

To Be a Member of The Governing Board

Diane Ravitch

March 2009

Paper Commissioned for the 20th Anniversary of the
National Assessment Governing Board
1988–2008

Diane Ravitch is Research Professor of Education at New York University and a senior fellow at both the Brookings and Hoover Institutions. From 1991 to 1993 she was Assistant U.S. Secretary of Education for Research and Improvement and Counselor to Education Secretary Lamar Alexander. She was appointed to the National Assessment Governing Board in 1997 by Education Secretary Richard Riley and served on the Board until 2004.

For more than 30 years, I have served on the boards of many different organizations, large and small. Given my experience, I thought I understood how a board of directors functions. Typically, the staff makes many decisions between board meetings; the board meets two or four times a year and decides big policy issues. In between meetings, the staff and the executive committee usually have a large amount of discretion. In many instances, the way that the staff and the executive committee frame the policy issues frequently turns the board into a “rubber stamp.” In essence, the board votes on decisions that have already been made, and, in some cases, implemented. Most board members seem to like it that way, because such an arrangement demands a minimal amount of time from their busy lives. Consequently, many organizations are staff-run, rather than board-run, or run by a small group on the executive committee.

This is not the case with the National Assessment Governing Board. I served as a member of the Board from November 1997 to September 2004. The Board is unlike any other board of which I have been a member. At the quarterly meetings, the Board not only discusses and makes major policy decisions but determines the inner workings of national assessments.

The Board’s responsibilities are spelled out in federal law. It is the policymaking board for the National Assessment of Educational Progress (NAEP), which examines samples of American students in a variety of subjects. The Board carefully discusses and decides which subjects will be tested, when tests will be given, how tests will be constructed, what kinds of standards will be used to gauge the results, and how the results will be reported to the public. The Board is also responsible for assuring that the assessments are valid, free of bias, and reliable.

The Board meets four times a year, usually for two full days, and each day is packed with committee meetings, public meetings, presentations, discussions, hearings, and deliberations. Board members are expected to attend every meeting and to work hard to fulfill their public obligations. Although outsiders might expect that the Board would be a partisan battlefield, in light of its sensitive responsibilities, this is not the case. Most extraordinarily, the Board operates in an atmosphere that, in my experience, has been entirely free of partisanship or rancor. While the Board frequently debates which course of action is best, the differing views expressed have never been related to party affiliations but to nonpartisan ideas about what might be most efficacious for the quality of the assessments and what might provide the best information to the public.

The absence of partisanship is written into the design of the Governing Board by law. Board members are appointed by the Secretary of Education. The Board consists of 25 members, plus a nonvoting, ex officio member (the Director of the Institute of Education Sciences). The membership of the Board is prescribed in the federal law. Its clear intent is to produce a Board that consists of a diverse group of people from different backgrounds, who are knowledgeable about education. The Board must consist of two state Governors (or former Governors), each from a different political party; two state legislators, each from a different political party; two chief state school officers; one local school superintendent; one local school board member; three classroom teachers, one

from each of the grade levels typically assessed (grades 4, 8, and 12); one member from business or industry; two curriculum specialists; three testing experts; one member from the nonpublic school sector; two school principals, one from an elementary school, the other from a secondary school; two parents who are not employed by a local, state, or federal education agency; and two representatives of the general public who are not employed by a public education agency.

The term of service for Board members is 4 years, and no member may serve more than two terms. The terms are staggered, so that the Board is regularly refreshed with new faces. In making appointments to the Board, the Secretary of Education is directed by law to assure that the Board “reflects regional, racial, gender, and cultural balance and diversity,” and that the Board “exercises its independent judgment, free from inappropriate influences and special interests.”

These legislative requirements produce a Board composed of earnest people from across the nation who join together as Board members with the sole purpose of improving oversight of the quality of the National Assessments. Because of the diversity and breadth of the Board, there are no cliques or factions. Seldom are there two members from the same state. It is rare that Board members know one another until they join the Board. Members generally do not think of themselves as representatives of an interest group, but as individuals who have been called upon to exercise their best judgment in the interest of American education.

My own service as a member of the Board began unexpectedly. It did not occur through the usual process. In his State of the Union address in 1997, President Bill Clinton proposed voluntary national tests in reading and mathematics in grades 4 and 8. To support that objective, the U.S. Department of Education awarded a contract worth \$50 million to a consortium of testing companies. It was my view that this delicate and controversial proposal should be removed as far as possible from any hint of partisan politics. I wrote an opinion piece in *The Washington Post* urging the Clinton administration to shift control of the national testing program from the U.S. Department of Education to the Board, so that the process of test development would be shielded from any appearance of partisanship. The new national tests, I contended, would have credibility only if they were developed outside the control of any particular administration.

