

DRAFT MEMO ON EDUCATION INITIATIVES:

Elementary and Secondary Education

The ongoing reauthorization of the Elementary and Secondary Education Act coupled with the need to provide increases to several elementary and secondary base programs has led us to develop initiatives that complement these efforts while supporting innovation, high standards, and efforts to fix failing schools.

Significant Existing Programs

Title I – We recommend increasing funding for the Title I accountability set-aside to \$250 million for FY01 and increasing the funding for the basic Title I grants by at least an equal amount.

After-School – We are analyzing options and developing a proposal that will make after-school programs universal for students in schools identified as failing under Title I. We also recommend allowing 30 percent of funds for after-school and other extended learning programs to be granted directly to non-profit community-based organizations (CBOs)—including faith-based organizations. These programs would be run in partnership with public schools at the school site. This change would allow local organizations that have proven track records in after-school and can offer innovative educational strategies and often better connections to parents and the broader community to get involved the program. At present, no funds can be granted to CBOs, although the ESEA reauthorization proposal would allow up to 10% of funds in a given year to be awarded to nonprofit CBOs.

Class Size – We recommend a significant increase to stay on track toward hiring 100,000 teachers. The program is now funded at \$1.3 billion and we recommend increasing funding to at least [Eric: check with Bruce and get his latest thinking here, this has been a moving target]

Charter Schools – With an increase of approximately \$120 million the charter schools program will support the President's goal of 3000 charter schools. We recommend funding the program at this level.

Elementary and Secondary Education Act Proposals

The President's ESEA reauthorization proposal contains items that we recommend including in the budget for FY01.

Teaching to High Standards: The President's ESEA reauthorization proposal would consolidate Goals 2000, the Eisenhower Professional Development Program, and Title VI into a standards-based-reform grant program. We recommend that the FY01 budget request reflect this proposal. At their FY00 enacted levels the Eisenhower program and Goals 2000 are a combined \$ 830 million. Although the Administration does not

recommend funding the Title VI program in its budget request, the Teaching to High Standards initiative should reflect an inclusion of current Title VI funding (\$380 million FY00) and be funded at a minimum of \$1.21 billion. In order to reach this level while funding our new initiatives, we recommend including five earmarks in the Teaching to High Standards funding: Teacher Academies (OMB proposal for \$100 million), Troops to Teachers – Transition to Teaching (\$50 million), Pay for Performance Demonstration Program (\$50 million), Dual Accredited Teachers (\$30 million), and School Leadership (\$40). *These earmarks are described in more detail in this memo.*

Public School Choice OPTIONS Proposal –The OPTIONS initiative would create a grant program that would fund the demonstration, development, implementation, evaluation and dissemination of information of innovative public school choice projects. Funded projects would focus on areas such as new choice options for students within districts, partnering choice schools with other schools to disseminate information and stimulate improvement, and overcome barriers to choice. The \$30 million request covers \$28 million for approximately 40 new grants \$1.3 million for evaluation, and \$200,000 for peer review.

Troops to Teachers – The Troops to Teachers program was created in 1994 to help improve public school education by injecting the talent, skills, and experience of military service members and other federal civilian personnel into high-poverty schools. Recent studies have found that a significant number of mid-career professionals who possess strong subject matter skills are interested in beginning a teaching career-this program taps this valuable resource. The new proposal will continue the recruiting, preparing and supporting retired military personnel as teachers in high-need subject areas and school districts. It will also support similar programs for other mid-career professionals by awarding grants to public agencies, institutions of higher education, and nonprofit organizations to recruit, prepare, and support career-changing professionals from diverse fields whose knowledge and experience could help them become successful teachers. \$50 million for FY01.

New Initiatives

Fund for Future Schools – During our discussions with educators and education leaders this fall, a common theme was the lack of funding available for large-scale innovation. Our proposal would establish an independent and bipartisan fund to support innovative regional, state, and local ideas that require up-front infusions of funding. As opposed to school-based only reforms these resources would support structural changes at the school district, state, and/or regional level. For example, a recent Education Commission of the States called for changes in school governance, especially in large school districts. The report recommended that school districts either move toward complete site-based management or a portfolio approach where different schools are accountable to the school board for performance but operate independently essentially as charter schools. Hugh Price, President of the Urban League recently called for a similar change in an editorial in Education Week and a few school districts are moving in this direction. For example Seattle has implemented site-based management and Cincinnati is slowly

moving toward the portfolio model. However, changes of this nature require both fiscal and political will. By providing funding we would help school districts that are struggling to maintain services with current funding adopt changes that would otherwise be out of their reach and also help generate political will by making funds available for changes of this nature. Of course, this funding could also support school based or other changes that the board deemed worthy of support. A local matching requirement would encourage local support and investment. Prime Minister Blair has started a similar fund in England and British officials report that in its second year their Education Action Zone initiative is showing positive results however they believe that the independence of the fund is a key characteristic. \$50 million would be distributed by competitive grants in the first year of this program.

Performance-Based Pay and Peer Review – As part of the 1999 National Education Summit in Palisades, New York, the governors and assembled business leaders pledged to sponsor “pay for performance” initiatives for teacher salary in 10 locations as a strategy to make teacher pay more competitive with other professions. Denver Colorado had adopted such a measure just prior to the summit. Despite considerable attention to the issue of teacher quality, teacher pay still remains far below other professions with the national average starting salary only approximately \$25,000, average teacher salary overall only approximately \$38,000, and the average maximum teacher salary in any state only about \$51,000. In addition, the Vice President and others have called for developing more equitable and expeditious ways to reward good teachers and get low-performing teachers out of the classroom. Specifically, in his May 16th speech at Graceland College, the Vice President called for teacher evaluations by experienced peers and administrators every 5 years to ensure, “faster but fair ways to identify, improve, and when necessary remove low performing teachers. California in addition to several large school districts have programs similar to what the Vice President proposed. Our proposal would fund 10 pilot pay for performance plans similar to what Denver has developed. While details would be left up to the local school districts and their teachers, the pilots could support performance awards that are either individual or school based. We would also provide grants for 10 pilots of “peer review” programs where teachers are evaluated by their peers every 5 years and low-performing teachers are given support to improve or ultimately removed from the system in a fair but fast manner. 10 pilots for each strategy could be funded for \$50 million in the first year.

Rewarding Leaders and High Performance –The federal government’s role in education has evolved to one that expects all students to achieve, holds schools accountable for performance, supports improvement, refuses to tolerate persistent failure, and rewards success. However, we have not yet put into place a viable rewards program. This proposal calls for a Rewards Fund for Leaders and High Performance that would award funding to states that met federal performance criteria and would support teacher performance pay demonstration projects (*described above*). From FY2001-2003, states would receive awards for developing accountability systems, including high school exit exams, meeting teacher quality requirements, and developing and disseminating report cards ahead of the timetable the President is calling for in his ESEA proposal. After

2003, the Reward Fund would move to a system based on actual student performance measures.

Initially, this fund would also provide seed money to states to help them develop school report cards including the guidelines that the President specified in his Education Accountability Act. This would help undermine Republican criticism of our ESEA requirements as unfunded mandates to the states. Beginning with \$125 million in FY01 and including our peer review and pay for performance proposals, we would be on a track to spend \$1 billion rewarding performance over five years.

Leadership for 21st Century Schools –Without quality school leaders, school reform is destined to fail. To address this, our initiative would fund independent School Leadership Centers. The Centers would focus on leadership development opportunities for existing schools leaders (principals, superintendents, etc) in areas such as effective management, school design, technology, and district governance. In addition, recruiting and training nontraditional candidates for school and district leadership would also be part of the Center's mission. Centers would be run by independent nonprofits or public-private partnerships on a state or regional basis and they would be required to partner with leadership from the public and private sector. Just as Presidential leadership has increased the national focus on improving teacher quality, federal leadership is needed to push states to make greater investments in developing quality school leaders prepared for accountability, the devolution of decision making, and the focus on school-based reform. At present, there is no federal program focused solely on school leadership and only a handful of programs allow funds to be used for it. There is broad political support for a focus on improving school leadership, as evidenced by the inclusion of a smaller provision in the House-passed ESEA bill. We recommend \$40 million in competitive grants to fund approximately 20 state-level or regional centers.

Schools without Walls: Ensuring Access through Distance Technology –This initiative aims to ensure challenging coursework for all students by providing access to high-quality on-line academic courses to students in under-served areas. While this Administration has shaped a federal education role that holds all students to high standards and expects they can achieve, the reality of school finance in America prevents all students from having access to challenging educational opportunities. Small and poor schools often lack the resources and teachers needed to offer challenging courses, including Advanced Placement courses at the high school level. Today, less than 60% of US high schools offer AP courses to their students. A lack of qualified teachers in these areas, also mean that students don't have access to challenging math and science courses or classes in English as Second Language.

This proposal complements existing federal efforts in technology to fund internet connections, equipment acquisition and teacher training, by funding a competitive grant program to districts that will allow schools to purchase on-line curriculum, train teachers to use on-line curriculum, or purchase course materials. Although many states are developing virtual school capacities to offer courses to students throughout their state, the quality of on-line courses has yet to match the capacity of technology. In order to ensure

high-quality content, the proposal also calls for a partnership with leading course software developers like APEX (run by Microsoft co-founder, Paul Allen), which would subsidize the cost of high quality Web-based course development in return for cut-rate prices for high poverty school districts. The proposal also calls for increases in the Technology Literacy Challenge Fund, which allows computer equipment purchases, and the Technology Teacher Training Program, and continued commitment to E-Rate. The initiative should win broad support as Republicans and Democrats have supported AP fee assistance and Senate Democrats and Republicans have consistently supported Star Schools, a distance learning initiative. Total funding for the initiative would be \$225 million, including \$50 million for grants, \$25 million for course development, and \$125 million in increases in existing programs.

The Backyard Teacher Recruitment Initiative –The problems of teacher recruitment and out-of-field teaching are issues that loom large for public education. For schools and districts in rural areas and inner cities across the country they are especially difficult. Despite federal and state efforts, these areas continue to have great difficulty finding qualified candidates in areas such as math, science and special education. This proposal aims to help high-need poor school districts with large percentages of out-of-field teaching by paying for high-quality teachers in non-shortage field to be trained and certified in shortage areas. This program would focus on teachers who have already made the decision to teach in high-need areas. Recognizing that they already have teaching expertise, the program would provide bonuses and tuition assistance for content training in a high-need subject area for 10,000 teachers across the country. In return, they would commit to teach for at least three years in their district.

While this can certainly not be considered a fix for teacher recruitment, it can provide some immediate relief for districts with chronic teacher shortages in critical subject areas. And by focusing on a population already in schools, it adds a needed dimension to a federal recruitment effort that focuses largely on efforts to bring new people into the profession. Teacher recruitment is a key issue in states across the nation and could attract bipartisan support. The initiative would cost \$30 million.

Second-Chance Schools – This proposal would create new second-chance schools through an infusion of additional funds to the Department of Education's Charter Schools program earmarked for this purpose. The initiative would build on the existing charter schools program, maintaining priority for states with well-designed charter school laws that hold schools accountable for results. Second-chance schools would be intended to serve students who have been removed from their previous schools because of discipline problems. This funding would support schools that provide alternative settings for disruptive students while allowing them to continue academic work toward state academic standards. This initiative would cost \$50 million in the first year.

Small Schools – This proposal would build on an initiative in the FY00 appropriations bill supported by Congressman Obey to establish smaller learning communities in America's high schools. The program would offer competitive grants to school districts

that submit a plan to create smaller learning environments for students by opening smaller schools (including charter schools), or by breaking up larger schools using strategies such as creating schools-within-schools, establishing career academies, or restructuring the school day. Allowable uses of federal funds could include planning and implementation costs, professional development, team building, minor facility renovation, additional planning time, legal and accounting consulting, supplies and other relevant costs. These funds could also be used to seed small charter high schools in the district. The Obey initiative is funded at \$45 million in the FY 2000 appropriations bill and we recommend expanding funding to \$200 million for FY01.

SAT and ACT Test Preparation – Historically poor and minority students have not scored well on college entrance exams such as the SAT and ACT. At a time when these scores are a very influential factor in college admissions and when backlash against affirmative action is decreasing minority enrollment in some top institutions, it is appropriate and necessary to offer low income students the same college test prep courses that boost the scores of their more privileged classmates. This program would be modeled after California's College Preparation Partnership Program, which the NAACP has recently recommended replicating across the country. The program would support partnerships of high schools, proven providers of college test prep courses (such as Kaplan and Princeton Review), and community based organizations that would offer high-need students college test preparation and other services related to college admissions. Funding for these partnerships would be distributed by competitive grant. \$30 million would serve approximately 50,000 students.

Teacher Recruitment -- This program would provide competitive grants to high-poverty school districts to develop programs to "grow their own" teachers as a means of addressing their shortage of qualified teachers. Programs supported by this grant would begin to cultivate and recruit students as they move through high school. Upon high school graduation, participants would attend universities to gain expertise and teacher certification in a high-need field. After college, the students would return to high-poverty districts to guaranteed teaching positions. Throughout the program, students would receive exposure and training experiences at summer camps and as teaching assistants. "Home grown" teachers would also receive salaries for work during high school and college, scholarships for tuition, recruitment signing bonuses, and high quality professional development. This program could be started in 100 school districts with \$100 million in FY 2001.



Bruce N. Reed
12/02/99 01:12:55 PM

Record Type: Record

To: John B. Buxton/OPD/EOP@EOP, Bethany Little/OPD/EOP@EOP

cc:

Subject: we need talking pts on goals panel report. see ap story

States Fall Short of Teaching Goals

By Anjetta McQueen
AP Education Writer
Thursday, Dec. 2, 1999; 11:32 a.m. EST

WASHINGTON — States have made some progress in improving their educational systems, but overall the nation will not meet by 2000 the goals in reading, math, graduation rates and other measures that a national commission set a decade ago, a panel reported today.

The good news, however, the National Educational Goals panel said, is that more U.S. preschool children are starting school in better health and with better reading skills, more middle school students are also reading better, and everyone is making strides in math.

More older students also are going to college and more are earning degrees in math and science, said the panel, charged with measuring the eight education goals set over the decade by Presidents Bush and Clinton and the nation's governors.

"Our mission is not complete," Kentucky Gov. Paul Patton said today. Patton, who heads the bipartisan panel of governors, state lawmakers and members of Congress, added, "But it is clear our goals have moved the U.S. in the right direction."

But progress toward goals for improved teacher preparation, school safety and parent participation has stalled or worsened since the goals were established, according to information gleaned from the states for the panel's last report of the decade.

In 1990, after Bush and the nation's governors and business leaders held an education summit, six goals were adopted to target preschool preparation, high school graduation, student achievement in basic subjects, achievement in math and science compared with other nations, school safety, and adult literacy and job skills.

Two goals for teacher training and parent involvement were added in 1994 under the Clinton administration.

Serious challenges remain on all eight goals, panel members acknowledged, but many educators said that didn't mean the mission was a failure.

"I am not deeply troubled that America did not conquer all these goals during the 10-year period," said Sharon Lynn Keegan, a Yale University early childhood learning researcher. "The national education goals panel has created hope, dreams and opportunities for millions of children in this nation."

Rep. Bill Goodling, R-Pa., a goals panel member and chairman of the House Education and the Workforce Committee, said the goals have prompted states to improve individually.

"They took the goals as a real challenge," Goodling said. "The whole idea of setting goals and getting the American people to think about education has helped a lot of enlightened governors win support from their state assemblies in order to move reform programs."

The panel's report found that Colorado, Connecticut, Kentucky, North Carolina, South Carolina and Texas have shown the biggest improvement in meeting the goals during the 1990s. Maine, Connecticut and North Dakota are the states that consistently do the best, as measured by how closely they meet the goals. Connecticut, the District of Columbia and North Carolina were cited for strides made in numerous categories.

Despite the praise for states, some observers have asked whether policy-makers set unrealistic goals or whether states failed an easy task.

Chester Finn, a former assistant secretary of education in the Reagan administration, argues for fundamental change in the way schools are run, including giving more freedom from regulations, yet more responsibility to graduate successful students.

"After all the celebrations, millions of American kids will remain badly educated," he said. "How characteristic of American education that we try to solve problems by hurling new structures and programs at them. We end up keeping the structures but not solving the problems."

The full report is available on the panel's Web site at www.negp.gov.

© Copyright 1999 The Associated Press

Back to the top

DRAFT

November 18, 1999

MEMORANDUM TO BRUCE REED

FROM: Bethany Little

SUBJECT: Reforming the American High School

Let { "Fourth, we should commit ourselves to fundamentally changing the American high school. I envision a new American high school -- with smaller classes, smaller schools, and principals with the power to lead. We've done some things wrong in education, and here's one of them: herding all students in a 25-square-mile area into overcrowded, factory-style high schools. When teachers and principals must practice crowd control, it becomes impossible to spot the early warning signs of violence, depression, or academic failure -- and it becomes even harder to do something about it. We should provide incentives to create smaller high schools. And for those that have already gotten too big, let's break them down by creating smaller "schools within schools. Classes are also way too big. Teachers need fewer students in each classroom so they can pay close attention to each one. We should begin with a national commitment to reduce class size to an average of 18 students in the early grades -- and then aim at average class sizes of twenty students or less across all grades. This will be a major commitment, but it will be worth it, because it will give all our students the individual attention they need to succeed. And we should empower principals, advised by teams of teachers, to hire their own staffs, regardless of seniority, and manage their own budgets. This would enable schools to choose teachers that fit the mission and approach of the school, including teachers who have entered the profession through high-quality alternative certification pathways. It is time to focus on results, not just process -- on measured competency, not check marks from gatekeepers."

