Improving Education Through Standards-Based Reform

A Report by the
National Academy of Education
Panel on Standards-Based Education Reform

Milbrey W. McLaughlin & Lorrie A. Shepard
with Jennifer A. O'Day

The preparation of this report was supported by grants from the Carnegie Corporation of New York and the John D. and Catherine T. MacArthur Foundation. The opinions expressed in the report do not necessarily reflect the position, policy, or endorsement of either the Carnegie Corporation or the MacArthur Foundation.

Copyright by The National Academy of Education, 1995
ISBN 0-942469-08-9

For additional copies contact:
The National Academy of Education
Stanford University
CERAS Building, Room 108
Stanford, CA 94305-3084
(415) 725-1003

The National Academy of Education

Panel on Standards-Based Education Reform

Co-chairs

Milbrey W. McLaughlin, Stanford University

Lorrie A. Shepard, University of Colorado, Boulder

Members

David K. Cohen, University of Michigan

Linda Darling-Hammond, Columbia University

Patricia Gándara, University of California, Davis

Edmund Gordon, Yale University and City University of New York/ City College of New York

Kenji Hakuta, Stanford University

Shirley Brice Heath, Stanford University

Carl F. Kaestle, University of Chicago, *ex officio*

Michael W. Kirst, Stanford University

Robert L. Linn, University of Colorado, Boulder

Andrew Porter, University of Wisconsin-Madison

Project Associate

Jennifer A. O'Day, Stanford University

Editor

Deborah Fife

Staff

Mona Neveu, Administrative Assistant

Debbie Leong-Childs, Executive Director

Table of Contents

Preface xi

Executive Summary xv

Part I
The National Context of Standards-Based Reform 1

Part II
The Vision and Challenges of Standards-Based Reform 7

Part III
Issues and Recommendations 19

Part IV
Conclusion 63

Glossary 69

Detailed Table of Contents

Preface xi

Executive Summary xv

Part I

THE NATIONAL CONTEXT OF STANDARDS-BASED REFORM 1

The 1989 Education Summit 1

Professionally Set Subject-Matter Standards 2

Linking Standards, Assessments, and Opportunity to Learn 3

Goals 2000: Educate America Act 4

The NAE Panel on Standards-Based Education Reform 5

An Overview of the Report 5

Part II

THE VISION AND CHALLENGES OF STANDARDS-BASED REFORM 7

The Vision of Standards-Based Reform 7

Challenging standards 8

Educational equity 10

Criticism of Standards-Based Reform 12

Diverting attention from more fundamental educational needs 12

Undermining professional and local responsibility for student learning 13

Heightening educational inequities 14

Threatening basic-skills instruction 14

An Unprecedented Task 15

Guidelines for Developing Standards-Based Strategies 16

Use existing research knowledge 17

Pay constant attention to equity issues 17

Foster a shared sense of responsibility 17

Ongoing evaluation and improvement of the system 18

Part III

ISSUES AND RECOMMENDATIONS 19
Content Standards 19
Definition 19
Figure 1. A sample from CA 1985 Model Curriculum Guide 21
Uses of content standards 21
Framing the conversation 21
Guiding curriculum and instruction 22
Allocating instructional resources 23
Key issues 23
One standard or multiple standards? 23
Who should define the standards? 25
Professional and public involvement 25
Level of system 28
How to design standards as tools for instructional improvement 28
Level of specificity 29
Scope 29
Links to performance 29
Figure 2. Examples of student work give meaning to standards 30
Summary of recommendations for content standards 31
Performance Standards 32
Definition 32
Figure 3. A sample performance standard from the Australian curriculum for English 33
Uses of performance standards 34
Exemplification of content standards 34
Figure 4. A sample task and high-quality response for eighth-grade mathematics 35
Accountability and certification 36
Key issues 37
Accountability and certification requires greater field experience 37
The value of multiple benchmarks 39
The defensibility of rewards vs. punishments 40
Summary of recommendations for performance standards 41

Part III continued

Opportunity-to-Learn Standards 42
Definition 42
Uses of opportunity-to-learn standards 43
Provide vision 43
Guide school review 43
Guide system accountability 44
Key issues 45
Defining what really matters 45
Measuring OTL 46
Objections to OTL standards 47
Summary of recommendations for opportunity-to-learn standards 48
Assessment 49
Definition 49
Uses of assessment 49
System-level monitoring of educational achievement 49
System-level monitoring for accountability 50
Student certification 50
Instructional benefits 50
Compatible or contradictory purposes of assessment 51
Key issues 52
Aligning assessments with content standards 52
Providing evidence of validity, reliability, and fairness 53
Allowing for demonstration of proficience by multiple methods 54
Assessing limited-English-proficient (LEP) students 55
Figure 5. A sample task from a Spanish version of the CLAS Grade 4 Math Assessment 56
Figure 6. A sample task from a CLAS Grade 4 Math Assessment 57
Making assessments count while guarding against unfairness 58
Summary of recommendations for assessments 61

Part IV

CONCLUSION 63

Implementation Concerns 63

Getting Started with Caution 64

Building Capacity 66

Professional development 66

Research and evaluation 66

Communication and feedback 67

A national organization 67

Refrain: The Promise and Challenges of Standards-Based Reform 68

Glossary 69

PREFACE

In the past dozen years Americans have conducted a vigorous and important debate about education. They have investigated and bemoaned the quality, the structure, and the purposes of their schools. They have initiated many reforms at the local, state, and national levels. One of the animating ideas of this period of reform has been higher academic standards. Many elected officials, business-people, and educators concluded that we should try to reach consensus on national standards about what students should know and be able to do in each school subject. Toward that end the federal government, philanthropic foundations, and educational organizations launched efforts to develop standards that would combine more effective learning of traditional content with greater emphasis on deep understanding and problem-solving.

Other reformers focused on the state and local levels, where responsibility for curriculum and instruction has traditionally resided; the federal government supported activities at these levels as well, through the legislation called Goals 2000. Most states have begun efforts to improve academic standards or to reform education through innovative assessments, or both. Because national standards have proved controversial, and would in any case be voluntary, the most consequential actions on higher standards will probably occur at the state and local levels.

Regardless of who develops the academic standards, their implementation will be difficult. This is particularly so because the watchword for the standards-based reform movement of recent years is "high standards for *all* students." Those who have articulated this ideal have indeed set high standards for the movement itself. The effort to design and gain consensus around new, challenging academic stan-

dards—and to implement them in ways that include all students—commits reformers to a series of complicated policy problems and choices.

In view of these challenges, representatives of the Carnegie Corporation of New York, the MacArthur Foundation, and the Pew Charitable Trusts proposed in the spring of 1994 that The National Academy of Education establish a panel of education scholars to think about what's at stake in the standards movement and to write a report about the subject. The Academy is a group of about 135 scholars and noted practitioners in the field of education, founded in 1965 "to promote scholarly inquiry and discussion concerning the ends and means of education." Among its most important functions has been the sponsorship of study panels on various education issues, such as the funding and uses of education research, the evolution of the National Assessment of Educational Progress, and the teaching of reading. As in those earlier efforts, the present report relates research to an important issue of educational practice.

The Academy was fortunate that two superb scholars agreed to chair this Panel: Milbrey W. McLaughlin of Stanford University, one of the nation's leading education policy analysts, and Lorrie A. Shepard of the University of Colorado at Boulder, one of the top measurement experts in the country. Together we prepared a prospectus for the project and recruited the remaining members of the Panel. Most, but not all, are members of the National Academy of Education. We sought a productive mix of policy analysts, assessment experts, and subject-matter specialists. We also sought people with a wide range of views on standards issues. The group met three times during the past year, and members exchanged many ideas between meetings, under the energetic leadership of the co-chairs. The committee achieved consensus on several difficult issues; all members of the Panel endorse the report and its recommendations. The report has also been reviewed by the Executive Council of The National Academy of Education, sitting as a committee on reports, and they have approved its publication as a report of the Academy.

Among the readers of this report there may be some standards advocates who find cautionary messages and some skeptics who discover some redeeming features of the movement. The Panel's intention, however, is not to argue for or against standards but to define the issues that are embedded in the standards movement and to make recommendations about how people committed to this movement can proceed in ways that will be both equitable and challenging for students, teachers, and communities. Most readers will already appreciate the

complexity of this reform movement and its purposes, and we hope that they will find our discussion of the challenges useful.

