test. Myriad important considerations first occupied our attention: the number and distribution of item types (machine- or non-machine-scored, multiple choice, gridded, drawn, short- or long-extended response), scoring rubrics, accessibility issues, reporting, scheduling, venue, administration, etc. The initial meetings focused on these technical, logistical, and administrative issues and featured sophisticated discussions by people whose professional lives they occupy.

I was impressed by the professionalism of this work and by this rapid exposure to the technology of assessment that I had not before closely witnessed. At the same time, I was worried that the attention to such concerns might subordinate the issues of mathematical quality of content, which were paramount in the minds of mathematicians like me. Moreover, I suspected that the knowledge of and sensitivity to such content issues were significantly higher among the few mathematicians present than among most of the other Committee members, and so I was concerned that these priorities enjoy a receptive hearing across the boundaries of the professional cultures represented on the Committee.

I was therefore gratified to see a very open, generous, and respectful working ethos develop within this Committee. Everyone became intellectually engaged with all of the work, both critiquing and learning from the work of the other professionals. A number of issues, particularly the position on use of calculators, were vigorously debated. But the debates were thoughtful and nonconfrontational. Moreover, the views expressed by mathematicians on matters of mathematical substance and quality were generally received with respect and generosity.

I have learned that most important agendas in mathematics education require multidisciplinary perspectives and resources, yet we have few good models of deeply probing collaborative work across the different professional cultures involved. The work of the Mathematics Committee provided for me a gratifying example of what can be achieved when such collaboration is effectively mobilized.

The specifications imposed on our work—adherence to the NAEP framework, the 90-minute test length, the large number of machine-scored items, etc.—were a source of frustration to many of us (mathematicians, teachers, educators), who felt that it would be difficult to represent sufficiently ambitious learning expectations under those constraints. Moreover, with due attention to considerations of accessibility, we intended that the VNMT be designed to express growing expectations over time. So, while honoring the specifications, we pushed their envelope toward allowance for more complex and higher-order problem-solving items and toward content coverage that creatively interpreted the content strands that govern NAEP. These efforts were supported by a broad consensus in the Committee.

A couple of (related) issues precipitated some lively debate within the Committee. These had to do with the allowance for use of calculators on the test and with the amount (and nature) of attention given to "basic skills" on the test. Consensual resolutions of both issues were achieved. In the case of basic skills, the inclination favored designing more complex or multistep problems whose solutions demonstrably entail various basic skills, computational and other, rather than directly posing routine computational tasks.

External Input and Public Response

The Mathematics Committee was instructed to solicit broad input from the public, and it aggressively sought to do so. Two full-day public hearings were held, one in Denver in June, the other in San Francisco in August. Moreover, John Dossey disseminated e-mail requests to many interested parties, including a broad solicitation to mathematicians, for suggestions of candidates for exemplar test items. Richard Askey was among several contributors of very valuable resource material, and he strongly urged attention to the practices of other high-performance countries, notably Japan.

At the time of the Denver meeting there was little available documentation to react to at the public hearing, and the public notice was regrettably short. I personally had to miss that meeting because of a family illness. It was at that time that the first cut on the selection of exemplar items was made, the process with which I was most concerned. These were submitted in the first publicly distributed draft, to which people responded at the San Francisco hearing. When I first saw the selections, I was dismayed at the poor quality, in my opinion, of several of them, and I was eager to help correct this.

The San Francisco hearing was, in many ways, a remarkable event. It was an open and dignified occasion, with the Committee very receptive and responsive to a wide range of opinions and concerns expressed there. A substantial and diverse group of persons, many representing important organizations, came to testify. Some were school system administrators expressing their concerns and hopes for the VNMT. Others made passionate appeals for sensitivity to the potential for aggravating inequities in opportunities to learn in this country. Some of them pleaded for multilingual versions of the test.