— Vice President Al Gore

BACKGROUND

Incidents of school violence, like the recent tragedy in Littleton, Colorado, have caused serious alarm among parents and students who are unsure what has caused such tremendous alienation and aggression in our teenagers. Although there have been calls for tougher gun control laws and stricter discipline policies, more and more educators are pointing to a systemic problem -- the model of the American high school. Experts have noted that the American high school is one of the most enduring and unchanging institutions in our society, and some have gone as far as calling the institution "obsolete". As Secretary Riley recently observed, "Many of our high schools do a very fine job of preparing students for college. At the other extreme, some high schools are little more than a way station for young people who already know that life is unfair". One indicator of our failure to adjust high schools to the changing needs of the population is the alarmingly high Hispanic dropout rate (30 percent for students born in the United States and 44 percent for those born outside the country). Hispanics are the fastest growing segment of our society, yet our high schools are clearly not engaging or serving these students. Finally, we are failing to prepare many students for the academic challenge of higher education and lifelong learning. While record numbers of high school graduates are going on

*Themes =
Smaller schools
and classes
- herding
- Revolving door
- Be the school for some schools
- School facilities
- Asking more from
students
- Exit plan
- 2010
- 139 in*

to college, almost half of entering college freshmen drop out by the end of their second year. Close to thirty percent of students entering a four-year college have to take remedial courses.

Although experts cite a variety of factors that contribute to the decline of our high schools, including academic mediocrity, poor teaching, insufficient counseling services and waning parental involvement, most cite the same remedy for these ills – create smaller schools with smaller classes. Since the end of World War II, the number of schools nationwide has declined seventy percent, while average enrollment has grown fivefold. More than 25 percent of American high schools enroll more than 1,000 students, and enrollments of 2,000 and 3,000 are common. The conventional wisdom in building these enormous schools was that they would offer students more academic and extracurricular opportunities. However, more and more large schools are being built because of the need to educate rapidly growing numbers of high school students. There are currently 13.5 million students attending secondary school in America. In the next decade, our high schools will enroll an additional 1.3 million students. Although larger schools may be more economically efficient (and evidence of this is mixed), they are frequently a detriment to student achievement, especially for our most disadvantaged students. One study found that large schools are actually more expensive to operate, and the additional layers of bureaucracy resulted in less flexibility and innovation in meeting students' needs. Contrary to prevailing wisdom, research shows that small schools can offer a strong core curriculum and, in most cases, a level of academically advanced courses comparable to large schools. Research also shows that smaller schools have better attendance records, lower dropout rates and fewer discipline problems. A 1996 study found that students in smaller schools were more likely to be involved in extracurricular activities and to hold positions of responsibility. Finally, research backs up the conventional wisdom that smaller schools are safer and more productive because students feel less alienated, more nurtured and more connected to caring adults, while teachers feel that they have more opportunity to get to know and support their students.

Although research suggests that large schools are more detrimental for disadvantaged students, the recent rash of violence in suburban schools is making the option of smaller schools more attractive to all parents. A 1997 study of charter schools showed that the most common reason parents chose a charter school over a traditional public school was the smaller size of the charter school. Fifty-three percent of parents surveyed cited the schools' small size, ranking that factor ahead of higher standards, greater educational philosophy, greater parental involvement and better teachers.

There is no agreement on what a "small school" looks like, but there is a general consensus that a high school should have no more than 1,000 students. The University of Illinois *Small Schools Workshop* describes "good" small schools as having a curriculum integrated around a coherent focus and philosophy of education, teachers that collaborate and discuss the needs of their students, a sense of leadership and investment in the school, and the involvement of families and community. Models of small schools include freestanding small schools, autonomous schools-within-schools, and scatterplexes, with two or more small schools at different sites that share a principal.

POLICY OPTIONS

These policies are not all intended to be mutually exclusive, but instead should offer some complementary options. An ideal high school reform proposal would probably encompass some combination of these policies as well as relevant rhetoric around important things outside the scope of

the federal government (e.g. a curriculum that actively engages students). Almost all of these proposals could target funds to high-need districts.

Option 1: Create a new focus on smaller high schools through an infusion of additional funds to the Department of Education's Charter Schools program to create smaller high schools. The initiative would build on the existing program, maintaining its priority for funding states with well-designed charter school laws that hold schools accountable for results. Other characteristics of a well-designed charter school law as defined by the Department include building-level control over budget and expenditures, and either a chartering entity other than the local school board, or an appeals process for charters rejected by the local school board. The small schools piece could be achieved either by creating an additional priority and investment for small schools, or by earmarking a portion of funds to be used only for the development of small high schools. Governance and staffing for the schools would remain subject to state law. There is some consensus that creating a high school is generally more expensive than an elementary school. Currently the average charter school grant is about \$100,000 a year for three years. An effective program would probably have to offer more money to ensure that schools are well designed and implemented, and to give states an incentive to break up large schools or make an effort to build new schools as smaller schools. An effective small high schools program could expect to make grants averaging \$250,000 a year for three years. This model includes an incentive of an additional 25 percent of the grant received by the school to be sent to the LEA for "transition costs" including transportation, facilitation and site management. Allowable uses of school funds include professional development, team building, minimal facility renovation (see construction proposal), additional planning time, extended learning time, legal and accounting consulting, supplies and other start-up costs.

Advantages -- Would create high schools that are not only smaller, but have the other positive characteristics of charter schools including a common mission; teacher, parent and student buy-in; and increased autonomy. Would also expand public school choice, and may be a vehicle to build on the Vice President's proposal for second-chance schools. Would not require an additional program, and would probably only require minimal changes in the current statutes.

Disadvantages – Not all states have charter laws, which would limit participation (although it might encourage more states to develop them). Also, the teachers unions are generally unhappy with the concept of charter schools, and might resist an effort to expand them

Option 2: Create a program run out of the Department of Education to give grants to create small schools. This program could include new schools, or the conversion of existing schools into small schools through schools-within-schools or multiplexes. It could include funds for facility development, professional development, team building, additional planning days, legal and accounting consulting, supplies or other start-up costs.

Advantages – A program designed from scratch could be engineered to meet our specific objectives.

Disadvantages – Must get legislation creating a new program, and would feed the perception that we create a program to solve every problem rather than fixing the programs we have.

Option 3: Implement a direct or guaranteed loan program for construction of small schools.

This model would establish a loan program that would use appropriated discretionary funds to “subsidize” loans from either Treasury (direct loan program) or private lenders (guaranteed loan program). The assumption OMB uses is approximately 25 cents per dollar loaned must be deposited in the “risk fund” by the Department of Education.

Yes
Advantages – This program would allow us to set the interest rate based on different factors. For instance, we could offer a lower rate to needier districts or to districts building small schools. It would also cost less over 10 years, with the bond proposal costing about \$10 billion over 10 years to leverage \$24 billion of funds, and the direct/guaranteed loan costing about \$6 billion over ten years to leverage the same amount. This proposal would probably be more complementary to local school finance, because it is extremely difficult to pass a bond issue.

Disadvantages – Money must be appropriated from discretionary funds.

No
Option 4: Expand or focus the current School Modernization Bond proposal to encourage smaller schools. Our current school construction program has a set-aside for the 100 neediest school districts, but other than that, the bonding authority goes to states with few strings attached. It would probably be possible to require that states give priority to proposals to build small schools or break up existing large schools into small schools. Our current proposal gives the Department of Education about half of the total bonding authority to allocate to the 100 neediest districts, and to an additional 25 districts at the Secretary’s discretion. One way to focus this would be to direct the Secretary’s discretionary piece to small, or to give some additional bonding authority to the Department of Education to allocate directly in response to applications from districts building small high schools.

Advantages – This spending is on the tax side, and is not discretionary.

Disadvantages – We have yet to pass our school construction proposal. The more strings we attach, the less likely it may be to ever pass.

No
Option 5: Establish a revolving loan fund for school construction and renovation. At one point, a revolving loan fund similar to the regional transportation banks was considered for the Administration’s school construction proposal. Under this model, the federal government appropriates money to set up financial institutions that make loans to districts building small schools. As these loans are repaid, the money is used to make more loans.

Advantage – More complementary to the needs of districts, because it does not require that they pass a bond issue.

Disadvantage – This option does not leverage funds as effectively as either the bond proposal or the direct/guaranteed loan proposal. Requires a heavier infusion of capital in the early years, before the fund really starts “revolving”.

20
Option 6: Create smaller classes in high schools. Expand our investment in the current class size reduction program, ~~mandating~~ that after states have reached their goals in the early grades they use the money to reach a similar goal in their high schools.

yes
Option 7: Incentives for a High School Exit Exam. Give additional funds to hire teachers for smaller high school classes to those states that institute (or have) a high school exit exam. These exams would be developed by the state, and tied to state standards and accountability systems. Alternative incentives for states to implement these exams might be funds to develop and implement the tests, and to cover additional learning time for students struggling to pass the test.

✓
Option 8: High School Readiness Assessments. A high school readiness assessment would be administered in the year before a student's matriculation in high school. The exam would identify students who lack the basic skills to succeed in a challenging high school curriculum, but would not be used to determine whether or not a student was eligible to enter high school. States that agreed to implement such an exam could be offered money to provide summer courses for students lacking basic skills to ensure they are prepared to succeed in high school. Another incentive might be money to provide "curriculum counselors" -- counselors or teachers trained to work with students entering high school to ensure that they take the courses they need to pursue higher education, satisfying jobs and other goals.

Advantage – This would relieve the burden on high school teachers who feel overwhelmed by the enormity of educating students who lack basic skills. It would also ease the transition into the more demanding high school environment, and identify those students who need extra encouragement and support to succeed.

Disadvantage – It could inadvertently contribute to the "labeling" and "tracking" of students.

Being
Option 9: Use technology to help ensure that every student has access to the challenging curriculum and basic skills training that they need to succeed. Many educators and parents complain that students in poor urban schools and small rural schools don't have access to the same high-level coursework as other students, and that this adversely affects their ability to get into good colleges and universities. Software and technology are being developed that would allow disadvantaged students to access the same AP courses, honors courses, and SAT or ACT prep courses that their more advantaged peers enjoy, but research is advancing slowly because of market fragmentation. A useful federal role might be to invest in the development of both the software and the advanced technology for this next generation of distance learning through private or non-profit corporations. In exchange, corporations could offer the products to needy districts at a much lower price. The same principle could be applied to the development of remedial and ESL software, to help students still struggling to master the basics.

OTHER

Other proposals the Vice President has made that would complement efforts to reform the American high school:

SAT Comee money for
Money Minuties.

DRAFT

Keeping Kids In School. As President, Al Gore will challenge every state to raise their compulsory school age to 18 and would provide matching grants to help those states that meet this challenge and adopt aggressive, proven strategies to prevent students from dropping out of school.

Strengthen Existing Efforts To Prepare Students For College. As President, Gore would also propose significant funding to strengthen the "GEAR-Up" and TRIO programs to help young people prepare for success in college.

Helping Disadvantaged High School Graduates Succeed in College. As President, Al Gore will propose new "Opportunity Academies" to provide disadvantaged high school graduates with the extra academic preparation and support needed to succeed in college.

Safety and Discipline. "I propose the creation of second-chance schools -- where kids headed for trouble, and those caught with guns, can receive the strict discipline and intensive services they need. For all schools, there should be a simple policy toward guns: zero tolerance, period. All schools should be gun-free, drug-free, and safe and secure. We should increase our commitment to after-school care this year, so children have a place to learn in those afternoon hours when most juvenile crime, teen pregnancy, alcohol and drug use occur. Then let's get a commitment from schools to add civic and character education courses -- to teach students that the same values they learn in good homes apply in our schools and in our society."

Increasing Parental Involvement. "If we want parents to be more involved, we must help them have the time to do it. We should extend the Family Leave law, so parents can take time off work to attend not only the beginning-of-the-year meeting, but all parent-teacher conferences -- without fear of losing their jobs. Either we're serious about this or we're not."

WHITE HOUSE STAFFING MEMORANDUM

Date: _____ ACTION / CONCURRENCE / COMMENT DUE BY: _____

3 Reports Prepared by the National Assessment
Governing Board

Subject: _____

	ACTION	FYI		ACTION	FYI
VICE PRESIDENT	☐	☐	MOORE	☐	☐
PODESTA	☐	☐	NASH	☐	☐
ECHAVESTE	☐	☐	REED	☒	☐
RICCHETTI	☐	☐	RUFF	☐	☐
LEW	☐	☐	SOSNIK	☐	☐
BERGER	☐	☐	SPERLING	☐	☐
BLUMENTHAL	☐	☐	STEIN	☐	☐
CAHILL	☐	☐	STREETT	☐	☐
FRAMPTON	☐	☐	TRAMONTANO	☐	☐
IBARRA	☐	☐	UCELLI	☐	☐
JOHNSON, B.	☐	☐	VERVEER	☐	☐
JOHNSON, J.	☐	☐	WALDMAN	☐	☐
KLAIN	☐	☐	YELLEN	☐	☐
LANE	☐	☐	_____	☐	☐
LEWIS	☐	☐	_____	☐	☐
LINDSEY	☐	☐	_____	☐	☐
LOCKHART	☐	☐	_____	☐	☐
MARSHALL	☐	☐			

REMARKS: Please return this report to the Staff Secretary's Office after review. If you would like a copy of this report to keep, please contact this office. Thank you.

RESPONSE: _____

ID# 310486

THE WHITE HOUSE
CORRESPONDENCE TRACKING WORKSHEET

'99 JUL 19 PM5:50

INCOMING

DATE RECEIVED: JULY 09, 1999

NAME OF CORRESPONDENT: MARK D. MUSICK

SUBJECT: TRANSMITS THE ENCLOSED 3 REPORTS PREPARED
BY THE NATIONAL ASSESSMENT GOVERNING BOARD

		ACTION		DISPOSITION	
ROUTE TO:		ACT	DATE	TYPE	C COMPLETED
OFFICE/AGENCY	(STAFF NAME)	CODE	YY/MM/DD	RESP	D YY/MM/DD
TIM SAUNDERS		ORG	99/07/09		694,7,19
Sean Maloney	REFERRAL NOTE:	A	99/7/19		1/1
	REFERRAL NOTE:		1/1		1/1
	REFERRAL NOTE:		1/1		1/1
	REFERRAL NOTE:		1/1		1/1
	REFERRAL NOTE:		1/1		1/1

COMMENTS: ENCLOSURE *Not for President's xmittal to the Congress*

ADDITIONAL CORRESPONDENTS: MEDIA:L INDIVIDUAL CODES: _____

MI MAIL USER CODES: (A) _____ (B) _____ (C) _____

*ACTION CODES:	*DISPOSITION	*OUTGOING	*
*	*	*CORRESPONDENCE:	*
*A-APPROPRIATE ACTION	*A-ANSWERED	*TYPE RESP=INITIALS	*
*C-COMMENT/RECOM	*B-NON-SPEC-REFERRAL	*OF SIGNER	*
*D-DRAFT RESPONSE	*C-COMPLETED	*CODE = A	*
*F-FURNISH FACT SHEET	*S-SUSPENDED	*COMPLETED = DATE OF	*
I-INFO COPY/NO ACT NEC		*OUTGOING	*
*R-DIRECT REPLY W/COPY *			*
*S-FOR-SIGNATURE *			*
*X-INTERIM REPLY *			*

REFER QUESTIONS AND ROUTING UPDATES TO CENTRAL REFERENCE
(ROOM 75, OEOB) EXT-2590
KEEP THIS WORKSHEET ATTACHED TO THE ORIGINAL INCOMING
LETTER AT ALL TIMES AND SEND COMPLETED RECORD TO RECORDS
MANAGEMENT.



National Assessment Governing Board

National Assessment of Educational Progress

June 29, 1999

The Honorable William Jefferson Clinton
President
The White House
1600 Pennsylvania Avenue
Washington, DC 20500

Dear President Clinton:

On behalf of the National Assessment Governing Board, I am pleased to transmit the attached three reports prepared by the Board:

- The Voluntary National Test: Purpose, Intended Use, Definition of "Voluntary" and Reporting
- Student Performance Standards on the National Assessment of Educational Progress: Affirmation and Improvements
- National Assessment of Educational Progress: Design 2000-2010

The reports to Congress and the President on the voluntary national test and on student performance standards are required under the FY 1999 Omnibus Budget Act (the Act) and are due by September 30, 1999.