On behalf of the Academy, I wish to express thanks to the members of the Panel, and to the several staff members who contributed to this effort, for their careful thought and hard work; in particular, Jennifer O'Day made signal contributions in conceptual work and drafting sections of the report. Fritz Mosher of the Carnegie Corporation coordinated the sponsors' efforts and participated as a colleague in every phase of our deliberations, to our benefit. Funds for the Panel's work came from the Carnegie Corporation of New York and from the John D. and Catherine T. MacArthur Foundation. Peter Gerber, education officer at the MacArthur Foundation, attended one of our sessions and provided helpful advice. Because of our timing, we were not able to submit the proposal to the Pew Charitable Trusts, but Robert Schwartz, education officer at Pew, helped initiate the project and provided good advice. Debbie Leong-Childs, the Executive Director of the Academy, coordinated the work of the Panel and staff in myriad, daily ways; she joins me in expressing our fulsome if insufficient gratitude to co-chairs Milbrey McLaughlin and Lorrie Shepard for their leadership and for their devotion of a prodigious amount of time to this project. It is, in the end, their report.

Carl F. Kaestle
President, The National Academy of Education
Professor of Education, University of Chicago

Executive Summary

Education Goals and Standards-Based Reform

The current educational reform movement arose from public dissatisfaction with the low level of achievement demonstrated by American students, particularly in comparison with the performance of students from other industrialized nations. In an effort to improve education and raise student achievement, President Bush and the nation's governors agreed in 1989 to a set of broad education goals to be reached by the year 2000. Those goals and the standards-based reform movement developed out of the commonsense notion that student effort and achievement are directly affected by expectations set by parents, teachers, schools, and the society at large. Thus, standards-based reform calls for the setting of challenging standards in academic subject areas as an important means of improving student achievement. In 1994, Congress passed the Goals 2000: Educate America Act, which enacted the education goals into law and provided resources for the development of standards and assessments.

Purpose of this Report

The National Academy of Education (NAE) Panel on Standards-Based Education Reform was invited to comment on the prospect of national and state standards in education and, in particular, to discuss some of the problems that would be faced by the National Education Standards and Improvement Council (NESIC), the body provided for in Goals 2000 to certify national content and performance

standards as well as state assessment systems. Although Congressional sentiment about NESIC has changed since this Panel began its work, making it improbable that appointments to NESIC will be made, fundamental problems nevertheless remain for the professional associations and state and local agencies engaged in developing standards and assessments.

From the beginning, Panel members agreed that our purpose is not to endorse a standards-based approach as the single best means to improve public education, but rather to make recommendations, to those involved in implementing standards-based reform, about how to proceed, given what is known from available research and what the consequences of various policy choices are likely to be.

Competing Views on Standards-Based Reform

In this report, we summarize both the vision and intentions of standards-based reform and the arguments of its critics. These competing perspectives not only help to frame a set of guidelines for how standard setters should proceed, but they also create an evaluation framework by which reformers can test for intended and unintended effects and continue to improve the system.

The vision of standards-based reform

The intentions of standards-based reform are best captured in the slogan "high standards for all students." Unlike previous reforms focused on the attainment of minimum competencies, setting higher academic expectations is expected to ensure that students have more complete mastery of challenging subject matter and the ability to apply what they know to solve real-world problems. Standards-based reform also reflects a strong commitment to educational equity: The same high expectations are to be established for all students, even groups who have traditionally performed poorly and received watered-down curricula.

Criticisms of standards-based reform

Arguments against standards-based reform are varied and tend to cluster around the following lines: Focusing on standards and assessments (1) diverts attention

from more fundamental educational needs, (2) undermines professional and local responsibility for student learning, (3) heightens inequities so long as resources are unequal, and (4) for some threatens a basic-skills conception of subject matter. People holding very different philosophical positions may nonetheless be allied in opposing standards-based reform.

Guidelines for Attempting an Unprecedented Task

The implications of standards-based reform are unprecedented and monumental. Teachers must learn new content and new ways of teaching, and school systems must become dramatically more effective in delivering resources. Although there exists an extensive research base that informs our understanding of student learning and educational practice, our knowledge about significant aspects of standards-based reform is fragmentary. The challenge remains to implement on a broad scale and for all students what has proven effective in smaller-scale individual cases.

The Panel offers the following guidelines to support the development of reform strategies:

- Standards-based strategies should be informed by existing research knowledge.
- Standards-based strategies should be implemented with constant attention to equity issues.
- Standards-based strategies should foster shared responsibility for student learning in the community and at all levels of the educational system.
- Evaluation and ongoing improvement of the educational system should be a central feature of standards-based reform.

Summary of Issues and Recommendations

Specific elements in a system of standards and assessments are content standards, performance standards, opportunity-to-learn standards, and assessments. Definitions of the elements, and the Panel's recommendations regarding each one, are listed below:

Content standards are broad descriptions of the knowledge and skills students should acquire and be able to do in a particular subject area.

- Because there is not one best way to organize subject matter in a given field of study, rigorous national standards should not be restricted to one set of standards per subject area.

- Content standards should embody a coherent, professionally defensible conception of how a field can be framed for purposes of instruction. They should not be an exhaustive, incoherent compendium of every group's desired content.

- Content standards should be consistent with a research-based model of learning.

- All relevant stakeholders, especially the profession and the public, should be informed about and involved in the adoptions of state and local content standards.

- Although content standards in each subject should be challenging, it should nonetheless be possible to teach them effectively within the constraints of the normal school day and year.

- To the extent possible, connections between content standards and benchmarks for student performance should be evident throughout content standards documents.

Performance standards are the more specific concrete examples and explicit definitions of what students must know and be able to do to demonstrate mastery of the content standards.

- Since using performance standards for accountability and certification imposes more difficult technical requirements, initial efforts should focus on developing performance standards to exemplify specific learning goals.

- Performance standards should include multiple benchmarks—along a continuum leading to high standards of proficiency—so as to demonstrate progress for students at all levels of achievement.

- A national organization or consortium of states should sponsor more in-depth study of technical, legal, and fairness issues associated with performance standards.

Opportunity-to-learn standards refer to the availability of resources, programs, and qualified teachers needed to enable all students to meet challenging content and performance standards.

- Opportunity-to-learn (OTL) standards should focus on those elements of schooling that are directly related to student achievement, rather than relying solely on traditional "input" indicators.

- To the extent possible, OTL standards should focus on the enacted curriculum rather than the reported or official curriculum.

- In the absence of proven reliable and trustworthy technology for measuring OTL, tentative indicators should be tried out and evaluated to determine if they are reasonably connected to student achievement.

- Student assessment should not be postponed until OTL standards have been met, but there should be no high-stakes consequences based on assessment outcomes until OTL has been addressed satisfactorily.

- The U.S. Department of Education as well as state agencies and consortia should support efforts to investigate ways to ensure opportunity to learn.

Assessment refers to the measurement of what students know and can do.

- Assessments should be compatible with and exemplify the content standards.

- Assessments should be accompanied by evidence of validity, reliability, and fairness.

- Assessments should allow students to demonstrate proficiency by multiple methods.

- Limited-English-proficient (LEP) students should be assessed in challenging subject matter in the language of instruction, with special assistance in their first language to ensure they are able to show what they know. The language and literacy development of LEP students should be assessed in both their first and second languages.

- Assessments should be visible indicators of school effectiveness. However, external assessments should not be used to make decisions with severe negative consequences for individual students or teachers without evidence of validity and consideration of opportunity to learn.

CONCLUSIONS

Critical implementation concerns remain. How will new challenging standards in each subject area fit together within the instructional day and school year? What substantive and technical solutions will be found for difficult problems arising in the development of performance standards and new forms of assessment? How will the capacity of states, districts, schools, and teachers be developed so as to reach a fundamentally different level in the quality of education?

Rather than waiting until all technical and implementation questions have been resolved, the Panel recommends a cautious, "learn-as-you-go" approach. As long as there are safeguards to protect individual students and teachers from imperfect and incomplete standards and assessments, it makes sense to try out standards, assessments, and accompanying changes in curriculum and instruction, because their effects can only be meaningfully evaluated in real school contexts.

The need for ongoing research and evaluation pervades the Panel's recommendations. In addition, we emphasize the need to build the capacity of the system by providing more and different opportunities for the professional development of educators and by communicating knowledge gained—from mistakes as well as from successes—to and among standards implementors and the general public.