Notable among those testifying were several distinguished mathematicians. John Ewing, executive director of the AMS, eloquently expressed the concerns in the mathematical community that the test represent substantial and high-standard mathematical expectations and that mathematicians remain engaged with the development and oversight of the test. Richard Schoen and James Milgram from Stanford presented very detailed and forceful but constructive critiques of several of the items in the then current draft report from the Committee. In fact, they focused on many of the same characteristics with which I had been concerned. Their views were gratefully received. When Schoen suggested that he knew many mathematicians who could construct very good test items, John Dossey eagerly invited him to have such materials submitted to the Committee.

I for one paid great heed to the critiques offered by Schoen, Milgram, and others. Their analyses were detailed and carefully reasoned, and they accorded pretty much with my own earlier reactions to the draft items. I made an express effort to address them in the redesign and, in some cases, reselection of the exemplar test items. I strongly urged inclusion of some items from TIMSS and from other sources, such as the Japanese. And the Committee adopted many of these recommendations. A review of the final report from the Committee will show that the test items selected come from a great diversity of sources and that many of them call for substantial computational skills, often embedded in more complex tasks. The point that I wish to emphasize here is not only the substance of the final report but the fact that the process for its generation, with substantial public feedback, in fact worked very effectively. The Committee did not have a preordained agenda, apart from the overall goals of the test described in the introduction above and the framework for its work imposed by the DOE, and it represented with integrity the constructive views transmitted to it by the public.

Politics

California was then and now remains the site of some volatile and highly polarized backlash to educational reform initiatives. Surprisingly little of this was in direct evidence at the San Francisco hearing, apart from the testimony of Madalyn McDaniel, a parent from Atascadero, CA, who was greatly displeased by IMP (the Interactive Mathematics Program) that had been adopted in her son's school. She had joined forces with anti-reform organizations and was even featured in a widely cited *Weekly Standard* article by Lynne Cheney. Her testimony was emotionally charged, and she flatly accused the Committee of being beholden to the support of IMP and other educational agendas, which she described in almost conspiratorial terms. Following her testimony, several Committee members raised their hands to ask what IMP was.

This was an admittedly extreme example of condemnation of the VNMT based on a priori judgments of the Committee members, insinuating either incompetence or misguided philosophies of education or conflict of interest. This was quite interesting to me, for I was an inside observer who had met many of the Committee members for the first time in this work. They represented a wide range of professional knowledge that seemed necessary for our task. And through direct participation and observation, I gained a great respect for the professionalism and constructive collaboration that the Committee achieved. Further, I found it ironic that concerned observers would condemn an activity expressly designed for generous and open feedback from the public, to which some of them contributed, before they had seen the final product of that process.

This protest found one public expression in the form of an open letter to President Clinton composed in August 1997—before the final draft of the Committee report was prepared—by Mike McKeown and Paul Clopton, founders of Mathematically Correct, a San Diego-based organization opposed to many of the current mathematics reform efforts in California. This letter advanced several claims that I know to be groundless, and it asserted many things about the inevitable design of the test before the final report was even written and which were demonstrably at variance with what was finally produced. Attacks were made on the integrity of the Committee members. I found the letter to be irresponsible and destructive in both tone and substance. I was therefore saddened to find among those enlisted to sign it some of the very people who offered criticism and guidance to the Committee and whose suggestions were incorporated in the final version of the report.

Not everything in our report is what I personally might have chosen, but it represents a thoughtful and intelligently negotiated design that had to accommodate many conflicting realities and ambitions in current U.S. education. Having entered this effort with considerable skepticism, I stand now firmly in support of the report that we have produced and, equally importantly, in support of the process of cross-disciplinary collaboration and receptivity to community feedback that produced it.

There remain, I recognize, good reasons for caution and skepticism about the wisdom of having such a national test at all. Can we be fully assured that its design and development will be well aligned with appropriate educational goals? Will it support or undermine the ongoing efforts to improve teaching and learning in classrooms and the production of high-quality curricular materials? Will such a test further aggravate the unequal opportunities to learn that characterize U.S. education? The desirability of the test is contingent on its effect on all of these other issues, and that is ultimately the context in which it should be judged.