With respect to the first report, the Act directs the Governing Board to "determine and clearly articulate the purpose and intended use" of the proposed voluntary national tests. This charge to the Governing Board includes defining the term "voluntary" and describing the achievement levels and reporting methods to be used in reporting results. The Act also directs the Governing Board to address the assertion in a National Academy of Sciences report that the process for setting the achievement levels used to report results on the National Assessment of Educational Progress is "flawed." The intent of these reports is to inform deliberations on the proposed voluntary national tests.

Although not required by law, the Governing Board is submitting the third report, on the future design of the National Assessment of Educational Progress, in order to inform deliberations on the reauthorization of that program, expected this year or next.

The Governing Board understands that the issues addressed in these reports are complex and some are controversial. We have sought comments from many sources and attempted to be careful and deliberative in preparing these reports. Please contact me or Roy Truby, Executive Director of the Governing Board, if you have any questions about them.

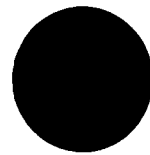
Sincerely,

Mark D. Musick
Chairman

Enclosures

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

310486
Spencer
JUL 9 1999





National Assessment Governing Board

National Assessment of Educational Progress

The Voluntary National Test:

Purpose, Intended Use, Definition of Voluntary and Reporting

Adopted Unanimously by
The National Assessment Governing Board, June 23, 1999

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

The Voluntary National Test: Purpose, Intended Use, Definition of Voluntary and Reporting

Background

Purpose

The purpose of this report is to fulfill one of the requirements of the FY 1999 appropriation act for the Department of Education (the Act). Specifically, with respect to the proposed voluntary national tests in 4th grade reading and 8th grade mathematics, the Act requires the National Assessment Governing Board to

“...determine and clearly articulate the purpose and intended use of any proposed federally sponsored national test. Such report shall also include—(A) a definition of the meaning of the term “voluntary” in regards to the administration of any national test; and (B) a description of the achievement levels and reporting methods to be used in grading any national test.”

This report addresses the four required areas: purpose, intended use, definition of “voluntary,” and reporting. Although the legislation states that the Governing Board shall “determine” these matters, the Governing Board recognizes that this report is advisory to Congress and the President. Any final determination on these matters will be made in legislation enacted by Congress and signed by the President.

The Act contains other provisions related to the voluntary national test. One provision amends the General Education Provisions Act, creating a new section 447, prohibiting pilot testing and field testing of any federally sponsored national test unless specifically authorized in enacted legislation. However, another provision permits the development of test questions for the voluntary national tests, giving the National Assessment Governing Board exclusive authority for such test development.

In order to carry out the congressional assignment to prepare this report, the Governing Board had to envision a situation in which there was authority to conduct voluntary national tests, while recognizing that the Act prohibits such tests at this time. Further, the Governing Board had to envision how national testing could work, given that schools in the United States are governed by states, localities and non-public authorities. The Governing Board attempted to answer the question: If there are to be voluntary national tests, what is a feasible, coherent plan that would be beneficial to parents, students, and teachers? Thus, while not advocating for or against the voluntary national test initiative, the Governing Board interprets the congressional assignment to be to present a sound and logical case for the potential purpose and use of the voluntary national tests.

The Act sets September 30, 1999 as the deadline for submitting this report to Congress and the President. However, to assist Congress and the President in deliberating on the future of the voluntary national test, help promote a timely decision, and avoid a full

year's delay in pilot testing should Congress and the President decide to proceed with the project, the Governing Board is submitting its report in June.

Report Preparation Process

In November 1998, the Governing Board established a special ad hoc committee to assist in drafting the report. The committee was composed of both veteran and new Board members. Chaired by Michael Nettles, the committee included Wilmer Cody, Thomas Fisher, Michael Guerra, Nancy Kopp, Debra Paulson, Diane Ravitch, and John Stevens.

The committee developed a plan for preparing the report, engaging the Governing Board in related policy deliberations, and obtaining public comment. At the March 1999 Board meeting, the committee presented materials that were developed for public comment. These included an explanatory statement; two possible scenarios addressing purpose, use, definition of voluntary, and the methods for reporting; and a set of questions related to the scenarios. The purpose of these materials was to provide a framework for public comment. They did not represent the positions of the Governing Board at the time.

The Governing Board discussed these materials at length, made several changes, and authorized the committee to proceed to obtain public comment. The materials and an invitation to provide written comments and/or oral testimony at four public hearings during March and April were disseminated.

Taking the comments received into account, the committee then prepared a draft report for review at the May 1999 Governing Board meeting. The Governing Board discussed and revised the draft report and authorized the committee to obtain comment on the draft report. The draft report was disseminated by mail, on the Governing Board's web site, and in the Federal Register. A hearing on the draft report was conducted June 12 at the annual Large Scale Assessment Conference with state and district testing experts.

After reviewing the comments from the public and the June 12 hearing, the committee revised the draft report for presentation to the Board at a special meeting on June 23. At the June 23 meeting, the Governing Board discussed the draft and approved a final version for submission to the President and Congress.

Overview

This report is in three sections. The first section is in the form of a story. It is intended to put a "human face" on the details in the section that follows. The second section describes the Governing Board's recommendations on purpose, intended use, definition of "voluntary," and reporting for the proposed voluntary national tests. The third section is a summary with recommendations.

Congress directed the Governing Board to “determine and clearly articulate the purpose and intended use” of the proposed voluntary national tests. In order to do this, the Governing Board had to envision a feasible, coherent approach. The following is intended to indicate how the voluntary national test program, if authorized, might work.

The Voluntary National Test: An Illustration

It is March 19; the year is 2006. Fourth grader Maria Johnson, along with her classmates and many other 4th graders across the nation, will be taking the voluntary national test in reading tomorrow. Eighth graders will be taking the mathematics test.

Maria started kindergarten in September 2001; the first voluntary national test was administered the following March. That year and each year since, a national Sunday supplement magazine devoted an early April article to the test. The test questions were published, along with the answers. For questions that require students to write their own answers, samples of student work from the national tryout of the test the year before were included to illustrate different levels of student performance. These levels of student performance are based on the achievement levels set for the National Assessment of Educational Progress (NAEP). Similar materials were made available following each year’s tests in newspapers, magazines aimed at parents and teachers, on the Internet, and on the Public Broadcasting System. Reading and mathematics achievement levels posters are displayed in pediatrician’s offices across the country. January through March of each year, restaurant chains print sample test questions on placemats and supermarkets print them on grocery bags and food containers.

Maria’s school district decided to volunteer to participate in the national test in 4th grade reading. The school district administration had examined the test framework, specifications, and sample test and determined that they were consistent with the district’s reading program. They knew that the results would belong to the district and the families. The federal government would not report or maintain any of the data resulting from testing nor require the district to report any of the data to the federal government.

Maria’s school provided copies of the Sunday supplement article to each of the families. In the school district, the policy is for all students to participate in testing unless a parent specifically objects. When Maria’s parents finished reading the article, they had a clear picture of what a proficient reader in the fourth grade should know and be able to do. They understood that proficiency would not come overnight, but with many small steps and that each year of school would mark progress toward the goal of reading proficiency. Maria’s parents decided that having a clear goal and following progress toward that goal are good things to do and wanted their child to participate.

Having this initial knowledge, the Johnsons wanted to learn more and did their homework. They attended a school-sponsored seminar on the reading program. They learned what they could do at home to reinforce what Maria was learning in school. The Johnsons obtained a special version of the NAEP framework, written for parents, to deepen their understanding of the material covered by the test. The Johnsons now had a frame of reference for talking with Maria’s teachers in specific terms about the reading program and for monitoring Maria’s progress each year toward 4th grade reading proficiency. Maria, with her parents’ encouragement and teachers’ support, has worked hard in school and at home on her reading assignments and enjoys reading on her own.

With this shared understanding and common language about reading proficiency, the school was helped in its efforts to involve parents. The school had developed its own testing program to track the reading progress of each student each year toward 4th grade reading proficiency. Thus, students who needed extra help were identified early and remediation occurred before it was too late.

The school liked using the achievement levels. They were found to be meaningful in relation to the state's performance standards for reading. They helped keep the school staff focused as they worked day-by-day, making hundreds of decisions about materials, instruction, and curricula to achieve the many incremental steps needed for each student to progress.

Parents and teachers also like the fact that the test booklet is returned. This permits parents and teachers to review with the student all of the test questions and the student's answers. The student gets reinforcement on what was done well. Parents and teachers can see which questions were answered well and which were missed, probe the reasons why with the student, and, from the student's response and other knowledge of the student, explore whether advanced activities, diagnostic testing, or any other intervention should be considered.

Together with the district's on-going assessment program and the state's standards and assessments, the school and parents found that the voluntary national test adds in a unique way to the range of methods for monitoring individual student progress. The teachers and principals found that the achievement levels used to report voluntary national test results were much easier for parents to understand than percentiles, stanines, or mean scores. Also, the voluntary national test provides parents and schools a single basis of comparison for individual student performance across states that is generally not available from classroom developed tests or state-wide assessments. Most of all, parents have a clear and very specific understanding of how their child has performed in comparison to rigorous standards.

Although the test was designed to provide individual results, the school district has decided that it will compile the individual student results that were provided by the voluntary national testing program. The district administrators want to know how the district overall compares with the students in the sample who participated in the national trial run of the test conducted the year before the test is administered.

The district has joined a consortium of similar districts that have agreed among themselves to follow the guidelines for compiling and reporting voluntary national test data developed by the National Assessment Governing Board (NAGB). Following these guidelines ensures that the data analyses are done properly, comparisons between and among districts and schools are fair, and inferences about achievement are defensible. When the district reports these results to the public, it makes a big point of saying that it has followed these guidelines to the letter and spirit, as a means for establishing credibility and trust.

The story presents one plausible scenario for how the voluntary national test might be implemented in public schools, but other scenarios are possible as well. The story is focused on the future because effects of the proposed voluntary national test would not be fully achieved in its first year. But two things are clear. If there is to be such a test, it should be made available to all who would find value in it, whether state, public school, private school, home school, or individual parent. And, while the federal government would provide resources to make the tests available, there should be no federal coercion, sanctions, or rewards for participating.

The story emphasizes that, while having widely recognized standards and assessments can provide focus for planning and a common language for students, parents and teachers, what is most important is what parents, students, and educators actually do with that knowledge. The story, implicitly, also suggests that a wide voluntary mobilization of private resources in society reinforcing the value and importance of learning (e.g., business, media, etc.) would be important.

The Purpose of the Voluntary National Test

As the Governing Board worked on this report, it became evident that purpose, intended use, the definition of voluntary, and means for reporting are, to a large degree, interdependent. A change in any one of these could affect the others. Therefore, it is important that these four areas be coherent.

In addition, the test should serve a unique purpose. The proposed voluntary national test would provide the only consistent measure of how individual students are performing, across states and school systems, against rigorous standards based on the National Assessment of Educational Progress. If the same purpose were already being fulfilled by another testing program, there would be no need for the voluntary national test. If the same purpose could easily be fulfilled by another testing program, it would be prudent to consider that possibility in weighing the pros and cons before proceeding with full implementation.

The National Assessment Governing Board suggests that Congress and the President consider the following as the purpose of the proposed voluntary national test:

To measure individual student achievement in 4th grade reading and 8th grade mathematics, based on the content and rigorous performance standards of the National Assessment of Educational Progress (NAEP), as set by the National Assessment Governing Board (NAGB).

Rationale

The legislation giving responsibility for voluntary national test development to the Governing Board does not specify or limit the subjects and grades to be tested. However,

the accompanying conference report does direct that the tests be based on NAEP content and NAEP performance standards and be linked to NAEP to the maximum extent possible. The Governing Board in August 1996 adopted a policy on NAEP redesign. The redesign policy provides for testing at grades 4, 8, and 12 at the national level in 11 subjects and, based on the needs and interests expressed by states, at grades 4 and 8 at the state level in reading, writing, mathematics and science.

Grades 4, 8, and 12 are transition points in American schooling. Consistent with the National Assessment redesign policy and the congressional directive that the voluntary national tests be designed to parallel NAEP, the Governing Board limited the test development contract to cover grade 4 reading and grade 8 mathematics. Proficiency in these subjects, by these grades, is considered to be fundamental to academic success.

Most importantly, measuring individual student achievement based on the National Assessment affords this proposed testing program a unique niche among K-12 academic testing programs in the United States. For 30 years, the National Assessment has reported the status and progress of student achievement on nationally representative samples of students. It has done so with credibility, technical competence, and widespread acceptance. For the last ten years, the National Assessment also has reported on state-representative samples of students in volunteering states, providing participating states with the only available comparable measure of student achievement.

However, the National Assessment, by law, does not provide individual student results. It provides only group-level results (e.g., for students overall, by gender, by race, by type of school, etc.). The NAEP state-level assessments represented a watershed event. Ten years ago, state-level assessments were begun amidst fears of encroachment on state and local autonomy and worry that a national curriculum would result. The promise that the NAEP state-level assessment program would serve a unique function—to provide comparable state results, trends over time, and an external validity check for state standards and assessments—has been realized. The fears have not. This is because there are checks and balances built into the governance of the program.

Today, similar fears of federal encroachment and the emergence of a national curriculum arising from the voluntary national test are being expressed and must be addressed. As with the NAEP state assessments, checks and balances can be provided for in the governance and operation of the voluntary national testing program to address these reasonable concerns.

Definition of the Term ‘Voluntary’

There are two dimensions to the definition of the term “voluntary” as it would apply in the administration of the voluntary national tests. The first dimension has to do with the role of the federal government. The second dimension has to do with who makes the decision to participate in the voluntary national tests.

Federal Role

The role of the federal government in the proposed voluntary national tests should be limited. The federal government should not make any individual take the voluntary national tests or require any school to administer the tests. The federal government should have no control or authority over any data resulting from the administration of the voluntary national tests, nor should participation in the voluntary national tests be a condition for receiving federal funds.

The National Assessment Governing Board suggests that Congress and the President consider the following as part of the definition for the term “voluntary”:

The federal government shall not require participation by any state, school district, public or private school, organization, or individual in voluntary national tests. The federal government also shall not make participation in voluntary national tests a specified condition for receiving federal funds or require participants to report voluntary national test results to the federal government.

Rationale

The definition of the term “voluntary” must include stringent limits on the role of the federal government. The limits on the federal role should be specified in legislation and designed to prevent any encroachment on state, local, and private school autonomy. Several witnesses in the Governing Board’s public hearings argued that the 55 mile-per-hour speed limit was voluntary, too, but became universally implemented by states (and in that sense was “mandatory”) because it was a specified condition required to receive federal highway funding. The definition of “voluntary” provided here would foreclose such an outcome. However, it would not foreclose any federal grantee from choosing to use the voluntary national test to meet a general reporting requirement, assuming both that other options are available and the requirements could be fulfilled validly and appropriately by the voluntary national tests. On the one hand, it is not fair to require that the VNT be used. On the other hand, it is not fair to foreclose its use as long as the decision to use the VNT is made without coercion and solely at the participant’s discretion.

Who Decides to Participate

Since the federal government will not coerce participation, it will be up to others to decide whether to participate. Education governance of public schools in the United States, about 88 percent of K-12 school enrollment, is vested in state and local public authorities. Responsibility for the remaining 12 percent of K-12 school enrollment resides with private school authorities and parents.

The definition of “voluntary” needs to accommodate a wide range and diversity of governance authority. For example, there is great variation among state laws in the

degree of central authority and responsibility for education and the degree of local district autonomy. Similarly, there are differences among private schools in how they are governed as well as among state laws regarding the oversight of private schools and home schooling. Provisions for who decides to participate should accommodate this range and diversity of authority and must be made in a manner that does not conflict with state and local law and policy.

With respect to who decides to participate in voluntary national tests, the National Assessment Governing Board suggests that Congress and the President consider the following:

Public and private school authorities should be afforded the option to participate in the voluntary national tests. For public schools, state and/or local law and policy should determine whether the initial decision to participate is made at the state level or at the local district level. Where state law or policy provides that the initial decision be made at the state level, and the state decides not to participate, school districts should be afforded the opportunity to decide whether to participate, to the extent permitted by state and local law and policy.

For private schools, the decision to participate should be made by the appropriate governing authority.

Parents may have their children excused from testing as determined by state and local law and policy in the case of public schools. In the case of private schools, parents may have their children excused from testing as determined by the policy of the appropriate governing authority.

Parents whose schools are not participating but want their children to take the voluntary national tests should have access to the tests either through a qualified individual or testing organization before the tests are released to the public or through dissemination procedures at no or minimal cost (e.g., public libraries and the Internet) after the tests are released to the public.

Rationale

The definition of “voluntary” adopted by the Governing Board is intended to align with state and local law and policy regarding the authority to make decisions about testing. The definition is designed to allow for choice in providing the opportunity to participate, but without exceeding the authority of the federal government in this sensitive area, without coercion by the federal government, and without intruding on the prerogatives of states, school districts, private schools, and parents.