Given the implementation concerns and the need for capacity building, the Panel believes that a national or quasi-national organization—although not necessarily a federal agency—may be valuable to inform standards-based efforts across disciplines and states. The most important purpose to be served by such a body, whether a public consortium or in the private sector, would be to foster a national conversation and learning venture so that states and districts will not have to struggle in isolation to address similar technical, institutional, and political problems.

Most centrally, standards-based reform must attend to the inequities that exist throughout the system and compromise opportunity to learn. Ultimately the goal of high standards is to continually improve the educational system—to learn what "opportunity to learn" means and how to provide it—so that more and more students can master challenging curricula and lead productive lives.

PART I

THE NATIONAL CONTEXT OF STANDARDS-BASED REFORM

The current educational reform movement arose from public dissatisfaction with the low level of achievement demonstrated by American students, particularly compared with the performance of students from other industrialized nations, and from the belief that such lack of achievement was largely the result of low expectations and mediocre instruction. Although, contrary to popular belief, academic achievement has not declined over the past four decades, nevertheless, there is evidence that even the best American students do not fare well in international comparisons, and many are ill prepared academically for the demands of citizenship and employment in America today. Standards-based reform developed out of the commonsense notion that student effort and level of achievement are directly affected by the expectations that have been set. Thus, standards-based reform calls for the setting of standards in academic subject areas as an important means of improving student achievement. Once agreed upon, standards are expected to affect performance by focusing the effort of students, teachers, and schools and by providing a yardstick to monitor progress.

THE 1989 EDUCATION SUMMIT

The current reform movement has many precursors, but the 1989 Education Summit in Charlottesville, Virginia, is commonly cited as the beginning of the effort to create a national agenda to improve education and raise student achievement. President Bush and the nation's governors agreed on six broad goals to be reached by the year 2000. Two of those goals focused on student achievement:

> **Goal 3:** By the year 2000, American students will leave grades 4, 8, and 12 having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy. [The list of subject areas was subsequently expanded to include foreign languages, civics and government, economics, and the arts.]
>
> **Goal 4:** By the year 2000, U.S. students will be first in the world in science and mathematics achievement.[1]

Motivated by concerns that the U.S. was losing its competitive advantage in world markets, reformers conveyed their intentions in the rhetoric of "world class" standards. At the same time, use of this terminology and emphasis on challenging subject-matter standards represented a rejection of previous educational reforms focused on the concept of minimum competency.

Professionally Set Subject-Matter Standards

Even before politicians began wrestling with how to focus efforts and resources to make schooling more effective, subject-matter specialists have been addressing needed reforms in curriculum and instruction. These substantive reforms are based on new knowledge from research about how students learn and what they need to know to meet the demands of contemporary society and an ever-changing, globally competitive workplace. Subject-matter specialists have focused curricular and instructional reform on higher-order thinking, conceptual understanding, and problem solving because, unlike rote learning, these learning goals seek to foster students' ability to master more advanced content and to apply their knowledge in real-world contexts. Thus, subject-matter specialists in each of the disciplines share policy makers' desire that rigorous and challenging standards be set for student learning.

Mathematics is the subject area that, to date, has received the most attention. The National Council of Teachers of Mathematics (NCTM) created curriculum

[1] National Educational Goals Panel. 1991. *The National Education Goals Report: Building a Nation of Learners.* Washington, D.C.: Author.

standards, with accompanying teaching and assessment standards, that serve as a model for the type of professionally set standards that can reshape what should be taught and what students should be expected to know and do. Reform of mathematics education, calling for higher expectations and greater mathematical literacy, has also been supported by studies conducted for the National Research Council (NRC) by the Mathematical Sciences Education Board, the Board on Mathematical Sciences, and the Committee on the Mathematical Sciences in the Year 2000. In other disciplines, professional associations and the NRC are in the process of developing and promulgating national standards for teaching and learning.

These national, discipline-based efforts to improve curriculum and instruction have also been reflected in standard-setting activity at the state level. Forty-nine of 50 states have now developed or are in the process of developing curriculum frameworks; some of these have been impressive enough to influence national standard-setting efforts. It is important to acknowledge that these standard-setting endeavors are as varied as they are numerous; on occasion, one group of standard-setters may hold views that are fundamentally incompatible with those of another group regarding the nature of subject matter or the best means to change education.

Linking Standards, Assessments, and Opportunity to Learn

When national educational goals were promulgated in 1989, it was unclear how standards were to be defined or how progress was to be assessed. In 1991, Congress created the National Council on Education Standards and Testing (NCEST) to consider whether there should be national standards and examinations. The NCEST report, *Raising Standards for American Education,* did much to codify and clarify the underlying assumptions shared by proponents of national goals and standards-based reform.

First, NCEST defined content standards and performance standards. Content standards are broad curricular goal statements that "describe the knowledge, skills, and other understandings that schools should teach in order for students to attain high levels of competency in challenging subject matter."[2] For example,

[2]The National Council on Education Standards and Testing. 1992. *Raising Standards for American Education: A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American people.* Washington, D.C.: Author, p. 13.

the NCTM Standards comprise only 13 or 14 broad standards for mathematics education at each level of schooling. Performance standards are more specific. They stipulate what students must know and be able to do.

The NCEST report also identified two other key components—assessment and opportunity to learn—now assumed to be integral parts of standards-based reform. Traditional assessment measures have tended to focus more on lowest-common-denominator basic skills. Challenging standards will require new forms of assessment fashioned after Advanced Placement examinations and direct performance measures, focusing on more advanced-level thinking. NCEST also called for the development of school delivery standards, now called opportunity-to-learn standards, in response to the criticism that implementation of challenging standards in the absence of adequate teaching resources would widen the already existing achievement gap between advantaged and disadvantaged students. "School delivery standards should provide a metric for determining whether a school 'delivers' to students the 'opportunity' to learn well the material in the content standards."[3] For example, opportunity-to-learn standards should address whether schools provide high-quality instructional materials and well-prepared teachers.

Goals 2000: Educate America Act

In March 1994 Congress passed the Goals 2000: Educate America Act, enacting into law the education goals set forth by the Education Summit and adding two new goals addressing parental participation and teacher preparation. Goals 2000 supported the concept of standards-based reform laid out by NCEST and provided resources for developing standards and assessments and for augmenting state and local improvement initiatives. It established in law the National Education Goals Panel, originally created by President Bush and the National Governors' Association to report on progress toward the goals, and created a new body, the National Education Standards and Improvement Council (NESIC), authorized to develop criteria and review national content and performance standards as well as state assessment systems and opportunity-to-learn standards submitted on a voluntary basis. At the time of this writing, however, it is unlikely that Congress will approve appointments to NESIC or provide funds for its support.

[3]Ibid., p. E-5.

The NAE Panel on Standards-Based Education Reform

Implementation of standards-based reform is far from straightforward. Because the U.S. has never before had formal national education standards, fundamental questions about the role of standards, the process by which they will be developed, how they will be used in the educational system, and how attainment of standards will be assessed remain to be addressed. In the summer of 1994, the Carnegie Corporation, MacArthur Foundation, and the Pew Charitable Trusts originally asked the National Academy of Education (NAE) to convene an expert panel to examine relevant research and make recommendations to NESIC about the issues it would face.

However, from the beginning the NAE "Panel on NESIC" recognized that its audience was much broader and the issues more far-reaching than those assigned to a single yet-to-be-named federal body. National legislation did not arise in a vacuum. Governors participated in the announcement of national goals in 1989 and many states have already launched reform efforts centered on challenging standards and new forms of assessment. Draft national content standards in several subject areas have been released recently, and others are near completion. Public debate over the content and quality of these standards has begun and is, in some cases, vociferous. How will states and local communities respond to these efforts and resolve disputes? How will they design, revise, and implement their own assessment systems? These questions remain with or without a NESIC.

The Panel, now more appropriately named the NAE Panel on Standards-Based Education Reform, was broadly constituted to include members who differ markedly in their views of standards-based reform. In undertaking their work, Panel members agreed that the purpose of this report is not to endorse a standards-based approach as the single best means to improve public education. Rather, its purpose is to advise those engaged in implementing such reforms about which decisions and actions are most likely to foster realization of the reform's grandest claims, and what pitfalls should be avoided.