Whether I was right or wrong, I do not know, but a few weeks later, I got a call from Secretary Richard Riley’s office asking me if I would be willing to accept an appointment to the Board. I readily agreed, and in November 1997, I was appointed by Secretary Riley to fill a vacancy. Because my appointment did not occur at the regular time, I was not sworn in by the Secretary at a meeting of the Board. I was informed that I should be sworn in by a notary public before the next meeting. I took the paper with the oath of office to Court Street Office Supply in Brooklyn, near my home, where I knew there was a notary. I asked him if he would administer the oath of office to me for a federal position. A tall Hasidic man with a beard and a skullcap, he did his duty efficiently. I pulled out two one-dollar bills, the usual fee for having a document

notarized. He said, “No, no, please don’t pay me. It was an honor.” I was very touched by what was surely a simple act of patriotism at my neighborhood stationery store.

Soon after I joined the Board, President Clinton announced that he was transferring development of the Voluntary National Test to it. I got my wish, and the Board got the biggest “hot potato” in American education. For the duration of the Clinton administration, the Board worked through the complex issues involved in creating national tests.

The Board, I learned, has five committees: In addition to the Executive Committee, there is a Committee on Standards, Design and Methodology (the testing experts are always on this committee, along with other Board members); a Reporting and Dissemination Committee; a Nominations Committee; and an Assessment Development Committee. I was assigned to the Assessment Development Committee, the one that was responsible for the actual content of assessment frameworks and test items.

For the next two years, the Board wrestled with the difficult issues associated with developing national tests. The problems did not inhere in the subjects to be tested. NAEP was already assessing students in the subjects of reading and math in grades 4 and 8, so it was not difficult to plan the expansion of the number of items to be tested. That did not involve breaking new ground. No, the real challenge with which the Board wrestled was how to define the term “voluntary” in the President’s call for “voluntary national tests.” For whom were the tests to be voluntary? For states? For school districts? For individual schools? For parents? For students? Would the validity of national data be compromised if some states or districts or schools or students decided that they did not want to be tested? Supposing a significant number of parents did not want their children to be tested; would their removal make it impossible to assign a score to their school? Or suppose a number of schools or districts chose not to be included? How would that affect the standing of their state? We discussed these issues at length. Nothing was decided in advance of our meetings.

We wrestled candidly with the questions of how to apply the slippery concept of “voluntary” and how to avoid the politically charged concept of “mandatory” that might be attached to national testing. Everyone had views, views were subject to change, and everyone was engaged in trying to decide what would be the wisest course of action.

One of President Clinton’s closest advisers took me aside and told me that he had not wanted the Board to take over national testing, because the Board “would just talk it to death.” In truth, the Board needed to talk and keep talking because there were so many unresolved issues. And the fact was that the political consensus needed for decisive action was not available. It was clear over time that Congress was very nervous about national testing. And eventually, even as the Board and the test publishers moved forward to develop the tests, events overtook the process. Congress refused to authorize field testing of the national tests, and the project died. All was not lost, however, as the many new reading and math items that had been created were added to NAEP’s ever-needy bank for future use.

All was not lost for me, personally, as I learned about the existence of bias and sensitivity review during the Board's meetings with the representative of the test publishers. Our committee was given a 30-page list of topics and words that could not be used on a standardized test. That list eventually prompted me to write a book about censorship in testing and textbooks called *The Language Police*. The book began with a summary of fourth grade reading stories that were removed from NAEP because they were allegedly biased or insensitive. I asked for and obtained permission from the Department of Education's ethics staff to use the materials. Had I never been a member of the Board, that book would never have been written!

Over the years in which I was a member of the Board, the Assessment Development Committee made many consequential decisions. All of the committee members were fully involved in every decision, which were almost always reached by consensus. It was rare indeed to have a divided vote, as the practice was to keep discussing until everyone felt comfortable with the ultimate decision. In the early years of my service, the committee was led by the late Marilyn Whirry, a fabulous 12th grade English teacher from California. During her time on the Board, she was selected as National Teacher of the Year, and she made a great contribution to our deliberations. Marilyn cared passionately about language, accuracy, and literary quality. After her retirement, the other members of the committee tried to carry on her legacy. Our committee encouraged the company that prepared the tests to seek out passages of high literary quality, and we often rejected passages for reading assessments because, even though they may have been technically sound, they were boring, trivial, and not worth reading. When someone on the outside questioned why the reading assessment included poetry, we reaffirmed our commitment to the inclusion of poetry by outstanding poets.

One of the most interesting assignments of the Assessment Development Committee was to read and review the items on the assessments. We read them as a committee, watching not only for the quality of the passages, but the accuracy and coherence of the questions and answers. We were familiar with the complaint that standardized test questions sometimes have more than one correct answer, and that in some cases, really thoughtful or imaginative students may choose an answer that the test developers consider wrong. We tried to weed out all such ambiguities.

When we read the questions for the United States history assessment, we watched to be sure that no question was politically biased, that it did not favor one party or another, and that it did not propagandize for one side of an issue over another. Some of us were insistent that history questions should actually test whether students knew history, not just test whether they were good readers or whether they could deduce an answer from the evidence presented in the question.