Typically, if not universally, determinations about testing are made by school authorities, whether state, local, or private (including home schools). They determine what should be

tested, what grades should be tested, the time of year for testing, the content of reports on test results and the use of the results. These authorities decide whether tests will be taken by all students or by a sample of students. Therefore, the definition of “voluntary” is designed to account for the fact that schools are the most likely venue through which the proposed voluntary national tests would be administered and that school authorities decide which tests will be given. At the same time, the definition of “voluntary” recognizes and accommodates the variation in responsibility and authority for education governance that exists across state boundaries among states and schools.

School authorities also decide the extent to which official policies will provide for parental intervention to have their children excused from testing. The definition of “voluntary” intends to accommodate this variability as well, again, without intruding on local prerogatives.

Finally, the definition of “voluntary” recognizes that there could be instances in which school authorities decide not to participate in the voluntary national tests, but certain parents want their children tested. In such cases, parents may elect to have their children tested by appropriately licensed or recognized individuals or organizations. Because all parents who may wish to have their children take the test may not have the resources to pay for private testing, the test and scoring guides could be made available for free, or at a minimal charge, after the period for conducting the testing is completed.

Intended Use of the Voluntary National Tests

The intended use of the voluntary national tests is related to the statement of purpose and definition of “voluntary” suggested above. The Governing Board suggests that Congress and the President consider the following as the intended use of the proposed voluntary national tests:

To provide information to parents, students, and authorized educators about the achievement of the individual student in relation to the content and the rigorous performance standards for the National Assessment, as set by the National Assessment Governing Board for 4th grade reading and 8th grade mathematics.

Rationale

The proposed intended use of the voluntary national tests is purposely narrow, and appropriately so. Consistent with the purpose statement, which is to measure individual student achievement, the intended use is to provide information describing the achievement of the individual student. Upon receiving the results of the test, parents, students and teachers will have an overall measure of the individual student’s achievement in 4th grade reading or 8th grade mathematics. As described in the following section on reporting, they will have information on the performance standard reached by the student and other detailed related information.

With information in hand from the voluntary national tests and other sources about the child and the school program, it is expected that: (1) parents could become more involved with the child's education, (2) students could study hard and learn more, (3) teachers could work more to emphasize important skills and knowledge in the subjects tested without narrowing or limiting their curricula, and (4) parents, students, and teachers could have a means for better communication about the child's achievement.

While such outcomes can be hoped for, their achievement relies on local effort, resources, skill, and persistence. No testing program alone can determine, ensure, or constrain what will be done with the information it provides. However, when the values of a society at large are focused on a clear goal widely recognized as important, with consistent methods for monitoring progress toward that goal, the likelihood is increased that local effort, resources, skill, and persistence will voluntarily be brought to bear to achieve that goal.

The Governing Board does not assume that uses of data from voluntary national tests beyond the intended use described above are necessarily inappropriate or should be prohibited to states, districts, and private schools. Any such additional use of voluntary national test data would be done at the discretion of the participating state, district, or private school authorities, who would be responsible for following appropriate technical standards and validation procedures.

However, the voluntary national tests are not tied to a preferred curriculum, teaching method or approach. The voluntary national tests are based on the content of the National Assessment of Educational Progress. The content of each NAEP test is developed by the Governing Board through a national consensus process involving hundreds of educators, curriculum specialists, school administrators, parents, and members of the public. The content of NAEP is designed to assess what students know and can do, not how they are taught.

The voluntary national tests also are not designed to diagnose specific learning problems or English language proficiency. Tests for such diagnostic purposes are specifically tailored. For example, a test of English language proficiency may involve speaking and listening as well as reading. A test to diagnose specific learning problems may include motor coordination and perception, but may or may not include mathematics skills. Tests for the general population, such as the voluntary national tests, are inappropriate for these diagnostic purposes.

The voluntary national tests are not intended to be used as the sole criterion in making "high stakes" decisions (e.g., placement or promotion) about individual students. This is consistent with professional standards for tests of the American Psychological Association/National Council of Measurement in Education/American Educational Research Association and recommendations in the National Academy of Sciences/National Research Council (NAS/NRC) report "High Stakes: Testing for

Tracking, Promotion, and Graduation.” The primary consideration is that, when making decisions on matters such as promotion, retention, or placement, scores from large-scale assessments should only be used in combination with other information about student achievement.

The NAS/NRC report also recommends against the use of the voluntary national test in any “high stakes” decision for individual students under any circumstances, whether in association with other sources of information or not. However, it is conceivable that circumstances could exist in which state and/or school district authorities would find information from the voluntary national test useful, in concert with other valid information, in making “high stakes” decisions about individual students. Therefore, it seems prudent not to foreclose such use absolutely. The Governing Board recommends that participating states, districts and schools should be afforded the discretion to decide whether to use the voluntary national test for “high stakes” purposes and the responsibility for ascertaining that such uses are valid.

Reporting the Results of the Voluntary National Tests

Consistent with the purpose and intended use of the voluntary national tests, the National Assessment Governing Board suggests that results of the voluntary national tests be provided separately for each student. Parents, students, and authorized educators (those with direct responsibility for the education of the student) should receive the test results report for the student. Test results for the student should be reported according to the performance standards for the National Assessment of Educational Progress (NAEP). These are the NAEP achievement levels: Basic, Proficient, and Advanced. All test questions, student answers, and an answer key should be returned with the test results; it will be clear which questions were answered correctly and which were not. The achievement levels should be explained and illustrated in light of the questions on the test. Also, based on the nationally representative sample of students who participated in the national tryout of the test the year before, the percent of students nationally at each achievement level should be provided with the report.

There should be no compilations of student results provided automatically by the program. The program should not provide results for the nation as a whole or by state, district, school, or classroom, since the purpose and use of the testing program are directed at individual student level results.

However, it is virtually certain that compilations of student results will be desired and demanded by at least some of the state and district participants and possibly by private school participants as well. These participants should be permitted to obtain and compile the data at their own cost, but they will bear the full responsibility for using the data in appropriate ways and for ensuring that the uses they make of the data are valid.

The Governing Board would develop and provide guidelines and criteria for use by states, districts, and schools for compiling and reporting the data from the voluntary national tests in ways that are appropriate and valid. The guidelines and criteria would,

for example, explicitly require full and clear disclosure about exclusions and/or absences from testing, so that results and comparisons would be accurately portrayed. Access to the test data by external researchers would be made strictly at the discretion of the participating state, district, or private school, as it would with any other testing program, without prejudice because of federal support for the voluntary national test program.

Other Issues

There are several issues which the Governing Board would be remiss not to raise, although they are outside the requirements for this report set by Congress and no attempt is made to resolve them here.

Implementation

By law, the Governing Board has exclusive authority for test development. The Governing Board has been meticulous in staying within the law's boundaries. The Governing Board has focused its efforts on developing test questions and on associated activities. Appropriately, the Governing Board has not taken up implementation issues such as

- the process by which states, districts and schools commit to participate, to what entity the commitment is made, and in what form and of what nature the commitment should be
- how information about the test program and the opportunity to participate will be made available to parents, teachers, and students
- whether and how quality control monitoring of testing should occur
- how printing of test booklets, scoring of student responses, and reporting of test results would be handled
- whether the testing program should be controlled by a federal agency or private commercial interests
- whether all or part of the costs for the test program should be paid by the federal government

Linking the Voluntary National Tests to NAEP

Underlying the concept of the proposed Voluntary National Tests is the desire to measure and report individual student achievement based on the content and rigorous performance standards of NAEP. Indeed, the directive from Congress to the Governing Board is to link the VNT to NAEP "to the maximum extent possible." Accomplishing this linkage presents a significant challenge—one which affects the design of the VNT as well as the manner in which data are calculated and reported. Two tests can be linked to the degree that they have common characteristics, including types of questions, range of content, test administration procedures, etc. Thus, the first task facing the Governing Board is to forge a close relationship between the two tests as the VNT is being created.

Linking two tests also depends upon the particular statistical approach that can be used. Unless a strong statistical procedure can be used legitimately, the VNT results cannot be reported directly on NAEP scales. This would necessarily mean that the VNT may have to be reported without direct reference to NAEP.

Solutions to the challenge of linking will evolve as (and if) work on the VNT continues. The Governing Board intends to develop options to create a good linkage between the VNT and NAEP. If the linkage cannot be established, alternative reporting strategies for the VNT will be prepared. These alternatives would, of course, be based on NAEP content and performance standards to the maximum extent possible.

These questions of implementation and linking do not need to be settled immediately. They will, however, need to be considered and must be settled in a timely manner if Congress and the President decide that the voluntary national test program should go forward.

Summary

This report presents the Governing Board's response to the congressional assignment to determine the purpose and intended use of the proposed voluntary national tests, including the definition of the term "voluntary" and a description of the achievement levels and other means for reporting results. The Governing Board has prepared this report over an eight month period that included extensive deliberation, expert advice, four regional public hearings and two successive periods of public comment (the first to develop the draft report, the second to review the draft report).

Although the legislation requiring the report calls for a "determination," the Governing Board views this report as advisory. Any final determination on these matters would be made in legislation enacted by Congress and signed by the President.

In submitting this report, the Governing Board is neither advocating for or against a voluntary national test. Rather, the Governing Board interprets the assignment from Congress to be to present a sound and logical case about the potential purpose and use of the voluntary national tests.

Recommendation

The Governing Board is submitting this report in June, three months before the required due date of September 30, 1999. This is to assist the Congress and the President in deliberations toward a timely decision on the future of the voluntary national tests.

The Governing Board recommends that a decision be made before September 30. The schedule for the voluntary national test, if the decision is made to proceed, calls for a pilot test in March 2000 of test questions developed by the Governing Board. In order for the pilot test to be properly carried out in March 2000, a decision is needed before September 30, 1999. This will permit the test development contractor to proceed in an orderly and efficient manner to carry out activities that are essential to the pilot test, such as determining the sample of participating schools and arranging for the printing of booklets of test questions.

A decision to proceed that comes too late will set the schedule for the pilot test back one year, to March 2001. This is because pilot testing must occur in the same month that testing is to occur, which is March. If authorization to proceed does not come before September 30, it may not be possible to carry out all of the necessary steps that lead up to the pilot test in time for it to occur in March 2000.

If, on the other hand, the decision is made not to proceed, a decision prior to September 30 will allow for an orderly and cost-effective termination of the test development contract.

It is important to note the purpose of pilot testing. The purpose of pilot testing is to determine the quality of each individual test question. This includes evaluating each question for bias against particular subgroups of students. There are no individual student scores reported. In pilot testing, individual questions are evaluated singly. There are no overall test scores calculated, even though a student in the pilot test will respond to many test questions. The only data collected are statistics that relate to the specific test question, such as the percent of students who answered the question correctly. From the analysis of student responses on the individual test questions, three decisions are possible: drop the test question, keep the test question as is, or keep the test question with changes. Only from the set of test questions that remain after pilot testing will test booklets be constructed, which then will be tried out in field-testing. The field test stage, unlike the pilot test, is designed to simulate the plans for actual testing. If the decision is made to proceed, a field test would be conducted in March 2001.

The optimal outcome would be to have a timely final decision on whether or not there shall be voluntary national tests. Another possible outcome would be to have agreement to proceed with the pilot test of questions, while continuing to deliberate on the prospects for the voluntary national test program itself. If the pilot test proceeds, the test questions could be considered for use in the National Assessment of Educational Progress, should the ultimate outcome be the continuing prohibition of voluntary national tests.



National Assessment Governing Board

National Assessment of Educational Progress

Student Performance Standards on the National Assessment of Educational Progress: Affirmation and Improvements

*A Report to Congress on the Long Range Plans
For Setting Achievement Levels on NAEP*

Adopted Unanimously by
The National Assessment Governing Board, June 23, 1999

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

Student Performance Standards on the National Assessment of Educational Progress: Affirmation and Improvements

Background

In September 1998, the National Academy of Sciences (NAS) issued an evaluation of the National Assessment of Educational Progress (NAEP), titled *Grading the Nation's Report Card*. Chapter 5 of that evaluation focused on the Achievement Levels established by the National Assessment Governing Board (NAGB). The NAS evaluation maintained that the standard-setting process used in establishing the NAEP Achievement Levels is "fundamentally flawed."

On October 22, 1998, the President signed the FY 1999 appropriations act for the Department of Education (the Act) passed by Congress, which contained provisions related to the proposed Voluntary National Test. One provision of the Act requires NAGB to submit a report to the White House and the Congress, responding to the assertion in the NAS evaluation that the process for setting achievement levels is "fundamentally flawed." This report, required no later than September 30, 1999, is in response to that legislative requirement.

The NAS Report makes the following points with which the Governing Board agrees regarding achievement levels:

- **"No clearly proven alternative exists" to the current achievement level setting process.** We agree. The NAS acknowledges that the Governing Board has been prudent in selecting a technical process that is widely used for which there is much validation evidence.
- **"NAEP's current achievement levels should continue to be used on a developmental basis."** We agree. The NAS report does not recommend eliminating the achievement levels, but supports their continued use in reporting data. This point echoes Congress' explicit authorization to NAGB to report to the American public using achievement levels.
- **"Performance standards serve an important evaluative function for NAEP's users."** "Many educators, policy makers, and parents have come to expect reports that state whether observed performance levels are good enough ." We agree. Many states are now successfully using standards-based reporting as well. The media, the states participating in the NAEP program, and the general public can comprehend more easily an increase in the percentage of proficient students than a statistically significant increase on a 0-500 scale.
- **"Tracking changes in the percentages of students performing at or above those cutscores can be of use in describing changes in student performance."** We agree. For mathematics and reading, there now are three separate data points for the achievement levels for states to measure growth in achievement.

- **“Standard setting rests on informed judgment.”** We agree. There has long been professional consensus that standard setting on any assessment is a matter of informed judgment, not a scientific statistical process. Over the past decade, the Governing Board has worked with literally hundreds of teachers, curriculum specialists, and knowledgeable members of the public to develop the achievement levels. These achievement levels are intended to be challenging but reasonable standards of what students should know and be able to do in reading, mathematics, and other academic subjects assessed by NAEP.
- **“Preliminary achievement levels descriptions should guide the development of assessment items and exercises.”** We agree. These achievement levels descriptions have played an increasing part in the development of assessments since 1990.

However, the Board **does not agree** with the NAS that the current process for setting the NAEP performance standards is “fundamentally flawed.” Indeed, the method of aggregating item-by-item judgments of panels of raters, the Angoff method, may be the nation’s most widely-used formal process for setting performance standards on tests. It is employed for setting high school graduation standards in many states. In addition, it is used by national commercial test publishers. Over the past 20 years it has withstood many legal challenges.

The Governing Board’s efforts over the past decade to set standards have been thoughtfully and carefully considered. The achievement levels are set by the Board as part of a judgmental process. Moreover, the Governing Board has sought and continues to seek external advice from nationally and internationally known psychometricians. The Governing Board’s contractor, ACT Inc., has a world-class reputation for setting standards in education and many other fields as well. In addition, the Governing Board has taken seriously the charge to develop the current standards through a national consensus approach.

The Governing Board is firmly committed to continuous review and improvement of the entire process. This is in line with the Governing Board’s past actions to improve the standard setting process. In response to previous evaluations and other advice, the process of developing levels for the Governing Board’s consideration has been changed and improved over almost a decade. Some of the Governing Board’s efforts toward improving the process include:

- Since 1991, the initial part of the process of developing information on achievement levels has been managed by the Governing Board under contract to ACT, which has extensive experience in standard setting in education and professional licensure fields. Improvements have been made in each standard setting process during the past decade.
- The Governing Board has consulted on a regular basis with state testing directors, many of whom have used similar methods in setting standards on their own state examinations.
- In 1994, NAGB and the National Center for Education Statistics co-sponsored a Joint Conference on Standard Setting to discuss the best standard setting processes available at that time.

- In 1994 NAGB conducted a design competition for a contract to improve the standard setting procedures. In 1996 NCES conducted a similar competition as part of the grant proposals for NAEP. Both efforts produced either no proposals or those which were judged to be no improvement on the current process.
- As part of the standard setting contract with ACT, NAGB has solicited input from an ongoing Technical Advisory Committee on Standard Setting (TACSS). Based on the advice of TACSS experts as well as requests from NAGB's Achievement Levels Committee, ACT has conducted numerous studies of its standard setting process. Over the years, these have led to many refinements.

There are those who would wait for the "perfect" model for setting standards; there are those who do not wish to have standards at all and are not committed to educational standards of any kind. The Governing Board's position is that there is no "perfect" model and the one currently in place is state-of-the-art for this time and place. The 1994 Joint Conference on Standard Setting, sponsored by the National Center for Education Statistics and the National Assessment Governing Board affirmed that conclusion. Even as it criticized the current model, the NAS report failed to identify any proven alternative. The Board is not willing to discard a well-established method until there is good, solid evidence that a better method is available for the purpose of large-scale national assessment.

The Governing Board is fully committed to the value of standards-based reporting, a view held by many in American education today. The Governing Board believes that setting clear markers of what students should learn makes any assessment far more meaningful to parents, schools, and the public than simply reporting average scores or describing points along the bell-shaped curve.