An Overview of the Report

The next section of this report presents the vision and intentions of standards-based reform and a discussion of the opposition to the development of national and state standards. To the extent that conclusions, pro and con, are grounded in

research, supporting evidence is summarized. However, in most cases the effects of standards-based reforms cannot be predicted with certainty. What will happen when a particular set of standards or an assessment system is implemented will depend very much on the particular features of local programs. Therefore, we use competing claims and the likelihood of their being realized to inform our recommendations.

Part Three of this report is a discussion of the categories of standards identified by NCEST and Goals 2000—content standards, performance standards, and opportunity-to-learn standards—as well as a section on assessment. In each section we provide basic definitions and examples, an explanation of how each type of standard and assessment is expected to be used, and a discussion of issues leading to the Panel's recommendations. In Part Four we revisit the report's central theme—how to proceed with the development of standards and assessments, given incomplete technical and practical knowledge, with sufficient safeguards and ongoing evaluation so that states, local communities, and professional associations can constantly learn from each other as new knowledge is gained.

PART II

THE VISION AND CHALLENGES OF STANDARDS-BASED REFORM

Internationally competitive standards for what American students should know and be able to do are expected to improve the substance of school curricula and to increase the motivation and effort of students, teachers, and school systems. Opponents criticize the standards movement for a variety of reasons, including opposition to a national curriculum, fears regarding the negative effects of testing, and concerns that a standards-based system will exacerbate existing inequalities in educational resources and outcomes. In this section we present, first, the vision of the current educational reform movement, and then a summary of the criticisms leveled against it.

Our intention is not to determine who is right in this debate. All sides can point to evidence from current practice to support their claims and predictions, but all sides also make significant extrapolations, beyond current data and experience, as to what will happen if their version of reform is implemented. We attend to these competing perspectives because they help to frame a set of guidelines for how standard setters should proceed. In addition, the claims and counterclaims create an evaluation framework by which reformers can test for both intended and unintended effects and continue to improve the public education system.

THE VISION OF STANDARDS-BASED REFORM

The standards being developed by various professional groups, states, and local districts are by no means uniform in their conception of curricula or their philosophical assumptions. Nevertheless most standards-based approaches to

education reform have a common overarching goal: high standards for all students. This goal has two equally important strategic components: challenging standards for what students should know and do and an accompanying agenda for educational equity.

CHALLENGING STANDARDS

In international comparisons of academic achievement, American students consistently rank below students in many other industrialized nations. For example, in the 1991 International Assessment of Educational Progress in mathematics, 13-year-olds in the U.S. ranked 13th of 14 participating nations.[4] Using a high-level cutoff score, only 3 percent of American students scored at an "advanced" level reached by 24 percent of Taiwanese students and 15 percent of Koreans. Statistical comparisons of this type are buttressed by examples of curricula and examinations from other countries suggesting that more is expected of students elsewhere than in the U.S.

Policy makers and educators rightly fear that students are learning too little—either to be competitive with other nations or to lead productive lives in this country. At the same time, however, there is substantial evidence that schools have been "successful" in responding to the minimum-competency and basic-skills reforms of the 1970s and 1980s. For example, the U.S. was second only to Finland in international assessments of reading.[5] During the decade of the 1980s, National Assessment of Educational Progress data in various subjects showed an improvement in basic-skills achievement accompanied by a decline in problem solving and higher-order thinking. Taken together these strands of positive and negative evidence suggest not a lack of will or of accountability for public education, but rather that the standards and focus of effort have been set too low.

At the end of the 1980s, reformers pointed to widely used textbooks, standardized tests, and the prevalence of worksheets as evidence of a "de facto national curriculum" aimed at low-level reading and arithmetic skills. The goals of minimum standards, such as being able to balance a checkbook and read a bus schedule, have had a pervasive effect on how instruction is conceived and pro-

[4]A.E. Lapointe, N.A. Mead, and J.M. Askew. 1992. *Learning Mathematics, The International Assessment of Educational Progress*, Report No. 22-CAEP-01. Princeton, N.J.: Educational Testing Service.

[5]United States. Department of Education, National Center for Education Statistics. 1994. *Reading Literacy in the United States*. Technical Report. Washington, D.C.: Author.

vided, especially for students who are not college bound. Even the curriculum for college-preparatory students in many cases consists largely of memorizing facts and algorithms and answering fill-in-the-blank questions at the end of a chapter. If, in contrast, we want students to learn more and to accomplish more—if we expect students to be able to write well-organized and well-documented essays, use mathematical knowledge to solve real-world problems, and understand and apply scientific principles—then we must communicate these higher expectations to students, teachers, parents, and the community at large, and set very different standards supported by corresponding changes in instruction.

While promoting high expectations, however, standards-based reformers do not advocate that schools simply do more and better what they have always done so that students can reach a higher rung on the same achievement ladder. Rather, the curriculum and instructional changes intended by the new reforms differ substantially from current practice, based on changing economic and social conditions and on new understandings about teaching and learning. Two main differences are apparent:

One, the new standards focus more on depth of understanding—how well students can reason with and use what they have learned—rather than on regurgitation of isolated facts. Much of current instructional practice is based on behavioral-learning theory from the early part of this century. According to this theory, learning occurs by reinforcement of low-level skills that become the building blocks for more complex understandings. An unfortunate practical consequence of this approach is that thinking and reasoning about core concepts are postponed, for some students indefinitely. In more recent decades, learning researchers have demonstrated that memorizing the facts first does not lead automatically to an ability to analyze and apply what has been learned. Learning requires thinking. Even reading the simplest texts requires that students actively comprehend what the text is saying, or decide to reread if they don't. According to current theories, learning requires the active construction of mental models or representations by the learner. Each person has an organized set of ideas in his or her head about how concepts are related in math, science, or history; for learning to occur, when new information or ideas are encountered, the individual must think about and figure out how they fit within the existing structure or must reorganize their mental schema to accommodate the new knowledge.

Teaching for understanding implies a change in methods of instruction as well. Instead of lock-step curricula that stress rote memorization of facts and hierarchical acquisition of skills, the emerging standards emphasize meaningful

interaction with content, hands-on experiments and projects, and applications of knowledge and skills to real-world problems. Correspondingly, the implication for assessment is that skills not be evaluated in isolation. Rather than asking students merely to show factual knowledge, assessments should require the kinds of problem solving, applications of knowledge, and complex performances that are ultimately the goals of learning.

Two, the new standards also reflect changing conceptions of the types of knowledge and skills required for productive employment, responsible citizenship, and personal fulfillment in the 21st century. For some advocates of a standards-based accountability system, imposing standards means a reaffirmation of the traditional curriculum they recall from their own days in school. However, for most advocates, development of standards implies a rethinking of what students need to know based on the demands of contemporary society.

For example, The Business Roundtable, the National Governors' Association, the American Association for the Advancement of Science, and other reform-minded groups assume that the effects of technology, the explosion of knowledge, and new demands of the workplace will play a strong role in redefining what is considered core knowledge and skills. This core would include greater emphasis on complex thinking (including the ability to analyze novel situations, evaluate multiple perspectives, and create appropriate means for solving real-world problems); on communication skills and the ability to work in groups; on the ability to learn and adapt to fast-paced changes in the economy and society; and on technological knowledge and skills. Note that these economic- and workplace-based views, though different in origin, are completely compatible with the views of learning theorists in their emphasis on the need for more challenging academic work and the application of knowledge and real-world problem solving as integral parts of instruction.

Educational Equity

Evidence of inequality of opportunity in U.S. schools is profound. Systematic differences in available resources correspond closely to differences in students' social and cultural backgrounds and economic circumstances. Not only are the curriculum and instruction in college-preparatory tracks more challenging and rewarding than in lower tracks, and not only are students in advanced tracks disproportionately from more advantaged backgrounds, but important gate-keeping courses like calculus are simply unavailable in many schools with large numbers of

poor or minority students. Between communities and even within the same schools, students of differing circumstances have differential access to experienced and knowledgeable teachers, to computers and books, and a classroom atmosphere focused on learning; in some communities even the most rudimentary requirements of clean, safe, and healthy school facilities are not met.

Unequal resources contribute, in turn, to unequal results on high-stakes tests, with poor and minority children bearing the consequences of low scores in the form of delayed kindergarten entry, grade retention, placement in lower tracks, and denial of diplomas. Extensive research has shown many of these well-intended remedial treatments to be ineffective. Thus, initial inequalities are compounded by ineffective and unfair treatment within the education system.