When we reviewed questions for the mathematics assessment, I gained a great sense of humility. I felt fairly comfortable with the fourth grade questions and congratulated myself when I figured out the answers correctly. But on the eighth grade math questions, I could barely register above the "basic" level. At first, I attributed my

ignorance to forgetting the math that I had learned many decades earlier. After a while, however, I realized that the math children learn today is far more demanding than what I studied in public school in Texas in the 1950s.

As is well known, NAEP tests include more than multiple-choice questions. There also are questions which the student answers in a few words of his or her own, and there are questions that require an “extended constructed response,” where students write several sentences. There are rubrics for grading these open-ended responses, and our committee reviewed the rubrics for their quality and coherence. I recall that at one meeting, we were shown examples of written responses that were supposedly excellent, even though some contained errors of spelling and grammar. Led by crack teacher Marilyn Whirry, our committee told the testmakers that such errors should not be overlooked by the graders and that error-ridden responses should not be presented as exemplars.

To demonstrate the tasks that are encountered by Board members, I would like to refer to some of the high points of my own service. While I was a member, the Assessment Development Committee and the Board as a whole revised or commissioned national frameworks for assessing reading, mathematics, U.S. history, and foreign language. Board members were deeply involved in the details of each decision about what would be assessed, the objectives and the specifications of each assessment. In each case, the decisions made by the Board, acting collectively, shaped the nature and content of the assessment.

In 2000, the Board was called upon to decide whether to eliminate the Long Term Trend NAEP (LTT). Many people are not aware that there are two different versions of the NAEP. One is the “main NAEP,” which is periodically changed as new frameworks are adopted to reflect current teaching and testing. The other is the LTT, which consists of the same test questions in reading, mathematics, and science that were first offered in the 1970s. The tests of the main NAEP are administered periodically. Before 2002, the frequency of the tests depended mainly on Congressional funding. Since 2002, the No Child Left Behind law has specified that students in grades 4 and 8 must be tested in mathematics and reading every other year. By contrast, the LTT is given to samples of students every 4 years.

Some critics considered the LTT obsolete and unnecessary. After much discussion, and a careful review of the items on the LTT by the Assessment Development Committee, the Board decided to retain the LTT in reading and mathematics because it acts as an outside check on the main NAEP. However, the Board agreed to drop the LTT in science because the items were indeed obsolete in light of scientific and technological advances over the decades since the original items were written. Some items were removed from the reading test (for example, a question referring to “S&H green stamps”) on grounds of obsolescence. The writing LTT was also dropped because it had too few questions on which to base a reliable trend.

Before the enactment of No Child Left Behind in 2002, one of the big issues facing the Board was the problem of persuading states to participate in the National Assessment. Usually, about 35 to 40 states voluntarily agreed to have their students assessed, but there was always concern that cost and testing burden might diminish participation. However, after the law passed, NAEP was in a very different position. The Board no longer had to persuade states to participate, because the law required every state to permit NAEP to be administered to students on a regular basis. The intent of Congress was to use NAEP as an external gauge by which to check on the states' own claims about their progress. Of course, there are no "stakes" for states or districts or students based on NAEP scores, but they are now viewed by the public as an important independent measure of performance. (The psychometricians on the Board were adamantly opposed to attaching any direct consequences for students or states based on NAEP scores, for fear that such stakes might corrupt the value of NAEP.)

Another important decision during my tenure was the inclusion of urban districts in NAEP. This came about in response to a request by the Council of the Great City Schools, which wanted NAEP to assess big cities on a voluntary basis. The Council for the Great City Schools made its case to Congress and won authorization for the Trial Urban District Assessment (TUDA). The first TUDA in reading was conducted in 2002; six big-city districts participated. In 2003, 2005, and 2007, both reading and math were assessed; 10 urban districts participated in 2003, and 11 in 2005 and 2007. With additional funding, the program grew to 18 districts in 2009. Over the past several years, the participating districts have received valuable information about the performance of different groups of students and about the gaps among groups, as well as about their progress over time.

Virtually every meeting during my tenure involved important issues that the Board was expected to resolve. Each of the committees weighed competing judgments, listened to outside experts, reviewed criticisms, and considered alternatives. Interested members of the public and the press witnessed Board discussions and deliberations.

My years as a Board member taught me a great deal. I saw a citizen Board function responsibly and energetically. I saw the internal operation of a consensus process that produced a genuine enthusiasm for high standards and educational excellence. I was especially impressed by the quality of the professional staff who worked for the national assessment, regardless of which political party was in office. Their professionalism, complete lack of partisanship, and steadfast commitment to the goals of the organization kept us on a steady course. From what I observed, the Board and its staff developed a culture of professionalism that repelled any attempts to use it to advance anyone's political agenda.

If there should ever be a time when the United States again decides that national standards and national assessments are good ideas, the model created by the Board will stand forth as proof that such things can be accomplished without falling prey to political zealots, religious controversies, or the schemes of those who would use testing to advance their own narrow agenda.