It should be noted that some of the most respected expert practitioners in psychometrics and standard setting disagree with the premises, the logic, and the conclusions reached by the NAS. These experts serve as technical advisors to ACT on the NAEP standard-setting process. The review of Chapter 5 of the NAS report by Hambleton et al (1999) states that the NAS evaluation of the NAEP achievement levels is not credible:

“In this paper, we have provided a detailed response to many aspects of the review/evaluation of NAGB’s initiatives to set achievement levels on the NAEP. Our review of the evidence suggests that Chapter 5 on “Setting Reasonable and Useful Performance Standards” for NAEP constitutes a one-sided, incomplete and inaccurate accounting of the NAGB/ACT standard-settings conducted during this decade. As such, Chapter 5 is a disservice to NAGB, educational policy makers, educators and the public. In our opinion, in this instance, the NAS review and screening process failed to produce a trustworthy and credible evaluation.” (pp. 27-28)

The Governing Board acknowledges that its standards are challenging but considers them reasonable. The central idea in defining the proficient level for each grade is that students demonstrate "competency over challenging subject matter." This language is taken directly from the National Education Goals, proposed by President Bush and nation's governors

(Charlottesville Summit, 1989) and later enacted by Congress. The Board deliberately seeks not to ratify the status quo.

General Overview of Plans for Improvement

The Governing Board has always been committed to continual improvement of the standard setting process. The Governing Board has chosen to take this opportunity to conduct a review that goes well beyond what the NAS report calls for.

The Governing Board plans to initiate a thorough and thoughtful evaluation, and to reconsider all aspects of the standard setting process, including examining alternative standard setting methods. At the same time, the Governing Board remains mindful of the value of stability. Meaningful reporting of achievement trends requires a stable measurement scale. In addition, a stable benchmark or score scale can become more useful over time, acquiring additional meaning and interpretations as users acquire experience with the scale.

First, the Governing Board will commission a thorough review of its policies and practices, including the implications and interpretations of those policies and practices.

Second, the Governing Board will explore alternative methods for developing the standards. The Board will actively seek promising alternative models to be piloted in small studies, using data from a previous assessment.

Timeline

The Governing Board will initiate these studies now, and expects to complete them in two phases over the next two years. Governing Board deliberations on the initial phases will occur over the next three Governing Board meetings (August 1999 through March 2000).

Phase 1 examines fundamental policy issues, reviews Governing Board policies and practices, solicits public perceptions of the standards, and reviews alternative standard setting models.

Phase 2 consists of piloting one or more standard setting models that show potential for use in determining NAEP's achievement levels.

Phase 1 will continue through the November 2000 Governing Board meeting. Phase 2 will be initiated before the completion of Phase 1. The results of these deliberations will be used primarily to inform the Governing Board's decisions on future standard setting, both for NAEP (starting as early as mathematics in 2004), and for the proposed voluntary national tests (depending on Congressional action and starting no earlier than field testing in 2001). Therefore, the project phases have been scheduled to allow for a thoughtful and thorough reevaluation of this important work of the Governing Board. In addition, this schedule has been designed to foster broad public input from the primary audience for NAEP results, namely the American

public, and from the primary users of NAEP data, namely policymakers at the national, state, and local levels, and the research community.

Phase 1: Governing Board Review

Examination of Fundamental Conceptual Issues

In 1989 the Governing Board developed its initial policy position for setting student performance standards on NAEP. The first Governing Board policy titled *Setting Appropriate Achievement Levels for the National Assessment of Educational Progress: Policy Framework and Technical Procedures (May 11, 1990)* guided the inaugural effort to set the levels on the 1990 NAEP mathematics assessment. This document was intended solely as an initial policy statement that would apply to the 1990 NAEP mathematics assessment with the expectation that it would be revised in the future as the Governing Board learned more about standard setting in general, and specifically for NAEP in a large-scale assessment context.

Over the past ten years much has happened to change the meaning embedded in the original policy and in subsequent versions. The conceptual underpinnings of the levels were not fully articulated ten years ago, and, indeed, some have only come to light during the last decade of use and interpretations of the levels. Though not intended to be exhaustive, the following list of issues is illustrative of those that the Governing Board will be considering over the next several months. Note too, that not all of the following issues are separate and distinct. Also, some issues may be obviated by the resolution of others. The most salient issues include:

- Role and function of the Governing Board in the standard setting process;
- Clarity of the policy definitions;
- Role and meaning of achievement levels descriptions;
- Reporting of NAEP levels;
- Role of exemplar exercises in understanding and interpreting the levels; and
- Intended interpretations of NAEP reports based on achievement levels

Review of Governing Board Policies and Practices

A synthesis of the Governing Board's policies and practices from the inception of the standards in 1990 to the present time will be developed. The purpose of the synthesis is to clarify both the articulated and tacit policies and practices of the Governing Board. This review will also catalog, by subject area, various criticisms of the standards. This review will determine how policy and practice have changed in response to criticism over the past decade.

Public Perception of the Standards

Through focus groups, information will be gathered about public perceptions of the primary NAEP audiences with respect to the three legislated criteria, i.e., reasonable, valid, and informative to the public. Although there is much evidence to suggest that the levels are useful and informative, which was acknowledged in the NAS report, questions still remain which can

be explored in depth through focus groups designed for that purpose. Four of these information-gathering sessions will be held around the country with different audiences, including, but not limited to, parents and the general public, policymakers, educators and state assessment directors, and the business community.

In addition, this aspect of the Governing Board's study will include tracing media coverage of the levels over the past ten years and conducting a content analysis of the media releases to answer a variety of questions. What is being reported? Is it being reported accurately? Who is reporting and using the levels? What are the intended and unintended consequences of this reporting? How can reporting to the public be improved?

The review will also examine what kinds of standards commercial test publishers are using in their "off-the-shelf" tests. How are these standards similar to or different from the achievement levels? How are they developed? Which states and local jurisdictions are using them?

Finally, the review will examine State-level efforts to develop standards. The review will explore where NAEP-like standards are being used, and where the actual NAEP standards (e.g., the policy definitions of *Basic*, *Proficient*, and *Advanced*) have been used by states.

Review of Alternative Models for Developing Achievement Levels

A literature review will be prepared which examines other standard setting models developed since the Joint Conference on Standard Setting. This review will examine a number of important questions. Are there trade-offs with some models that are greater than those with other models? How does one weigh the costs and benefits of one model over another? Which models have been researched over the past decade that were found not to work so well in the NAEP setting? Are there theoretical grounds for favoring any of the newer models over the current model?

Based on the results of this review, designers of alternative models will be invited to prepare proposals employing their models and showing how they could be implemented in NAEP. An independent technical advisory group will review the proposals before piloting the models.

Governing Board Review

The goal of gathering this information is to assist the Governing Board in advancing the quality of standard setting in the future. An Achievement Levels Committee Workshop will be scheduled to inform the Governing Board about potential alternative standard setting procedures. The Committee will use the information coming from the various sources outlined above to revise its policies and practices, and to decide which procedures they judge likely to be successful for the future of NAEP standard setting.

Phase 2: Model Piloting

Once the Governing Board has determined which models to study more carefully, a series of implementation designs will be developed which will allow a full-scale pilot of the models with existing data from at least one grade level of one subject area. A common data set will be used

for each pilot so that comparisons can be made from one model to the next among the *products* of the standard setting, namely, the *cut scores*. The *process* for each model will be evaluated by NAGB for evidence of validity. The alternative designs will include any and all modifications anticipated during the operational implementation of the model.

It is hoped that this thorough review and empirical research on alternative models will be a fruitful activity and will advance the field of setting student performance standards beyond its current state of development.

Governing Board's Commitment

The Governing Board is committed to meeting the Congressional mandate to set student performance standards. However, we are convinced there is a fundamental tension between those who advocate continuing research, and the public's interest in sound information about what students know and can do in reading, writing, mathematics, science, and other academic subjects. The Governing Board believes that the search for the *perfect solution* should not undermine the use of *well-established state-of-the-art approaches currently available*.

No educational standard-setting process in history has been more carefully developed, more thoroughly researched, or more closely scrutinized than the process followed in establishing the NAEP Achievement Levels. It is the consensus of the National Assessment Governing Board, ACT, and the testing experts who advise NAGB and ACT that the procedures followed in establishing the NAEP achievement levels are the very best available to use in the unique context of the National Assessment of Educational Progress.

References

- Bourque, M.L. (Ed.) (1995). *Proceedings of the joint conference on standard setting*. Washington, DC: National Assessment Governing Board and the National Center for Education Statistics.
- Hambleton, R.K., Brennan, R.L., Brown, W., Dodd, B., Forsyth, R.A., Mehrens, W.A., Nellhaus, J., Reckase, M., Rindone, D., van der Linden, W. J., & Zwick, R. (1999). *A response to "Setting Reasonable and Useful Performance Standards" in the National Academy of Sciences' Grading the Nation's Report Card*. Report available from the first author at Laboratory of Psychometric and Evaluative Research, University of Massachusetts, Hills House South Rm. 154, Amherst MA 01003. E-mail: rkh@educ.umass.edu.
- National Assessment Governing Board. (May, 1990). *Setting appropriate achievement levels for the National Assessment of Educational Progress: Policy Framework and Technical Procedures*. Washington, DC: Author.
- Pellegrino, J.W., Jones, L.R., & Mitchell, K.J. (1998). *Grading the Nation's Report Card*. Washington, DC: National Academy Press.



National Assessment Governing Board

National Assessment of Educational Progress

CLINTON LIBRARY PHOTOCOPY

The National Assessment of Educational Progress:

Design 2000-2010

Adopted Unanimously by

The National Assessment Governing Board, June 23, 1999

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

The National Assessment of Educational Progress: Design 2000-2010

What should the Nation's Report Card on student achievement look like during the next decade? How can it most effectively help the public understand the academic readiness of our youth at grades 4, 8, and 12—key transition points in American education? Ultimately, how can the National Assessment of Educational Progress (NAEP) best be used as an indicator of national and state educational preparedness for the challenges facing our society?

The purpose of this report to Congress and the President is to describe the recommendations of the National Assessment Governing Board for answering these questions. The report will provide a summary of the Governing Board's policy to redesign the National Assessment, describe the status of implementation of the redesign policy, and address the implications for reauthorization of the National Assessment of Educational Progress.

Background

In 1996, prompted by increasing demand for more and more frequent information about the status and progress of student achievement in the United States, the National Assessment Governing Board, an independent, bipartisan citizen's group created by Congress to set policy for the National Assessment, charted a course for NAEP through the year 2010. The policy to redesign the National Assessment followed two years of study, expert advice, deliberation by the Governing Board, and public comment (Appendix A).

In 1997, the National Center for Education Statistics (NCES) developed a plan to implement the redesign policy. The plan has two phases. The first phase covers assessments in the years 1999-2003. In 1998, NCES awarded new contracts for NAEP covering this period. During this first phase, the Governing Board's annual schedule of assessments will be carried out (see Table 1), National Assessment student achievement data will be released more quickly, National Assessment reports will be redesigned for the general public, and research will be conducted to foster a streamlined design for the National Assessment. The second phase of National Assessment redesign, covering assessments for the years 2004-2007, will continue the earlier improvements and begin to implement the innovations aimed at streamlining the design of NAEP.

Even as redesign implementation begins under the new contracts, the Governing Board has continued to weigh new evidence that may bear on the shape of the NAEP redesign policy. Following the adoption of the redesign policy in 1996, there have been evaluation reports issued on the National Assessment, reviews by other experts, and papers prepared for the November 1998 Ten-Year Anniversary Conference sponsored by the Governing Board. The views expressed raise issues or concerns that bear on several areas of the redesign policy (e.g., when to change test frameworks, how to simplify NAEP's technical

design, how to improve the process for setting achievement levels, and the extent to which NAEP results can be used to explain the causes that underlie student achievement). The Governing Board decided to examine once again these areas of the redesign policy to determine whether any modifications to the policy were in order. These policy areas were reviewed in detail in a forum conducted by the Governing Board on April 15, 1999 with technical experts, consumers of NAEP data, representatives from the National Center for Education Statistics and the NAEP contractors. The results of the April 15 forum are incorporated in this report.

One area in which the Governing Board and NCES have worked diligently, but which is not addressed in this report, is policy and research regarding inclusion and accommodations for students with disabilities and students with limited English proficiency. This work is important and will continue on a parallel course with NAEP redesign.

National Assessment Redesign: A Summary and Status Report

Introduction: The Redesign Principles

Over its thirty-year history, the National Assessment has earned respect and credibility. The National Assessment is widely recognized for the comprehensiveness of its tests, the quality of its technical design, the accuracy of its reports, and innovation in its execution. The data produced by the National Assessment are unique. No other program provides regular reports on the status and progress of student achievement for our nation as a whole and that are comparable state-by-state.

Although its original purpose was to measure and report on the status of student achievement and on change over time, recognition of the quality and integrity of the National Assessment led to a multitude of demands and expectations beyond reporting on achievement. Meeting those expectations was done with good intentions and seemed right for the situation at the time. However, some additions that the National Assessment performs less effectively were “tacked on” to the original design.

The National Assessment was being asked to do too many things, some even beyond its reach to do well, and was attempting to serve too many audiences. For example, in contrast to the 1970’s in which a single 120 page report on mathematics was deemed sufficient, the 1992 NAEP mathematics reports numbered seven and totaled about 1,800 pages.

The result of attempting to respond to demands beyond NAEP’s central purpose was to overburden NAEP’s design, drive up costs and reduce the number of subjects that could be tested. For example, the National Assessment tested two or three subjects each year during the 1970’s, its first decade, but only every other year after the 1980’s. Another indicator that NAEP had too many distractions was that results could be released as many as two to three years after testing. This simply was not acceptable, particularly with the advent of state-level assessments in the 1990’s.

The Governing Board's solution was to focus NAEP on what it does best: measure and report on the status of student achievement and change over time. Focusing NAEP on what it does best would permit NAEP's design to be simplified and also would mean putting limits on demands that are outside NAEP's central purpose. Another part of focusing NAEP is to define the audience for reports. The Governing Board has determined that the NAEP program should not attempt to serve multiple audiences directly. The audience for reports should be the general public.

Specialized needs for NAEP data should be accommodated by making the NAEP data easily accessible for analysis by others—educators, researchers, policymakers, and the media, among others. In order to make data more understandable and useful to the general public, the Governing Board has determined that achievement levels, or performance standards, should be the primary means for reporting NAEP results.

Thus, five principles undergird the Governing Board's policy for the redesign of the National Assessment:

- Conduct assessments annually, following a dependable schedule
- Focus NAEP on what it does best
- Define the audience for NAEP reports
- Report results using performance standards
- Simplify NAEP's technical design

Details on these and other aspects of the redesign policy follow.

Annual Schedule

A centerpiece of the National Assessment redesign is a dependable annual schedule of assessments through the year 2010 (Table 1). In the past decade, the focus on education reform, new and revised state assessments, and the national education goals have led to demand for National Assessment testing more frequently than the biennial schedule of the 1980's and most of the 1990's. The schedule for the period 1996 through 2010 was adopted in March 1997 and revised in November 1998. It provides for annual assessments at the national level and state-level assessments in even-numbered years. The long-term trend assessments in reading, writing, mathematics, and science continue on a once per four-year cycle beginning in 1999.

At the national level, grades assessed will be 4, 8 and 12. Subjects covered will be reading, writing, mathematics, science, geography, U.S. history, world history, civics, economics, foreign language, and the arts. These are the subjects listed in the current national education goals. Reading, writing, mathematics and science will be assessed once every four years. Other subjects will be assessed less frequently, but there will generally be two assessments in a subject over a ten-year period.

Testing at the state level will occur in even-numbered years, with reading and writing in grades 4 and 8 alternating with mathematics and science in grades 4 and 8. Student

Table 1. Schedule for the National Assessment of Educational Progress

The following schedule was adopted by the National Assessment Governing Board on March 8, 1997 and revised in November 1998. Assessments shown as scheduled for 1996, 1997, and 1998 were approved previously by the Board.

YEAR	NATIONAL	STATE
1996	Mathematics Science Long-term trend* (reading, writing, mathematics, science)	Mathematics (4, 8) Science (8)
1997	Arts (8)	
1998	Reading Writing Civics	Reading (4, 8) Writing (8)
1999	Long-term trend*	
2000	Mathematics Science Reading (4)	Mathematics (4, 8) Science (4, 8)
2001	U.S. History Geography	
2002	Reading Writing	Reading (4, 8) Writing (4, 8)
2003	Civics FOREIGN LANGUAGE (12) Long-term trend*	
2004	MATHEMATICS Science	MATHEMATICS (4, 8) Science (4, 8)
2005	WORLD HISTORY (12) ECONOMICS (12)	
2006	READING Writing	READING (4, 8) Writing (4, 8)
2007	ARTS Long-term trend*	
2008	Mathematics SCIENCE	Mathematics (4, 8) SCIENCE (4, 8)
2009	U.S. HISTORY GEOGRAPHY	
2010	Reading WRITING	Reading (4, 8) WRITING (4, 8)

Note: Grades 4, 8, and 12 will be tested unless otherwise indicated. Comprehensive assessments are indicated in **BOLD ALL CAPS**; standard assessments are indicated in upper and lower case (see page 10 for discussion of comprehensive and standard assessments).