Current educational practices are still rooted in 100-year-old beliefs about the fixed and permanent nature of human abilities. For example, children who are seen as slow learners are given less to learn, and thus confirm everyone's low expectations by knowing very little at the end of a year of remedial worksheets. Similarly, instead of providing rich story-reading experiences to children who come to school with no prior literacy experiences, many schools use screening tests to encourage such children to stay home from kindergarten until they are "more ready to learn."

The slogan "high standards for all students" is fundamentally a refutation of these past practices. Supporters of standards-based reform argue that it could in fact be a tool for achieving greater educational equity. The intention of such reform is, first, to insist that high standards apply even to students who begin at a low level in order to forestall acceptance of a lower set of expectations for these students. Then, advocates believe, standards and accompanying assessments can be used to spotlight inadequacies in the system if certain groups of students do not achieve intended performance levels. As the NCEST report states, "[In] emphasizing their applicability to all students, standards and assessments will help assure that adequate resources are available and appropriately targeted to helping all students attain the standards."[6]

It should be emphasized that the commitment of standards-based reform to equity is more an aspiration than a certainty. The major industrialized nations with which the U.S. competes make no effort to educate all students, but ruthlessly select, seeking only to educate the best students to high standards. Thus,

---

[6]NCEST, op. cit., p. 10.

this expressed American ambition to bring all students to high standards is unprecedented in world history. Nevertheless, the rhetoric that all students can perform to the same high standards is deeply troubling to some educators, who recognize how much effort and attention will be required, given present circumstances, to make this claim a reality.

To be sure, the negative consequences of organizing an educational system around the belief that not all students can master challenging subject matter are well documented. There are also smaller-scale research studies that demonstrate that when low-performing students are exposed to sustained instruction aimed at comprehension and thinking, their academic achievement improves substantially. What remains to be seen, however, is whether these opportunities can be delivered on a broad scale throughout the educational system and whether the improvements will be great enough to ensure that high standards are met. Merely declaring higher expectations without implementing concomitant changes in curriculum, instructional practices, and resources may doubly victimize students if and when they fall short of the standards.

Criticism of Standards-Based Reform

There are more than two sides to the debate surrounding standards-based reform. For example, many educational researchers concur with new concepts of teaching and learning and with the need for more challenging curricula, but disagree strongly that standards should be imposed centrally by either a state or a national authority. Others want to see external standards imposed, but resist any departure from a traditional, skills-oriented definition of core content. Still others favor implementing standards only after ways are found to guarantee the opportunity to learn. Here we summarize briefly several key points of contention among those who are grappling with the issue of standards-based reform.

Diverting attention from more fundamental educational needs

Serious questions have been raised about whether the standards movement is what is needed most to improve public education. Excellence in academic achievement requires more than setting goals and expecting students to meet them. What if reformers focused instead on the necessary conditions of a highly literate and well-educated nation? What conditions would foster high-level academic outcomes for most of our citizens—A respect within the general populace

for intellect and its use? The guarantee of adequate educational resources? The availability of opportunities to use and benefit from high-level education? If these conditions prevailed, it is less likely that we would have to worry about students meeting high standards.

Setting high standards solves a specific and highly circumscribed problem—that previous expectations have been set too low. Although advocates for systemic reform intend more broadly that standards will be the vehicle for pervasive changes in education, it should be acknowledged that attention to standards may have political motives as well. An exclusive focus on outcomes may prevent debate about how instruction should be delivered or how learning could be supported by the community; and it could avoid attention to adequacy of resources which would be more costly than developing standards.

UNDERMINING PROFESSIONAL AND LOCAL RESPONSIBILITY FOR STUDENT LEARNING

In the Panel's judgment, curriculum standards developed by national professional associations would be helpful as tools for informing and guiding state and local curriculum building, teaching practice, and assessment development. Where they are well conceived and well constructed, they may help local educators reflect on and evaluate their own efforts at reshaping classroom and school-level practices, particularly when concepts of knowledge are changing. Many supporters of standards-based reform point to the coherence and efficiency that can be gained from an agreed-upon set of standards developed through a process of evolving professional consensus.

Critics of the standards movement are less sanguine, however, about the public education system's ability to use standards as tools rather than as mandates or prescriptions for instruction. They fear that national standards may be a step down the road to a highly specified national curriculum and argue that, given the diversity of human experiences and the need to build on prior knowledge in the learning process, it is neither wise nor reasonable to make centralized decisions about what, when, and how ideas should be taught or how student understanding should be tested. (Ironically, the same learning research that supports the substantive changes being made in content standards can be used to argue against a national curriculum and testing system: If students learn by making sense and "constructing" their own understandings, then teaching must be highly adaptive, able to elicit relevant prior knowledge, attuned to local contexts and experiences

to make connections, and open to pursuing questions that arise from students' emergent understandings.)

More specifically, critics worry that if national content standards are used narrowly as "test specifications" for national, state, or local accountability devices, teachers will then merely implement the test-defined curriculum rather than inventing instructional activities consistent with broad curriculum frameworks but responsive to their own students' learning.

Heightening Educational Inequities

Concerns about equity have already been presented as part of the positive vision of standards-based reform. Both advocates and critics agree that the present system is inequitable and that poor and minority children have disproportionately borne the burden of low expectations, inadequate curricula, and poorly prepared teachers. But they disagree about whether standards will remedy or exacerbate these inequities, especially when the standards are accompanied by high-stakes assessments.

Arguments that standards could worsen inequities have two key components. First, it is likely that traditionally disadvantaged groups will have the least access to the kind of instruction needed to reach high standards. The vision of new standards, exemplified by the NCTM math standards, requires that most teachers make substantial shifts in their teaching routines, in the kinds of work they expect of students, and in their assessment practices. Teachers in urban and poor school systems, where instruction is currently often characterized by drill and practice, will have the biggest changes to make but are the least likely to be provided with needed staff development resources. Second, negative consequences of failing to meet the standards will further impede already disadvantaged students. The intention in having students repeat a grade or be placed in a remedial class to prepare for retesting is to help them academically. Past research has consistently shown, however, that students are very often worse off not only academically but also socially, following these placements.

Threatening Basic-Skills Instruction

The intended changes in curriculum and instruction described in the section on challenging standards (Part Three) are opposed by a number of groups each with

slightly different reasons for their objections. Some fear that the deemphasis of skills instruction, especially in mathematics, will leave students ill prepared to take important tests such as the SAT. Some groups worry that advanced students will be slowed down by a common set of standards for all students. Others oppose the emphasis on a "thinking" curriculum because critical thinking and questioning threaten strongly held beliefs and introduce value issues into the school curriculum. In some local applications involving integrated curricula, people see goals aimed at effective communication and cooperative learning as a watering down of rigorous academic content. Specific objections differ from community to community, depending upon the particular features of proposed standards and curriculum and the beliefs and values of parents and community members. Occasionally, groups with philosophically incompatible perspectives form coalitions against the common enemy—standards-based reform.

The Panel concurs with the summary of research on learning presented as part of the vision of standards-based reform. Arguments favoring instruction aimed at helping students make sense of what they are learning are well grounded in empirical research. This does not mean, however, that the Panel endorses all versions of standards-based reform or extreme examples of curricula that abandon skills and content. Ongoing debate about the content of standards has serious implications for efforts to implement reforms. It is not obvious that consensus groups will easily come together and develop coherent curricular goals to improve student learning. In fact, it is a real possibility that they could adopt any of the following extreme, and ultimately detrimental, positions: (1) reify a concept of knowledge as superficial content coverage aimed at regurgitation of discrete disconnected facts; (2) discard content and focus only on process without attention to discipline-based knowledge; or (3) create an incoherent and infeasible set of standards by including everyone's desired content.

An Unprecedented Task

Standards-based reform attempts to raise expectations for students throughout the system while eliminating current inequalities—all at a time when demographic shifts and declining financial support present increasing challenges for already stretched school resources and personnel. This would be a monumental goal even if the proposed content and instructional approaches were familiar to teachers and parents, and educators were prepared for the work involved in realizing them.

But such is not the case. The proposed changes in instructional goals and in approaches to teaching and learning make this a massive educational undertaking for the adults as well as the children in the system. Teachers will be required to learn new content and ways of teaching, and to unlearn the old ways in which they were schooled and successful as students themselves. Moreover, schools and school systems are not organized to assist them in this task. Although some teachers are accustomed to evaluating student work in the context of more challenging performance standards, most large-scale testing and accountability programs reflect only a low-level skills curriculum.

Perhaps most important, the very concept of standards-based instruction is unprecedented in the U.S. We as a nation have little experience setting or maintaining high expectations for any but the most privileged of our students. As public institutions, schools have always been under considerable political pressure to avoid system failure, which is exactly how a finding of large numbers of students performing below standards would be interpreted. Consequently, standards in American education have tended to become minimum standards over time, set at levels that the vast majority of students can attain but that inspire neither student effort nor instructional improvement. As a nation or as individual states, we have little experience defining educational opportunity in terms of programmatic needs and outcomes, attending instead to court-ordered, minimally effective formulae for redistributing resources.

Guidelines for Developing Standards-Based Strategies

The Panel's approach to standards-based reform is based on two contrasting conclusions about the current state of knowledge. One, there is an extensive research base that informs our understanding about student learning and educational practice. Two, at the same time, there are significant gaps in the knowledge we will need to carry out standards-based reform. What is known is fragmentary—some teachers who do outstanding work in a subject, some schools that have been effective in educating poor children, some standards that seem promising, some teacher-education programs that have changed how they prepare teachers. But integrating these fragments into a coherent system remains to be done. The challenge for reformers is to implement, on a broad scale and for all students, what has proven effective in individual cases. With this in mind, the Panel offers the following four guidelines to help ensure that reform strategies continue to be developed and improved.

USE EXISTING RESEARCH KNOWLEDGE

While much about standards-based reform is unprecedented, we do have a substantial knowledge base on which to build. The results of the last 30 years of research in cognitive psychology, for example, are relevant to discussions about which curricula and instructional approaches are most likely to help students learn. Other equally important bodies of work inform our understanding of teacher learning, institutional change, and the ways that school structures, resources, and social conditions can either support or thwart learning.

While the Panel concurs with the view that standards should be developed through a public process, we believe that at its center, standards-based reform should be guided by existing research and professional knowledge.

PAY CONSTANT ATTENTION TO EQUITY ISSUES

Equity receives special attention in this report because, if standards-based reform falls short of its promises, the potential for harm to traditionally disadvantaged groups is great. We urge attention to equity issues both in the design of standards strategies and in the ongoing evaluation of program effects. This means there must be opportunities for individuals of diverse backgrounds and those who are knowledgeable about students with special needs to participate in designing and responding to the standards. It means that standards and assessments must be valid and fair—for example, by allowing special-needs student or those with limited proficiency in English the opportunity to demonstrate their knowledge in appropriate and multiple forms. It means that standards must be accompanied by explicit strategies and sufficient resources to facilitate their realization by students who have been traditionally underserved by the system. Finally, it means that the consequences of implementing standards and assessments—who fails, the effectiveness of remedial instruction, the effects on subsequent educational and job opportunities—should be closely monitored, with an eye toward eliminating any inequities in educational opportunity that arise. Particular attention should be paid to the consequences of standards and assessments for identifiable social groups.

FOSTER A SHARED SENSE OF RESPONSIBILITY

The types of changes called for by these reforms cannot be brought about by students and teachers working alone, and therefore students and teachers must not

be held solely accountable for improvement. Altering the expectations, assumptions, and structures that underlie current low achievement and unequal opportunities requires the participation and responsibility of individuals and agencies throughout the educational system. It also requires that individuals and organizations at all levels be held accountable for fulfilling their respective responsibilities. Students need adequate and appropriate instruction if they are to achieve higher standards, and teachers need support and opportunities to try out new approaches if they are to facilitate that learning. Providing that support is the responsibility of district and state agencies, a fact that should be kept in mind when student performance results are reported. Similarly, responsibility for protecting the general welfare and fostering non-school learning of children extends beyond the school walls to parents, community organizations, and social agencies. This perspective helps to make clear how broad-based the reform efforts must be if they are to have a chance of succeeding, and it suggests that student results should not be reported without information about the efforts of other agents responsible for public education—such as parents, district administrators and board members, and state-level policy makers.

Ongoing Evaluation and Improvement of the System

Despite the many issues and problems yet to be resolved, the position of the Panel is that standards-based efforts should not wait until a perfect technology is developed. For one thing, a complex system of standards, assessments, and supportive instruction has to be tried out and revised in the context of real schools, not in a laboratory. We argue that standards and assessments should be developed and implemented using two strategies—safeguards and ongoing evaluation—to acknowledge that initial efforts may have serious flaws. Safeguards, such as forbidding certain high-stakes uses of assessment results, can protect students, teachers, and schools from the negative consequences of such things as unreliable assessments or inadequate instructional support. At the same time, ongoing systematic evaluation and research, which any sound business would conduct, can prevent negative side effects greater than the intended benefits of standards-based reform. Evaluation data should not only be used to improve local systems but should also be made publicly available so that states, districts, and professional associations can learn what worked and what did not work in other settings.

With these guidelines in mind, the Panel has developed the specific recommendations that appear in Part Three of this report.

PART III

ISSUES AND RECOMMENDATIONS

The current reform movement in education introduces three types of standards for consideration: content standards, performance standards, and opportunity-to-learn (OTL) standards. Below we discuss the definitions, uses, and relevant recommendations for each of these types, as well as for assessments related to the standards.

Each standard type has issues specific to its own development and implementation. In practice, however, the three types are closely interrelated and mutually reinforcing. Without related standards for students' performance, for example, broad content standards remain vague and open to misinterpretation. Furthermore, in effectively implemented systems, we also assume that standards and assessments eventually will be linked, both substantively and practically. For this reason, issues involving the relationships between and among elements of the system will also be discussed.

CONTENT STANDARDS

DEFINITION

There is no single view of what content standards should include, how detailed they should be, or what format is most useful. The National Council on Educational Standards and Testing defined content standards as "the knowledge, skills, and other understandings that *schools should teach* in order for students to attain

high levels of competency in challenging subject matter" (NCEST, 1992, emphasis added). The Goals 2000: Educate America Act took a slightly different tack: "Content standards are broad descriptions of the knowledge and skills *students should acquire* in a particular subject area." (emphasis added) Here the focus was on student learning rather than curricular content.

Examples of content standards currently in use reflect these differences in emphasis. For example, California's 1985 Model Curriculum Guide for English Language Arts illustrates the first approach, emphasizing the nature of the curriculum and instruction (see Guideline 2, Figure 1). The Curriculum and Evaluation Standards for School Mathematics published by the National Council of Teachers of Mathematics in 1989, however, takes the second approach, focusing on what students should learn. Standard 11 for grades K-4, statistics and probability, reads:

> "In grades K-4, the mathematics curriculum should include experiences with data analysis and probability so that students can—
>
> - collect, organize, and describe data;
> - construct, read, and interpret displays of data;
> - formulate and solve problems that involve collecting and analyzing data;
> - explore concepts of chance."[7]

Formulated in terms of what students should be able to do with regard to mathematics, the NCTM standards have become the model for much of the current standards development work. In the NCTM standards, the implications for curriculum and instruction derive from the student goals and are explored briefly after explanation of the standard and its rationale. Also included are examples of problems illustrative of each standard and appropriate for specific age levels.

Content standards are not simply a listing of important knowledge and skills, however. The preface to the NCTM standards describes them as a vision of what the mathematics curriculum should include in terms of content priority and emphasis. Put another way, content standards frame a field for the purposes of learning and mastery. That field may be a traditional single discipline or a domain across disciplines; in either case, what is important is that the standards provide a coherent structure to guide curriculum and instruction.

---

[7]National Council of Teachers of Mathematics. 1989. *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author, p. 54.

Figure 1. A sample guideline from the California 1985 Model Curriculum Guide for English Language Arts.

Guideline Number Two

All students, individually and in small groups, read and respond in a variety of ways to literary works, selected with the help of their teacher, that extend or enhance the classroom study of core works.

Representative enabling activities:

K-3

The teacher, with the help of a library media specialist, assembles a classroom library of literary works that have commonalities with the core works that are used in the classroom. Groups of students select from among these works in the classroom library and each group decides on a method for dramatizing the selected work. Methods such as readers' theater, puppetry, pantomime, or choral reading might be employed. Each group presents its selection to the entire class. Some of the selections might then be presented to other classes, parents, or to other audiences, or they might be videotaped for future showings.

English-Language Arts

Reprinted from *English-Language Arts Model Curriculum Guide, Kindergarten Through Grade Eight.* Sacramento: California Department of Education, 1987. Used by permission.