* Long-term trend assessments are conducted in reading, writing, mathematics and science. These assessments provide trend data as far back as 1970 and use tests developed by the National Assessment at that time.

achievement results in these subjects and grades at the state level will be reported on a once per four-year basis.

Many of the other redesign policies, described below, are aimed at making the annual schedule affordable through cost-saving efficiencies.

Status of Implementation:

The work in the new NAEP contracts covers the schedule as adopted by the Governing Board for the years 1999-2003. The long-term trend assessments in reading, writing, mathematics, and science will be conducted in 1999 and 2003. In 2000, mathematics and science assessments will be conducted in grades 4 and 8 at the state level and at grades 4, 8, and 12 at the national level. In addition, a reading assessment at grade 4 at the national level will be conducted. In 2001, geography and U.S. history assessments will be conducted at grades 4, 8, and 12 at the national level. In 2002, reading and writing assessments will be conducted at the state level in grades 4 and 8 and at the national level in grades 4, 8, and 12. In 2003, assessments will be conducted at the national level in civics in grades 4, 8, and 12 and in foreign language at grade 12.

Define the Audience for NAEP Reports

The expanded demands and expectations noted above reflected the many varied audiences that NAEP was attempting to serve. Trying to serve too many audiences has meant that no audience is optimally served by the National Assessment. The NAEP redesign policy makes the distinction between the audience for reports prepared by the NAEP program and the users of NAEP data. The audience for NAEP reports is the American public. The primary users of NAEP data are national and state policymakers, educators, and researchers.

This distinction in the policy between the audience for reports and users of data is important. It is intended to address the needs of various groups and individuals interested in NAEP results, while providing an appropriate division of labor between them and the federal government.

National Assessment reports released by the National Center for Education Statistics should be objective, providing the facts about the status and progress of student achievement. Providing objective information about student achievement is an appropriate federal role. Since the public is the primary audience, NAEP reports should be understandable, jargon free, easy to use, widely disseminated, and timely.

On the other hand, the redesign policy suggests that interpreting NAEP data (e.g., developing hypotheses about achievement from relationships between test scores and background questions) is a role that falls primarily to those outside the Department of Education—the states that participate in NAEP, policymakers, curriculum specialists, researchers, and the media, to name a few. For the NAEP program itself to address the myriad of interests and questions of these diverse groups seems both impractical and inappropriate. However, the federal government should encourage and provide funds for

a wide range of individuals and organizations with varied interests and perspectives to analyze NAEP data and use the results to improve education. This is the point of the redesign policy. Thus, the redesign policy provides that National Assessment data are to be made available in easily accessible forms to support the efforts of states and others to analyze the data, interpret results to the public, and improve education performance.

Status of Implementation:

The National Center for Education Statistics is placing a high priority on “highlight” reports and national report cards for each subject, which are aimed at the general public. NAEP data will be accessible through a new Internet web site, customized for particular data users. Priorities for NAEP secondary analysis grants were revised to encourage wider use of NAEP data by national and state policy makers, educators, and researchers and to focus the analyses on interpretive and education improvement purposes. Also, NCES is continuing to develop and provide training on software for analyzing NAEP data.

Report Results Using Performance Standards

In 1988, Congress created the Governing Board and authorized it to set performance standards—called achievement levels—for reporting National Assessment results. Under the redesign policy, achievement levels are to be used as the primary (although not exclusive) means for reporting National Assessment results. The achievement levels describe “how good is good enough” on the various tests that make up the National Assessment. Previously, the National Assessment reported average scores on a 500-point scale. There was no way of knowing whether a particular score represented strong or weak performance and whether the amount of change from previous years’ assessments should give cause for concern or celebration. The National Assessment now also reports the percentage of students who are performing at or above “Basic,” “Proficient,” and “Advanced” levels of achievement.

The achievement levels have been the subject of several independent evaluations, some controversy, and conflicting recommendations. Recommendations have been carefully considered and some have been used to improve the standard-setting procedures. While the current procedures are based on what may be the nation’s most widely used formal process for setting standards on tests, the Governing Board remains committed to making continual improvements.

Status of Implementation

The Governing Board will continue to set achievement levels for reporting NAEP results. These achievement levels are to be used on a developmental basis until a determination is made that the levels are reasonable, valid, and informative to the public. At that point, the developmental designation will be removed.

The Governing Board views standard setting as a judgmental, not a scientific, process. However, the process must be conducted in a manner that is technically sound and

defensible. The Governing Board is preparing a report required by Congress to respond to the assertion that the process for setting the achievement levels is “flawed.” This report will include a detailed plan for reviewing the criticisms and compliments found in the evaluation reports that studied the achievement levels. The plan also will address alternatives to the current level-setting procedures.

Simplify the Technical Design for the National Assessment

The current design of the National Assessment is very complex. The redesign policy requires that the research and testing companies that compete for the contract to conduct the National Assessment must identify options to simplify the design of the National Assessment. Examples of NAEP’s complexity include: (1) National and state results are based on completely separate samples. (2) No student takes the complete set of test questions in a subject and as many as twenty-six different test booklets are used within a grade; thus scores on NAEP are calculated using very sophisticated statistical procedures. (3) Students, teachers, and principals complete separate background questionnaires, which may be submitted at different times, complicating their use in calculating assessment results. (4) The data for every background question collected must be compiled before any report can be produced, regardless of whether the data from the background question will be included in a report, lengthening the time from data collection to reporting.

Status of Implementation:

This is a “work in progress.” Options for combining the national and state samples are being developed by the contractors in collaboration with NCES and the Governing Board. Similarly, options to reduce the size of the state sample are being considered. An option to increase the precision of the state results will be implemented in the year 2000 mathematics and science state assessments. Progress also has been made in shortening the time between data collection and reporting by eliminating the requirement to link certain background questionnaires to student achievement data. Plans for a short-form of the National Assessment, using a single test booklet, are being implemented. The purpose of the short-form test is to enable faster, more understandable initial reporting of results and, possibly, for states to have access to test instruments allowing them to obtain NAEP assessment results in years in which NAEP assessments are not scheduled in particular subjects. Plans also are in the development stage for improving the quality, relevance, and efficiency of background questionnaires.

Measure Student Achievement at Grades 4, 8, and 12

The primary purpose of the National Assessment is to measure student achievement at grades 4, 8, and 12 in academic subjects at the state and national level and for subgroups, showing trends over time in the percent of students at or above each achievement level. The subjects to assess are those listed in the national educational goals—reading, writing, mathematics, science, U.S. history, geography, world history, civics, economics, the arts,

and foreign language. Grades 4, 8 and 12 are considered to be important transition points in American education.

Although grade 12 performance is important as an “exit” measure from the K-12 system, there are problems with grade 12 results. The problems are that student and school participation rates and student motivation at grade 12 are low. The Governing Board has considered whether to change NAEP to another grade at the high school level, examining both anecdotal and empirical evidence. Anecdotal evidence about the low motivation of high school students taking low stakes tests in the spring of their senior year raises serious questions about whether NAEP should test at grade 12. However, the empirical evidence in NAEP does not indicate that switching to grade 11 would result in higher motivation on the part of students or greater accuracy in the results. In fact, there is some evidence that twelfth graders taking NAEP may try harder in some cases than eleventh graders. The redesign policy asks the companies that compete for the NAEP contract to find ways to increase school and student participation rates and student motivation. Until they increase, National Assessment reports should include clear caveats about interpreting grade 12 results.

Status of Implementation:

Because the empirical evidence does not warrant a change at this time, NAEP should continue to test at grade 12. New NAEP contracts have been awarded for the conduct of assessments through the year 2003. The contracts are designed to measure student achievement at grades 4, 8, and 12; report state, national, and subgroup results; report trends over time; and use performance standards for reporting results. Caveats for interpreting grade 12 results have been added to reports. However, more attention needs to be placed on improving grade 12 participation rates and student motivation. Toward this end, NCES is planning a series of studies to examine the relationship between student achievement and motivation, including studies of the relationship between courses taken and NAEP results.

What NAEP is Not Designed to Do

The NAEP redesign policy attempts to focus NAEP on what it does best. What the National Assessment does best is measure student achievement. Focusing NAEP on what it does best comes with a related idea—recognizing and limiting what NAEP is not designed to do.

Although the National Assessment is well designed for measuring student achievement and trends over time, it is not a good source of data for drawing conclusions about or providing explanations for the level of performance that is reported. It also is not a measure of personal values, a national curriculum, an appropriate means for improving instruction in individual classrooms, or a basis for evaluating specific pedagogical approaches.

The National Assessment is what is known as a “cross-sectional survey,” an effective and cost-efficient means for gathering data on student achievement. A cross-sectional survey gathers data at one point in time. In the case of NAEP, data are gathered on national and state-representative samples of students at a particular time during the school year. The sample is large enough to permit reasonably accurate estimates of subgroup performance (e.g., by gender, race, and ethnicity). Change over time can be measured by administering the same survey again in later years, under the same testing conditions, with samples of students that are similar to the ones tested earlier. Comparisons can be made within and across the subgroups and for the whole sample.

However, a cross-sectional survey cannot provide answers about what causes the level of performance that is reported. Measuring the causes of achievement would involve an experimental design, with specific research questions to answer, pre- and post-testing of students, and comparisons of results between groups of students receiving a particular educational approach with those that are not. While some may view such research as a worthwhile part of NAEP, the need for pre-and post-testing alone would double the costs of NAEP testing. Because pre-and post-testing would require additional administrative burden on schools and more time away from instruction for students, it could severely hamper school and student participation rates in NAEP, especially with NAEP’s annual assessment schedule. Too few schools and students in the sample, in turn, would jeopardize NAEP’s ability to provide national and state-representative student achievement results.

The best that can be done regarding explanation or interpretation of results is to report on background variables that may be associated with achievement. However, in many cases, the data from background questions collected by NAEP are inconclusive. Even where the associations are stronger, the data are not adequate for supporting conclusions that explain why achievement is at the level reported. Clearly, the use of NAEP background data to explain or interpret achievement results should be done with caution.

Status of Implementation:

Under the new NAEP contracts, the collection of background information will be more focused. The plan is to collect a well-defined core of background information. For example, the well-defined core of background information will include the data that are required for every assessment—e.g., data on gender, race, ethnicity, whether the students are in public or private schools, etc. In addition, each assessment will have a set of background questions designed specifically for the subject being assessed, with each set being determined by policy. Therefore, the background questions for the mathematics assessment will vary from those for the science or reading assessments.

The intent is not only to be more purposeful about what is collected, but more strategic about how it is collected as well. For example, in the past, information on TV watching by students was collected regularly as a part of every assessment. In the same year, the same background questions could be asked of the students in each separate national sample. Clearly, where two or more subjects are being assessed in a particular year, it may not be necessary to ask identical questions across all of the assessments. Similarly,

it may not be necessary to ask certain questions every year. In addition, the background questions themselves will be pilot tested to reduce the possibility of misinterpretation.

Reporting NAEP Results

The redesign policy provides that National Assessment results should be released with the goal of reporting results six to nine months after testing. Reports should be written for the American public as the primary audience and should be understandable, free of jargon, easy to use and widely disseminated. National Assessment reports should be of high technical quality, with no erosion of reliability, validity, or accuracy.

The amount of detail in reporting should be varied. Comprehensive reports would be prepared to provide an in-depth look at a subject the first time it is assessed using a newly adopted test framework, testing many students and collecting background information. Although scale scores also will be used, achievement levels shall continue to be the primary method for reporting NAEP results. Test questions, scoring guides, and samples of student work that illustrate the achievement levels—Basic, Proficient, and Advanced—will receive prominence in reports. Data also would be reported by gender, race/ethnicity, socio-economic status, and for public and private schools; other reporting categories also are possible. Standard reports would be more modest, providing overall results in the same subject in subsequent years using achievement levels and average scores. Data could be reported by gender, race/ethnicity, socio-economic status, and for public and private schools, but would not be broken down further. The amount of background data collected and reported would be somewhat limited in comparison to a comprehensive report. Special, focused assessments on timely topics also would be conducted, exploring a particular question or issue and possibly limited to one or two grades.

Status of Implementation:

The new NAEP contracts provide for faster release of data, standards-based reporting, reports that are targeted to the general public, and three different kinds of reports: “comprehensive,” “standard,” and “focused.” The 1998 national reading results were released in 11 months of testing; the state results in 12 months. Although still short of the Board’s goal of reporting results in 6 to 9 months following testing, progress is being made.

Simplify Trend Reporting

The NAEP redesign policy requires the development of a carefully planned transition to enable “the main National Assessment” to become the primary way to measure trends in reading, writing, mathematics and science. This is because there are now two NAEP testing programs for reading, writing, mathematics and science. The two programs use different tests, draw different samples of students (i.e., one based on age—9, 13 and 17-year-olds, the other based on grade—4, 8 and 12), and report results in two different ways. Not surprisingly, the two different programs can yield different results. Having different results is not necessarily a problem, however, for it can raise important

questions about student achievement and prompt helpful discussions about NAEP results. However, a trade-off is that having two separate testing programs boosts costs, potentially limiting assessments in other subjects.

The first program, referred to as the “long-term trend assessments,” monitors change in student performance using tests developed during the 1960’s and 1970’s. The sample of students is based on age (i.e., 9, 13, and 17-year olds) for reading, mathematics, and science and on grade for writing (i.e., grades 4, 8 and 11). The age-based samples include students from two or more grades. For example, the 9-year-old sample has 3rd, 4th, and 5th grade students. Long-term trend assessment results are reported displaying changes over time in average scores. The NAEP long-term trend data is the only measure we have in the United States to gauge how well current students are performing compared to students who took the same tests in the 1970’s, 1980s, and the first half of this decade. The second program, referred to as “main NAEP,” uses tests developed more recently, reports results by grade, and employs performance standards for reporting whether achievement is good enough. As an example of the value of maintaining two separate programs, in 1996 the long-term trend assessment program declared mathematics results flat since 1990, while main NAEP reported significant gains.

Status of Implementation:

The Governing Board initiated the review of the long-term trend assessments in part because significant cost savings could be achieved if main NAEP is able to assume the purpose of the long-term trend assessments. This review continues as a “work in progress.” The National Center for Education Statistics is just beginning to develop options. Identifying options that are practical, affordable, and technically feasible is a significant challenge. The Governing Board has scheduled long-term trend assessments to be conducted in 1999, 2003, and 2007. This will afford adequate time to evaluate the viability of the options that may be proposed and at the same time maintain the long-term trend line. The immediate effect is to change the schedule for this part of the testing program from once every two years to once every four years.

Keep NAEP Assessment Frameworks Stable

The NAEP redesign policy states that assessment frameworks shall remain stable for at least ten years. The purpose is three-fold: to provide for measuring trends in student achievement, to allow for change to frameworks when the case for change is compelling, and to manage costs.

By law, National Assessment frameworks are developed by the Governing Board through a national consensus process involving hundreds of teachers, curriculum experts, state and local testing specialists, administrators, and members of the public. The assessment frameworks describe how an assessment will be constructed, provide for the subject area content to be covered, determine what will be reported, and influence the cost of an assessment.

Both current practice and important developments in each subject area are considered: How much algebra should be in the 8th grade mathematics assessment? Should there be both multiple choice and constructed response items and if so, what is the appropriate mix? How much of what is measured should students know and be able to do? The frameworks receive wide public review before adoption by the Governing Board.

Status of Implementation:

The Governing Board is solely responsible for developing and approving assessment frameworks and has been adhering to its policy of keeping the frameworks stable. With a decision to be made this year about whether to conduct a national consensus process for the 2004 mathematics assessment, the Governing Board is beginning to examine criteria for determining when a new framework is necessary. An important factor will be the impact of changing the framework on the measurement of trends in student achievement.

Use International Comparisons

The NAEP redesign policy states that National Assessment frameworks, test specifications, achievement levels, and data interpretations shall take into account, where feasible, curricula, standards, and student performance in other nations, and promote studies to “link” the National Assessment with international assessments.

The National Assessment is, and should be, an assessment of student achievement in the United States. It should be focused on subjects and content deemed important for the U.S. through the national consensus process used to develop NAEP frameworks. However, decisions on content, achievement levels, and interpretation of NAEP results, where feasible, should be informed, in part, by the expectations for education set by other industrialized countries, and comparative test results. Although there are technical hurdles to overcome, consideration of such information can be useful in determining “how good is good enough” in an assessment for U.S. students.

Status of Implementation:

The National Center for Education Statistics conducted a linking study of the 1996 NAEP science and mathematics assessments with the 1995 Third International Mathematics and Science Study (TIMSS). The Governing Board used information from this linking study in setting the achievement levels for the 1996 science assessment. NCES will be conducting TIMSS again in the spring of 1999 and thirteen states have agreed to participate to collect state-representative TIMSS data. NCES will be applying a methodology for relating TIMSS to NAEP and will be evaluating the strength of the relationship.