Uses of Content Standards

Framing the Conversation

One of the most important uses of content standards throughout the iterative process of development, implementation, and revision is to advance the public and professional conversation about what is important for schools to teach and for students to learn. The NCTM standards mentioned above provide numerous examples of this use. The standard on probability and statistics, for instance, places emphasis on an area of mathematics that traditionally has received little attention in the mathematics curriculum, particularly in the elementary grades, but that frequently is required in everyday applications of mathematical knowledge. Its presence in the standards has prompted schools and districts to rethink their curricula, and teachers to seek new knowledge.

Another example of this use can be seen in NCTM's more controversial emphasis on developing students' ability to communicate—discuss, read, and write—about mathematical ideas. The notion that communication is an important aspect of mathematical knowledge and skills is a marked change from adherence to the routine application of step-by-step procedures that has characterized traditional mathematics instruction. This standard has provoked many questions and concerns among teachers, parents, and the community, which can be welcomed as opportunities for educators to explore among themselves and with the public the connections between developing communication skills and enhancing reasoning abilities, or the use of mathematics in developing language skills.

Guiding curriculum and instruction

A second use of content standards is to guide instruction and assessment. To the extent that standards frame a field in useful and meaningful ways for instruction, teachers and curriculum planners can look to them to provide general direction as they create specific curricula relevant to the needs, strengths, and contexts of their particular students. The English department in one California middle school, as an example, used the state's language arts framework to articulate the curriculum across the grades, identify weaknesses in individual teachers' knowledge and instruction, and assist teachers in improving classroom instruction. In another example, teachers in an elementary school that bases its curriculum on interdisciplinary projects used a combination of NCTM standards, American Association for the Advancement of Science (AAAS) Project 2061 benchmarks[8], and state standards to evaluate the content covered across the projects, determine what was lacking, and develop new projects to ensure that all of the respective mathematics and science standards were being addressed.

Some reformers suggest that the linkage between content standards and performance standards will prove most useful for instructional guidance. As discussed in the next section, performance standards define and provide concrete examples of the desired levels of student achievement expected by the content standards. They also provide real targets at which students and teachers can aim, as well as a substantive basis on which teachers can evaluate student performance and identify areas for improvement—both in student work and in their own instruction.

---

[8]Project 2061, American Association for the Advancement of Science. 1993. *Benchmarks for Science Literacy*. New York: Oxford University Press.

Allocating instructional resources

The third intended use of content standards, particularly in combination with performance standards, is as a basis for the allocation of instructional resources. The underlying question here is, With the learning goals specifically in mind, what will it take to get us there? The "us" may be a state, a district, a school, or even an individual teacher. Marshaling resources for student learning consistent with the content standards might require new efforts to improve teachers' professional knowledge and skills, development of new instructional materials and content assessments, attention to student readiness and prior learning, support structures for other students and teachers, and new organizational arrangements. Opportunity-to-learn standards and strategies, discussed later in this report, are one manifestation of this use of content standards to redirect instructional resources.

The relative emphasis given to three uses of content standards may vary by locality, content area, or reform perspective. In practice, the use of content standards also will depend on whether they are distrusted as a strategy for dictating local curriculum or welcomed as a tool to create locally appropriate curriculum, professional development, and instructional materials.

Key issues

Questions of what content standards should comprise, who should define them, and how they should be used touch complex conceptual, practical, and political nerves. These concerns derive in part from our limited experience with national standards, but they also reflect the tensions inherent in trying to apply a new philosophy to an established system.

One standard or multiple standards?

Much of the discussion surrounding national standards has either assumed, or explicitly argued for, a single set of standards in each subject at the national level, often with some sort of official endorsement, which states and localities could voluntarily adopt for their use. The main argument in support of this approach is that it would promote a core of shared knowledge and skills to which all children in our diverse and mobile society have access. Additionally, advocates of a single-standard approach believe it would help to focus resources and thus increase the capacity of teachers, administrators, and the system as a whole to meet the desired learning goals. They often point to the impact of the NCTM standards on the discourse about mathematics instruction as an example of the potential

usefulness of a single, well-respected set of standards. Endorsement of more than one set of standards, in their view, would only add confusion to an already fragmented and incoherent system.

Underlying such reasoning, however, is the questionable assumption that there is one best way to define and structure knowledge in science, mathematics, or other fields, and that it will be a relatively easy matter to identify and reach consensus on that one way. In fact, there are deep divisions within the disciplines and among the public about what knowledge and skills students need, about the desired relative emphasis on different strands within a discipline, about the best ways to structure knowledge in a field, and about the relative merits of disciplinary versus interdisciplinary approaches. Evidence of these divisions was apparent in the vociferous debate engendered by the release of the draft history and science standards at the end of 1994.[9,10]

The politics of standards also suggest that efforts to minimize conflict among developers within a knowledge domain and to include as many views as possible will result in standards of unwieldy size and scope. Permitting only a single set of standards in a given field could result in "inclusive" standards that might accommodate competing perspectives, but would do little to provide an effective framework for the field.

Some advocate the development and acceptance of multiple sets of standards for a content domain, each of which must meet established criteria to ensure high quality. There might also be interdisciplinary standards—for example, combining mathematics, science, and technology, or history, geography, and economics—that present alternative, high-quality ways to structure knowledge and skills in the applicable domain.

We believe the nation has much to gain from a multiple-standards approach. Assuming that they have been shown to meet a demanding set of criteria, the several sets of endorsed national standards might then serve as high-quality exemplars from which states or localities could choose, either to guide their curriculum and instruction directly or to modify and adapt according to their own conditions and perspectives. Multiple standards could also provide a means of

---

[9]National Center for History in the Schools. 1994. *National Standards for United States History*. University of California at Los Angeles: Author.

[10]National Research Council. 1994. *National Science Education Standards*. Draft. Washington, D.C.: National Academy Press.

enhancing our knowledge base. Evaluating how well different standards operate in different contexts can enhance our understanding of more and less effective ways to organize instruction, and can keep the standards themselves dynamic and evolving.

One concern raised by the possibility of multiple standards in a field is that they will detract from and even preclude any real commitment to educational equity. Might not the most valued knowledge and skills remain reserved for certain selected groups of students, in the form of one set of standards, while less-valued outcomes were embodied in standards used for others? If such were the case, would not the acceptability of multiple sets of standards further promote and legitimize a multitiered system of education?

Recognition of this possibility leads the Panel to underscore the need for all content standards endorsed at the state or national level to pertain to all students within any jurisdiction that adopts them. *Nonetheless, because there is not one best way to organize subject-matter knowledge, the Panel recommends that voluntary standards recognized at the national level not be restricted to one set of standards per subject area.*

WHO SHOULD DEFINE THE STANDARDS?

There are two issues involved here: (1) the need to balance professional and public involvement in standards development and adoption, and (2) deciding at which level of the system standards are most appropriately developed.

Professional and public involvement. The United States is unique in the degree to which curriculum is a public as well as a professional domain. Disputes over sex education and the teaching of evolution attest to the fact that public debates about school curriculum have a long-running history. The centrality of curriculum to standards-based reform strategy, however, has the potential of accelerating public involvement and disagreements over instructional content. This is especially likely when the proposed standards differ markedly from what today's adults experienced in their own schooling and thus have come to expect from the American educational system. Reform efforts in some states and localities have been derailed, even halted, because of opposition from some sections of the public to inclusion of certain topics or values they believe to be undesirable or inappropriate for public education. Given the potential strength of community opposition, reformers have come to believe that the public, as well as educators, must "own" the standards if they are to be incorporated in a deep and lasting way into the instructional program.

To achieve this "ownership," parents and the general public must be informed about proposed standards and have opportunities to influence their adoption and implementation. Stakeholders who have a say in the standards set for their schools are more likely to support them, and to appreciate the challenges involved in crafting and implementing them. Stakeholder involvement also offers substantive benefits; input from parents, the business community, and other members of the general public can strengthen standards by fostering responsiveness to students' diverse needs and to broad public values and goals.

Some avenues for public participation will be direct, through forums and focus groups, public hearings, back-to-school nights, and other parent activities. Others will be more indirect, through public representation on panels and review boards or through the political process. Although it is time-consuming, an open and public process for standards development is essential to building broad support for the reforms. Furthermore, if heretofore disenfranchised groups are drawn into the process, they can ensure attention to equity issues.