Use Innovations in Measurement and Reporting

The NAEP redesign policy states that the National Assessment shall assess, and, where warranted, implement advances related to technology and the measurement and reporting of student achievement. In addition, the competition for NAEP contracts for assessments beginning around the year 2000 shall include a plan for conducting testing by computer in at least one subject and grade and for using technology to improve test administration, measurement, and reporting.

Status of Implementation:

The newly awarded NAEP contracts include plans for a short-form test (described above) in 4th grade mathematics in the year 2000 and for the development of a computer-based assessment.

Help States and Others Link to NAEP and Use NAEP Data to Improve Education Performance

The NAEP redesign policy states that the National Assessment shall assist states, districts and others, who want to do so at their own cost, to link their test results to the National Assessment. The policy also provides that NAEP shall be designed to permit access and use by others of NAEP data and materials. These include frameworks, specifications, scoring guides, results, questions, achievement levels, and background data. In addition, the policy provides that steps be taken to protect the integrity of the NAEP program and the privacy of individual test takers.

Status of Implementation:

The state of Maryland and the state of North Carolina have collaborated with the Governing Board on studies to examine the content of their respective state mathematics tests in light of the content of NAEP. The National Center for Education Statistics has a special grants program that provides funds to analyze NAEP data. The NCES has amended priorities for this grants program to encourage applications from states (and others) to conduct analyses that will be of practical benefit in interpreting NAEP results and in improving education performance. The National Academy of Sciences report "Uncommon Measures," describes the many technical difficulties involved in linking state results to NAEP. The NCES is planning a major conference with the states to provide a forum for discussing and addressing these difficulties. In addition, NCES is planning to conduct studies on various linking methodologies to provide insight on how the linking of NAEP and state assessments may best be done.

National Assessment Redesign: Implications for Reauthorization

The Governing Board's redesign policy is directed at the operation of the National Assessment program. It does not address governance of the National Assessment. While there are a number of areas in the current NAEP legislation for which change should be

considered, the NAEP redesign policy can, with two exceptions, be implemented within the current NAEP legislation.

The first exception has to do with the subjects to assess. Current law ties the subjects covered by NAEP to reading, writing, and the other subjects listed in the national education goals. The Governing Board agrees that these subjects should be assessed by the National Assessment and, accordingly, has adopted the schedule displayed in Table 1 above. The Governing Board recommends that, with respect to subjects to assess, the reauthorization of the National Assessment should be consistent with the schedule of assessments adopted by the Governing Board.

The second issue has to do with long-term trend assessments. Current law requires that assessments using age-based samples be conducted at least once every two years. Since the only assessments using age-based samples are the reading, science and mathematics long-term trend assessments, this provision is interpreted as requiring long-term trend assessments once every two years. In accordance with the schedule of assessments, the Governing Board recommends that the NAEP legislation be modified so that the frequency of the long-term trend assessments is changed to at least once every four years.

Conclusion

The National Assessment in the next century will provide student achievement results at the national level each year. State-level data will be provided every other year. Student achievement in reading, writing, mathematics and science will, appropriately, receive the most attention, with testing once every four years, but not to the exclusion of other important subjects. By continuing to report results using achievement levels and improving the process by which achievement levels are set, the National Assessment will help advance standards-based assessment and reporting in the United States. With a focus on its core purpose—measuring and reporting on the status of student achievement and change over time—the National Assessment design can be made more streamlined, more effective, and more efficient. With a clear sense of its primary audience—the general public—National Assessment reports will have more impact.

With a predictable schedule of assessments and reporting of National Assessment results, the public at regular intervals will discuss and debate education quality, states can plan ahead for their participation, and educators will have an external standard against which to compare their own efforts.

Appendix A



National Assessment Governing Board

National Assessment of Educational Progress

Policy Statement

on

Redesigning

The National Assessment of Educational Progress

Adopted Unanimously by
The National Assessment Governing Board, August 2, 1996

800 North Capitol Street, N.W.
Suite 825, Mailstop 7583
Washington, D.C. 20002-4233
Phone: (202) 357-6938
Fax: (202) 357-6945

Redesigning the National Assessment of Educational Progress

A Better Way to Measure Educational Progress in America

An effective democracy and a strong economy require well-educated citizens. A good education lays a foundation for getting a good job, leading a fulfilling life, and participating constructively in society.

But is the education provided in your state and in America good enough? How do our 12th graders compare with students in other nations in mathematics and science? Do our 8th grade students have an adequate understanding of the workings of our constitutional democracy? How well do our 4th grade students read, write, and compute? The National Assessment of Educational Progress is the only way for the public to know with accuracy how American students are achieving nationally and state-by-state.

The National Assessment tests at grades 4, 8, and 12. By law, it covers ten subjects, including reading, writing, math, and science. The National Assessment has performance standards that indicate whether student achievement is "good enough." The National Assessment is not a national exam taken by all students. In fact, only several thousand students are tested per grade, comprising carefully drawn samples that represent the nation and the participating states. Since its first test in 1969, the National Assessment has earned a trusted reputation for its quality and credibility. That reputation must be maintained.

The National Assessment is unique because of its national, state-by-state, and 12th grade results. State and local test results cannot be used to provide a national picture of student achievement. States and local schools use different tests that vary in many ways. The results cannot simply be "added up" to get a national score nor can state scores on their different tests be compared. The National Assessment Governing Board believes that twelfth grade achievement is important to monitor at the national level, because the 12th grade marks the end of elementary and secondary education, the transition point for most students from school to work, to college, or to technical training. The National Assessment is the only source of nationally representative data at the 12th grade. College entrance tests such as the ACT and the SAT are taken only by students planning on higher education; the results do not represent the achievement of the total 12th grade class. And to date, virtually no state-based assessment program tests 12th graders.

While there is much about the National Assessment that is working well, there is a problem. Under its current design, the National Assessment tests too few subjects, too infrequently, and reports achievement results too late—as much as 18 to 24 months after testing. Testing occurs every other year. During the 1990's, only reading and mathematics will be tested more than once using up-to-date tests and performance standards. Six subjects will be tested only once and two subjects not at all during the 1990's.

Why is the National Assessment testing so few subjects and fewer subjects now than years ago? Over the years, the National Assessment has become increasingly complex. Its quality and integrity have led to a multitude of demands and expectations beyond its central purpose. Meeting those expectations was done with good intentions and seemed right for the situation at the time. However, additions to the National Assessment have been "tacked on" without changing the basic design, reducing the number of subjects that can be tested and driving up costs.

For example, where a single 120 page mathematics report once sufficed, mathematics reporting in 1992 consisted of seven volumes totalling almost 1,800 pages, not including individual state reports. Also, there are now two separate testing programs for reading, writing, math, and science. One monitors trends using tests developed during the 1970's; the other reflects current views on instruction and uses performance standards to report whether achievement is good enough.

The current National Assessment design is overburdened, inefficient, and redundant. It is unable to provide the frequent, timely reports on student achievement the American public needs. The challenge is to supply more information, more quickly, with the funding available.

To meet this challenge, the National Assessment design must be changed, building on its strengths while making it more efficient. The design of the National Assessment must be simplified. The purpose of the National Assessment must be sharply focused and its principal audience clearly defined. Because the National Assessment cannot do all that some would have it do, trade-offs must be made among desirable activities. Useful but less important activities may have to be reduced, eliminated, or carried out by others. The National Assessment must "stick to its knitting" in order to be more cost-effective, reach more of the public, provide more information more promptly, and maintain its integrity.

National Assessment Redesign

To provide the American public with more frequent information in more subjects about the progress of student achievement, changes must be made in the way that the National Assessment is designed and the results are reported. These changes are described in this policy statement. Undergirding these changes is an explicit statement of the purposes, objectives, audiences, and limitations of the National Assessment.

While change is in order, many current policies should continue. For example, reliability, validity, and quality of data will remain hallmarks of the National Assessment. The sample of tested students will be as representative as possible, using policies and procedures that maximize the number of students included who are disabled or are of limited English proficiency. And reporting on trends over time will remain a central commitment of the National Assessment.

The intent of this policy statement is to guide current operations of the National Assessment, the development of new requests for proposals for contracts for conducting the National Assessment and the activities and structure of the National Assessment Governing Board. Contracts for current operations extend through assessments to be conducted in 1998. New contracts would cover assessments as early as 1999 and thereafter.

Purpose and Objectives of the National Assessment of Educational Progress

The purpose of the National Assessment is stated in its legislation:

to provide a fair and accurate presentation of educational achievement in reading, writing, and the other subjects included in the third National Education Goal, regarding student achievement and citizenship.

Thus, the central concern of the National Assessment is to inform the nation on the status of student achievement. The National Assessment Governing Board believes that this should be accomplished through the following objectives:

- (1) to measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations;
- (2) to develop, through a national consensus, sound assessments to measure what students know and can do as well what students should know and be able to do; and
- (3) to help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance.

The specific changes in the design of the National Assessment described below are discussed in relation to these objectives.

The Audience for the National Assessment

The primary audience for National Assessment results is the American public, including the general public in states that receive their own results from the National Assessment. Reports should be written for this audience. Results should be released within 6 months of testing. Reports should be understandable, jargon free, easy to use, and widely disseminated. Although more comprehensible, direct, and useful, the reports will not trade accuracy for simplicity. The tradition of high quality of National Assessment reports will be continued, with no erosion of validity and reliability. Assessment questions and samples of student work that illustrate performance standards are likely to receive heightened prominence in reports.

Principal users of National Assessment data are national and state policymakers and educators concerned with student achievement, curricula, testing, and standards. National Assessment data will be available to these users in forms that support their efforts to interpret results to the public, to improve education performance, and to perform secondary analysis.

Limitations: What the National Assessment Is Not

The National Assessment is intended to describe how well students are performing, but not to explain why. The National Assessment only provides group results; it is not an individual student test. The National Assessment tests academic subjects and does not collect information on individual students' personal values or attitudes. Each National Assessment test is developed through a national consensus process. This national consensus process takes into account education practices, the results of education research, and changes in the curricula. However, the National Assessment is independent of any particular curriculum and does not promote specific ideas, ideologies, or teaching techniques. Nor is the National Assessment an appropriate means, by itself, for improving instruction in individual classrooms, evaluating the effects of specific teaching practices, or determining whether particular approaches to curricula are working.

OBJECTIVE 1: To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations.

Assess all subjects specified by Congress: reading, writing, mathematics, science, history, geography, civics, the arts, foreign language, and economics

The gap must be closed between the number of subjects the National Assessment is required to assess and the number of subjects it can assess at the national level under the current design. By law, the National Assessment is required to assess ten subjects and report results and trends. In order to chart progress and report trends, subjects must be assessed more than once. However, during the 1990's only reading and mathematics will have been assessed more than once using up-to-date tests and performance standards to report how well students are doing.

Some have suggested that a solution is to combine into a single assessment several related subjects (e.g. reading and writing and/or history, geography, civics, and economics). Under such an approach, assessment data would be reported using both an overall score and

subscores for the respective disciplines. Although such an approach has the appeal of reducing the number of separate assessments, its feasibility, desirability, and costs are unknown. Also, such an approach has far-reaching implications for the test frameworks that guide the development of each assessment and for reporting results. These implications must be considered carefully. For the immediate future, subjects will continue to be assessed separately. However, the National Assessment Governing Board is committed to providing the public with more information as efficiently as possible. The Governing Board will consult with technical experts and education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.

- **The National Assessment shall be conducted annually, two or three subjects per year, in order to cover all required subjects at least twice a decade.**
- **The National Assessment shall assess all subjects listed in the third National Educational Goal—reading, writing, mathematics, science, history, geography, civics, the arts, foreign language and economics—according to a publicly released schedule adopted by the National Assessment Governing Board, covering eight to ten years, with reading, writing, mathematics, and science tested more frequently than the other subjects.**
- **The National Assessment Governing Board shall consult with technical experts and with education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.**

Provide National Assessment results for states

In 1988, testing at the state level was added to the National Assessment as a trial, with participation strictly voluntary, subjects and grades specified in law, and an independent evaluation required. Previously, the National Assessment had reported only national and regional results. For the first time, the information was relevant to individuals in states who make decisions about education funding, governance, and policy. As a result, states now are major users of National Assessment data.

Participation was strong in the first state-level assessment in 1990 and has grown to include even more states. In 1996, 44 states and 3 jurisdictions participated in the math assessments at grade 4 and 8 and the science assessment at grade 8. The independent evaluation concluded that the trial state assessments produced valid and reliable data. The evaluation report recommended, and Congress agreed, that state-level assessments, with continued evaluations, be included in the 1994 reauthorization of the National Assessment.

Currently, the National Assessment draws a separate sample to obtain national results in addition to the samples drawn for individual state reports. Keeping the schools drawn for national samples completely partitioned from the state samples increases costs and creates additional burdens on states, particularly small states. Options should be identified for making the national and state samples more efficient and less burdensome. For example, it may be possible to reduce the current state sample size of 100 schools to a smaller number (e.g. 65-75) without a great loss in precision.

States participate in the National Assessment for many reasons, including to have an unbiased, external benchmark to help them make judgments about their own tests and standards. National Assessment data are used to make comparisons to other states, to help determine if curriculum and standards are rigorous enough, to develop questions about curricular strengths and weaknesses, to make state to international comparisons, and to provide a general indicator of achievement.

There is a strong interest among states to participate in the National Assessment to get state level information at grades 4 and 8 in reading, writing, mathematics, and science. The level of interest in participating in the National Assessment varies with respect to the other subjects (i.e., history, geography, civics, economics, the arts, and foreign language) and at grade 12, where state officials say that obtaining cooperation from high schools and 12th grade students is difficult.

Some states, however, would like to be able to use National Assessment tests in the other subjects and at grade 12. Such use of National Assessment tests would be conducted as a service, with the reporting of results and maintenance of data under the control of the state. States will be able to use National Assessment tests if they adhere to requirements to protect the integrity of the National Assessment program and pay the additional costs. At the present time, states that participate in the National Assessment to get state level information at grades 4 and 8 in reading, writing, mathematics, and science provide in-kind support to cover the cost of in-state coordination and test administration. The National Assessment program covers the majority of costs, including test development, sampling, analysis and reporting. States that wish to use National Assessment tests in other subjects and at grade 12 would pay for much of these additional costs.

States are active partners in the National Assessment program. States help develop National Assessment test frameworks, review test items, and assist in conducting the tests. The National Assessment program is effective, to a great degree, because of the involvement of the states.

Because it is useful to them, and because they invest time and resources in it, states want a dependable schedule for National Assessment testing. With a dependable schedule, states that want to will be better able to coordinate the National Assessment with their own state testing program and make better use of the National Assessment as an external reference point.

- **National Assessment state-level assessments shall be conducted on a reliable, predictable schedule according to an eight to ten year plan adopted by the National Assessment Governing Board.**
- **Reading, writing, mathematics, and science at grades 4 and 8 shall be given priority for National Assessment state-level assessments.**
- **States shall have the option to use National Assessment tests in other subjects and at grade 12 by assuming a larger share of the costs and adhering to requirements that protect the integrity of the National Assessment program. However, the National Assessment Governing Board shall seek ways to make such use of National Assessment tests attractive and financially feasible.**
- **Where possible, changes in national and state sampling procedures shall be made that will reduce burden on states, increase efficiency, and save costs.**

Vary the amount of detail in testing and in reporting results

More subjects can be assessed if different strategies are used. Currently, each time the National Assessment is conducted, it uses a similar approach, regardless of the nature of the subject or the number of times an assessment in a subject has been administered. This approach is locked-in through 1998 under current contracts. Under this approach, a larger number of students is tested in order to provide not just overall results, but fine-grained details as well (e.g. the achievement scores of 4th grade students whose teachers that year had five hours or more of in-service training). The National Assessment also collects "background" information through questionnaires completed by students, teachers, and principals. The questionnaires ask about teaching practices, school policies, and television watching, to name a few. Data analyses are elaborate. Reports are detailed and exhaustive, involving as many as seven separate reports per subject. Although the National Assessment has been praised for this thoroughness, the cost of this thoroughness is that fewer subjects are assessed, assessments occur less frequently, and reports take longer to produce.

The different strategies needed might include several approaches to testing and reporting, all of which should be designed in ways that maintain the National Assessment's commitment to providing valid and reliable data of high quality. For example, these approaches could take the form of "standard report cards," "comprehensive reports," and special, focused assessments.

A standard report card would provide overall results in a subject with performance standards and average scores. Results for standard report cards could be reported by sex, race/ethnicity, socio-economic status, and for public and private schools, but would not be broken down further. This may reduce the number of students needed for testing and may

reduce associated costs. Generally, subcategories within a subject (e.g. algebra, measurement, and geometry within mathematics) would not be reported. However, data from the National Assessment would continue to be available to state and local educators and policymakers for additional analysis.

Comprehensive reports, like the current approach, would be an in-depth look at a subject, perhaps using a newly adopted test framework, many students, many test questions, and ample background information. In addition to overall results using performance standards and average scores, subcategories within a subject could be reported. Results would be reported by sex, race/ethnicity, socio-economic status, and for public and private schools, and might be broken down further as well. In some cases, more than one report may be issued in a subject. Comprehensive reporting in a particular subject would occur infrequently, perhaps once in ten years, but under a planned schedule of assessments.

Special, focused assessments on timely topics also would be conducted. They would explore a particular question or issue and may be limited to particular grades. Generally, the cost would be less than the cost of a standard report card. Examples of these smaller-scale, focused assessments include: (1) assessing subjects using targeted approaches (e.g. 8th grade arts), (2) testing special populations (e.g. in-school 12th graders versus out-of-school youth), and (3) examining skills and knowledge across several subjects (e.g. readiness for work).

The use of background surveys also would be varied. The three kinds of background surveys—student, teacher and principal questionnaires—would not necessarily all be employed each time a subject is assessed. Instead, the use of such surveys would be limited and selective, with reports of results focused on a core of background questions addressing the most essential issues. Also, background surveys used for standard report cards in a particular year would be designed to complement, rather than duplicate, background surveys used for comprehensive reports in the same year.

- **National Assessment testing and reporting shall vary, using standard report cards most frequently, comprehensive reporting in selected subjects about once every ten years, and special, focused assessments.**
- **National Assessment results shall be timely, with the goal being to release results within 6 months of the completion of testing for standard report cards and within 9 months for comprehensive reports.**

Simplify the National Assessment design

The current design of the National Assessment is very complex and, in fact, has grown more complex over the years. Here are just three examples of this complexity. (1) No student

takes the complete set of test questions in a subject and as many as twenty-six different test booklets are used within each grade. Scores are calculated using sophisticated statistical procedures. (2) Students, teachers, and principals complete separate background questionnaires and may submit them for scoring at different times. Data from the questionnaires are used in calculating results of the assessments. (3) Current requirements for data analysis demand that test scores be calculated for every background variable collected by the National Assessment before any report can be produced. This lengthens the time from data collection to reporting and adds significantly to cost.

The design became more complex, in part, because the National Assessment's purposes and audiences had proliferated and the amount of background information collected had expanded. Specifying the purposes, audiences, and limitations of the National Assessment, as well as providing for varied means for testing and reporting, will result in opportunities for simplifying the National Assessment design.

- **Options shall be identified to simplify the design of the National Assessment.**

Simplify the way the National Assessment reports trends in student achievement

From its beginning in 1969, monitoring achievement trends has been a central mission of the National Assessment of Educational Progress. Monitoring long-term trends in educational achievement, both for the population as a whole and for significant sub-groups, is a capacity unique to the National Assessment and should be continued as a central mission. However, as the National Assessment approaches its third decade, it must address the problem of how to assess trends in achievement when curricula continue to evolve and change. An assessment in a subject must be kept stable to monitor trends. However, stable assessments may not reflect important changes in curricula. Over time, there develops a legitimate concern about the relevance of the content of the assessment versus the ability to track change in achievement.

As a solution to this problem, since 1990, the National Assessment has reported achievement trends using two unconnected assessment programs. The tests, criteria for selecting students, and reporting are all different. The first program, "the main National Assessment," tests at grades 4, 8, and 12 and covers ten subjects. The assessments are based on a national consensus representing current views of each subject. Performance standards are used to report whether student achievement on the National Assessment is "good enough." The schedule of subjects to be assessed in the main National Assessment is unrelated to the schedule of subjects under the second testing program.

The second assessment program reports long-term trends that go as far back as 1970. Only four subjects are covered: reading, writing, mathematics, and science. The assessments are based on views of the curricula prevalent during the 1970's and have not been changed. Testing is at ages 9, 13, and 17 except for writing, which tests at grades 4, 8, and 11. Trends

are reported by average score; performance standards are not used. The long-term trend program has been valuable for documenting declines and increases in student achievement over time and a decrease in the achievement gap between minority and non-minority students.

It may be impractical and unnecessary to operate two separate assessment programs. However, it also is likely that curricula will continue to change and that current test frameworks may be less relevant in the future. The tension between the need for stable measures of student achievement and changing curricula should be recognized as a continuing policy matter for the National Assessment, requiring efficient and balanced design solutions. Among the factors to consider are: (1) setting a standard period of time for a long-term trend (e.g. 15-20 years) using a particular "metric" in a subject; (2) providing for overlapping administrations of old and new assessments and "bridge" studies to determine whether the new can be linked to the old assessment; and (3) periodic administration of older assessments (e.g. once every ten years once a new trend-line has been established so that it would be possible to compare performance in 2010 with that in 1970 on the old trend line and with that in 1990 on a new trend line).

- **A carefully planned transition shall be developed to enable "the main National Assessment," to become the primary way to measure trends in reading, writing, mathematics, and science in the National Assessment program.**

Use performance standards to report whether student achievement is "good enough"

In reporting on "educational progress," the National Assessment has, until recently, only considered current student performance compared to student achievement in previous years. Under this approach, the only standard was how well students had done previously, not how well they should be doing on what is measured by the National Assessment. Although this approach has been useful, it began to change in 1988 from a sole focus on "where we have been" to include "where we want to be" as well.

In 1988, Congress created a non-partisan citizen's group—the National Assessment Governing Board—and authorized it to set explicit performance standards, called achievement levels, for reporting National Assessment results.

The achievement levels describe "how good is good enough" on the various tests that make up the National Assessment. Previously, it might have been reported that the average math score of 4th graders went up (or down) four points on a five-hundred-point scale. There was no way of knowing whether the previous score represented strong or weak performance and whether the amount of change should give cause for concern or celebration. In contrast, the National Assessment now also reports the percentage of students who are performing at or above "basic," "proficient," and "advanced" levels of achievement. Proficient, the central

level, represents "competency over challenging subject matter," as demonstrated by how well students perform on the questions on each National Assessment test. Basic denotes partial mastery and advanced signifies superior performance on the National Assessment. Using achievement levels to report results and track changes allows readers to make judgments about whether performance is adequate, whether "progress" is sufficient, and how the National Assessment standards and results compare to those of other tests, such as state and local tests.

First employed in 1990, the achievement levels have been the subject of several independent evaluations and some controversy. Information from these evaluations, as well as from other experts, has been used over the last six years to improve and refine the procedures by which achievement levels are set. Although the current procedures may be among the most comprehensive and sophisticated standard-setting procedures used in education, the Governing Board remains committed to improving the process and to the continuing conduct of validity studies.

- **The National Assessment shall continue to report student achievement results based on performance standards.**

Use international comparisons

Looking at student performance and curriculum expectations in other nations is yet another way to consider the adequacy of U.S. student performance. The National Assessment is, and should be, a domestic assessment. However, decisions on the content of National Assessment tests, the achievement standards, and the interpretation of test results, where feasible, should be informed, in part, by the expectations for education set by other countries, such as Japan, Germany, and England. Although there are technical hurdles to overcome, consideration of such qualitative information can be used to good effect. In addition, the National Assessment should promote "linking" studies with international assessments, as has been done with the Third International Mathematics and Science Study, so that states that participate in the National Assessment can have state, national, and international comparisons. This, in turn, should take into account problems in making international comparisons truly comparable, such as differences in the samples of students tested, differences in the curricula, and differences in the translated test questions.

- **National Assessment test frameworks, test specifications, achievement levels, and data interpretations shall take into account, where feasible, curricula, standards, and student performance in other nations.**
- **The National Assessment shall promote "linking" studies with international assessments.**

Emphasize reporting for grades 4, 8, and 12

An aspect of the National Assessment design that needs reconsideration is age versus grade-based reporting. At its inception, the National Assessment tested only by age. Current law requires testing both by age (ages 9, 13, and 17) and by grade (grades 4, 8, and 12). Grade-based results are generally more useful than age-based results. Schools and curricula are organized by grade, not by age. Grades 4, 8, and 12 mark key transition points in American education. Grade 12 performance is particularly important as an "exit" measure from the K-12 education system. Grades 4, 8, and 12 are specified for monitoring in National Education Goal 3. Age-based samples may be more appropriate with respect to international comparisons and, given high school drop-out rates, would be more inclusive for age 17 than for grade 12 samples, which are limited to youth enrolled in school. However, assessing the knowledge and skills of out-of-school youth may properly fall under the purpose of another program, such as the National Adult Literacy Survey.

Although grade-based reporting is generally preferable, there is a problem about the accuracy of grade 12 National Assessment results. At grade 12, a smaller percentage of schools and students that are invited actually participate in testing than is the case with 4th and 8th graders. Also, more 12th graders fail to complete their tests than do 4th and 8th graders. In addition, when asked "How hard did you try on this test?" and "How important is doing well on this test?" many more 12th graders, than 4th or 8th graders, say that they didn't try hard and that the test wasn't important. Low participation rates, low completion rates, and indicators of low motivation suggest that the National Assessment may be underestimating what 12th graders know and can do.

One possible reason for low response and low motivation is that schools and students receive very little in return for their participation in the National Assessment beyond the knowledge that they are performing a public service. They do not receive test scores nor do they receive other information from the National Assessment that teachers and principals might wish to use as a part of the instructional program. This should be changed. The National Assessment design should use meaningful, practical incentives that will give school principals and teachers a greater reason to participate and students more of a reason to try harder. The underlying idea is clear: if principals and teachers see direct benefits, they are more likely to agree to participate in the National Assessment. Students may be more likely to take the assessment seriously if they see that their teachers and principals are enthusiastic about participating. Without practical incentives, even at grades 4 and 8, the willingness of district and school administrators and staff to participate in the National Assessment may diminish over time.

- **The National Assessment shall continue to test in and report results for grades 4, 8, and 12; however, in selected subjects, one or more of these grades may not be tested.**

- Age-based testing and reporting shall be permitted when deemed appropriate and when necessary for international comparisons and for long-term trends, should the National Assessment Governing Board decide to continue long-term trends in their current form.
- Grade 12 results shall be accompanied by clear, highlighted statements about school and student participation, student motivation, and cautions, where appropriate, about interpreting 12th grade achievement results.
- The National Assessment design shall seek to improve school and student participation rates and student motivation at grade 12.
- The National Assessment shall provide practical incentives for school and district participation at grades 4, 8, and 12.

Use innovations in measurement and reporting

The National Assessment has a record of innovations in large-scale testing. These include the early use of performance items, sampling both students and test questions, using standards describing what students should know and be able to do, and employing computers for such things as inventory control, scoring, data analysis, and reporting. The National Assessment should continue to incorporate promising innovative approaches to test administration and improved methods for measuring and reporting student achievement.

Technology can help improve National Assessment reporting and testing. For example, reports could be put on computer disc, transmitted electronically, and made available on the World Wide Web. Test questions could be catalogued and made available on-line for use by state assessment personnel and classroom teachers. Also, the National Assessment could be administered by computer, eliminating the need for costly test booklet systems and reducing steps related to data entry of student responses. Students could answer "performance items" in cost-effective, computerized formats. The increasing use of computers in schools may make it feasible to administer some parts of the National Assessment by computer under the next contract for the National Assessment, beginning around the year 2000.

Other examples of promising methods for measuring and reporting student achievement include adaptive testing and domain-score reporting. In adaptive testing, each student is given a short "pre-test" to estimate that student's level of achievement. Students are then administered test exercises that are in the range of difficulty indicated by the pre-test. Since the test is "adapted" to the individual, it is more precise and can be markedly more efficient than regular test administration. In domain-score reporting, a subject (or "domain") is well-defined, a goodly number of test questions are developed that encompass the subject, and student results are reported as a percentage of the "domain" that students "know and can

do." This is in contrast to reporting results using an arbitrary scale, such as the 0-500 scale used in the National Assessment.

- **The National Assessment shall assess the merits of advances related to technology and the measurement and reporting of student achievement.**
- **Where warranted, the National Assessment shall implement such advances in order to reduce costs and/or improve test administration, measurement, and reporting.**
- **The next competition for National Assessment contracts, for assessments beginning around the year 2000, shall ask bidders to provide a plan for (1) conducting testing by computer in at least one subject at one grade, and (2) making use of technology to improve test administration, measurement, and reporting.**

OBJECTIVE 2: To develop, through a national consensus, sound assessments to measure what students know and can do as well as what students should know and be able to do.

Keep test frameworks and specifications stable

Test frameworks spell out in general terms how an assessment will be put together. The frameworks also determine what will be reported and influence how expensive an assessment will be. Should 8th grade mathematics include algebra questions? Should there be both multiple choice questions and questions in which students show their work? What is the best mix of such types of questions for each grade? Which grades are appropriate for assessment in a subject area? Test specifications provide detailed instructions to the test writers about the specific content to be tested at each grade, how test questions will be scored, and the format for each test question (e.g. multiple choice, essay, etc.).

Since 1989, the National Assessment Governing Board has conducted a national consensus process to develop new test frameworks and specifications. The national consensus process involves hundreds of teachers, curriculum experts, directors of state and local testing programs, administrators, and members of the public. The national consensus process helps determine what is important for the National Assessment to test, how it should be measured, and how much of what is measured by the National Assessment students should know and be able to do in each subject.

Through the national consensus process, both current classroom teaching practices and important developments in each subject area are considered for inclusion in the National Assessment. In order to ensure that National Assessment data fairly represent student achievement, the test frameworks and specifications are subjected to wide public review before adoption and test questions developed for the National Assessment are reviewed for relevance and quality by representatives from participating states.

An important role of the National Assessment is to report on trends in student achievement over time. For the National Assessment to be able to measure trends, the frameworks (and hence the tests) must remain stable. However, as new knowledge is gained in subject areas and as teaching practices change and evolve, pressures arise to change the test frameworks and tests to keep them current. But, if frameworks, specifications, and tests change too frequently, trends may be lost, costs go up, and reporting time may increase.

- **Test frameworks and test specifications developed for the National Assessment generally shall remain stable for at least ten years.**
- **To ensure that trend results can be reported, the pool of test questions developed in each subject for the National Assessment shall provide a stable measure of student performance for at least ten years.**
- **In rare circumstances, such as where significant changes in curricula have occurred, the National Assessment Governing Board may consider making changes to test frameworks and specifications before ten years have elapsed.**
- **In developing new test frameworks and specifications, or in making major alterations to approved frameworks and specifications, the cost of the resulting assessment shall be estimated. The National Assessment Governing Board will consider the effect of that cost on the ability to test other subjects before approving a proposed test framework and/or specifications.**

Use an appropriate mix of multiple-choice and "performance" questions

To provide information about "what students know and can do," the National Assessment uses both multiple-choice questions and questions in which students are asked to produce their own answers, such as writing a response to an essay question or explaining how they solved a math problem. Questions of the latter type are sometimes called "performance items." Both types of questions can vary in difficulty and the richness of information they provide, and may require students to demonstrate different kinds of skills and knowledge.

Performance items are desired because they provide direct evidence of what students can do. They range in length of test taking time from a short-answer or fill-in-the-blank format requiring about a minute of response time, to items requiring about 5 minutes of response time, to writing exercises that may allow 15 to 50 minutes response time. Although they may be desirable, performance items are more expensive than multiple-choice to develop, administer, and score. In addition, much larger proportions of students fail to respond to performance items, particularly as the amount of required response time increases.

Multiple-choice questions can be challenging and are desired because they are efficient in collecting information about student knowledge. However, multiple-choice questions are more subject to guessing than are performance items.

Currently, all students tested by the National Assessment are given both types of questions. Generally, about half the testing time is devoted to each type of question, but the amount of time for each differs based on the skills and knowledge to be assessed, as established in the National Assessment test frameworks. For example, in a writing assessment, all students are asked to write their responses to specific exercises. In other subjects, the mix of multiple-choice and performance items varies. The appropriate mix of items for each subject should be determined by the nature of the subject, the range of skills to be assessed, and cost.

- **Both multiple-choice and performance items shall continue to be used in the National Assessment;**
- **In developing new test frameworks, specifications, and questions, decisions about the appropriate mix of multiple-choice and performance items shall take into account the nature of the subject, the range of skills to be assessed, and cost.**

OBJECTIVE 3: To help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance.

The primary job of the National Assessment is to report frequently and promptly to the American public on student achievement. The resources of the National Assessment must be focused on this central purpose if it is to be achieved. However, the products of the National Assessment—test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data—are widely regarded as being of high quality. They are developed with public funds and, therefore, should be available for public use as long as such

uses do not threaten the integrity of the National Assessment or its ability to report regularly on student achievement.

The National Assessment should be designed in a way that permits its use by others while protecting the privacy of students, teachers, and principals who have participated in the National Assessment. This should include making National Assessment test questions and data easy to access and use, and providing related technical assistance upon request. Generally, the costs of a project should be borne by the individual or group making the proposal, not by the National Assessment.

Examples of areas in which particular interest has been expressed for using the National Assessment include linking state and local tests with the National Assessment and performing in-depth analysis on National Assessment data. States that link their tests to the National Assessment would have an unbiased external benchmark to help make judgments about their own tests and standards and also would have a means for comparing their tests and standards with those of other states.

- **The National Assessment shall develop policies, practices, and procedures that assist states, school districts, and others who want to do so at their own cost to link their test results to the National Assessment.**
- **The National Assessment shall be designed so that others may access and use National Assessment test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data.**
- **The National Assessment shall employ safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers.**