Public involvement and discourse, however, must not mean that standards are subject to veto by anyone who disagrees with their content. Some have argued for the importance of consensus in standards development and adoption. But, as discussed earlier, efforts to reach consensus where disagreements are substantial and profound generally lead to superficial and incoherent attempts to incorporate all viewpoints. Standards resulting from such a process will not only be weak pedagogically, but will also almost inevitably reflect past practices rather than reform goals; that is, they will reflect those approaches to knowledge and skills that parents and the general adult population experienced in their own schooling and with which they are therefore comfortable. Basing standards primarily in past practices fails to prepare students for the increasingly complex demands they will confront in the workplace and in society.

The Panel thus concludes that while public involvement is valuable and necessary to the acceptance and long term viability of new content standards, the responsibility for guaranteeing that those standards present a coherent, up-to-date, and challenging vision of a field of study ultimately resides with the professionals engaged in ongoing work in that knowledge domain.

Moreover, professionals are in the best position to ensure that the standards also embody an educationally sound approach to instruction, which in turn implies a coherent and sound model of learning. We have learned much in recent

decades about how children learn and develop, and these new understandings have substantial implications for the design of curriculum and instruction.

Professional consensus on curriculum content, however, is no more fully attainable than is public consensus. Educators often hold differing but professionally reasonable perspectives about the knowledge and skills most essential to a knowledge domain. For example, scientists and science educators do not all agree on which scientific concepts are key or how to structure those concepts for teaching and learning. The National Science Teachers Association's Scope, Sequence, and Coordination Project holds that various subject matters of secondary school science (earth science, chemistry, physics, biology) must be integrated, or at least coordinated. In contrast, Project 2061 states only that students should have learned certain science concepts by the time they leave high school; organization of the science curriculum is immaterial. Given these kinds of disagreement, professional consensus about a set of content standards may be unrealistic as a criterion of acceptability.

But while consensus may be unrealizable, professional defensibility is another matter. In the absence of a universally held concept of a particular knowledge domain or a commonly accepted understanding of how children learn, we believe that the professional judgment of a significant segment of the subject matter and educational communities, based on and backed up by practical experience, sound research, and theory, is the best guide to evaluating the validity of any given approach. *This perspective leads the Panel to recommend that voluntary national content standards embody a coherent, professionally defensible concept of how a field can be framed for purposes of instruction, and a research-based model of learning.*

In recognition of the importance of public participation, the Panel also recommends that all relevant stakeholders be informed about, and included in, the adoption of state and local content standards. Balancing public input with professional judgment will take different forms depending upon the level at which standards are developed. Reform efforts in Vermont, for example, began, under the leadership of the State Department of Education, with a series of public town meetings and other forums to determine broad goals (called "vital results") based on commonly held values. This step was followed by a professional process of developing standards that incorporated those goals and advanced concepts of learning and interdisciplinary subject matter. Those standards were then circulated in draft form for continued public consideration and debate. Delaware, on the other hand, began first with mainly professionals developing the standards, but incorporated public input along the way.

In either case, the development of standards entailed continuing exchanges among professionals, policy makers, and the public. Initiating and sustaining this kind of ongoing dialogue constitutes an enormous task, and a major job of educating the public about this new approach to education.

Level of system. Some argue that it is inappropriate for standards to be developed at the national or even the state level, because a curriculum must be reflective of the specific needs and interests of the students and teachers in a particular school in a particular community. There are both pedagogical and equity aspects to this argument: Teachers must be able to respond professionally to the learning needs of their students; and compulsory public education in a democracy should be relevant to the diverse backgrounds and local contexts of the students it serves.

Those who argue for a more centralized approach to standards development point to the limited resources available at the local level and the need for consistency and commonality in curriculum frameworks among schools. Reinventing curricular specifics in every local district, they argue, is inefficient and disregards the largely common core of knowledge and skills to which all students should have access.

Some of this debate may boil down to a distinction between standards, which are broad descriptions of the field, and curriculum, which is the more specific delineation of scope and sequence necessary to attain the standards. While standards may be developed more centrally, local districts and schools must have the ability to mold the curriculum to fit their students and instructional contexts. Thus, standards should focus on the enduring ideas in a knowledge domain and provide overall direction without constraining local professional discretion either in specific curricular design or in pedagogical strategies. This approach provides a means for easing the tension between general guidelines and local responsiveness to the needs, interests, and strengths of educators and students alike.

How to design standards as tools for instructional improvement

In addition to the overall quality of the standards content, other aspects of their design will help to determine the degree to which they can useful tools for improving practice. We focus here on three of these: the level of specificity of the standards; their overall scope; and the degree to which they are linked to actual student performance.

Level of specificity. It is one thing to say that standards should focus on the "big ideas" in a field of study and quite another to define what this means. Some such ideas are very specific; others are very general. The level of specificity in a set of content standards will have implications for whether or not the standards are useful for instruction and contribute to the development of teachers' capacities. At one end of the continuum, standards could reflect very broad statements of educational goals (such as "students should be able to communicate effectively"); these provide little guidance for structuring the curriculum or designing effective instruction. At the other end, standards could be delineated so specifically as to preclude integration of curriculum or professional discretion in meeting needs of particular students (for example, "students should know the advantages and disadvantages of the city-manager form of government").

Scope. Closely tied to the level of specificity is the scope of the standards—that is, how much of a field of study should we expect schools to teach and students to learn? Answering this question addresses the need to balance making the standards challenging and making them feasible, given the multiple subject areas and the constraints of the school day and year. Some of the draft national standards have been criticized for including so much material to be taught and learned in some specified field (e.g., the arts or geography) that students would have little time to attend to learning in other areas. It runs counter to the broad objectives of standards-based reform to expect every child to master advanced levels of content and skills in every knowledge domain covered in school. Such curricular demands would be unmanageable for both teachers and students; coverage could consume more hours than are available in the school year. *Thus, the Panel recommends that rigorous standards for any content area must also be reasonably attainable in the context of the total curriculum.*

Links to performance. For standards to be useful for instruction, teachers and curriculum developers must be able to understand their meaning clearly and determine when a given activity or sample of student work is on target. For example, what does it mean for a second-grader to "link conceptual and procedural knowledge,"[11] or for a middle-school student to "understand the meaning of time and chronology,"[12] or for a third-year high-school French student to "under-

---

[11]NCTM, op. cit., p. 32.

[12]History-Social Science Curriculum Framework and Criteria Committee. 1988. *History-Social Science Framework for California Public Schools.* Sacramento: California State Department of Education, p. 13.

Figure 2. Examples of student work give meaning to content and performance standards.

The New Standards project is a partnership of states and urban school districts that is adopting a set of high national education standards and developing a new system of assessments. As part of this effort, New Standards is collecting samples of student work at the elementary, middle, and high school levels to illustrate work that meets the standards. Development of commentaries is proceeding to make clear the criteria on which judgments are based.

According to New Standards, evidence of work at the level of the elementary standard for Life Sciences Concepts would reside somewhere between the work samples below. The fourth-grade example shows that "students who are not proficient in English can demonstrate understanding in science." The sixth-grade sample shows elementary level understanding of the needs of organisms and, by including the concept of cells, shows progress towards the middle school life sciences standard.

Elementary Life Sciences Concepts Standard

Understand:

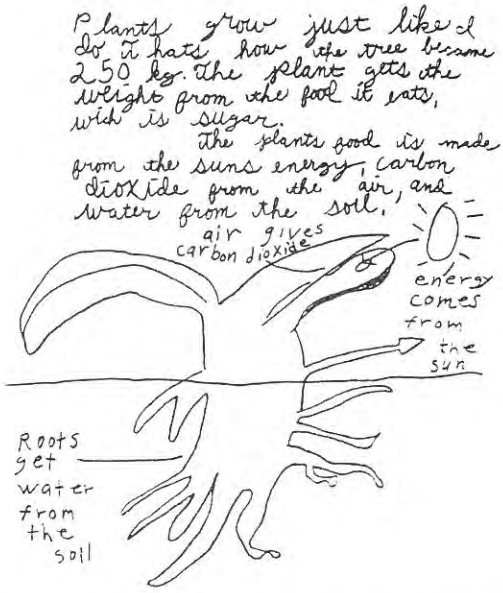
- **characteristics of organisms (needs, environments that meet them:** structures, especially senses; variation and behaviors, inherited and learned);
- life cycles (birth, development, reproduction);
- organisms and environments (food chains, populations, effects on the environment);
- change over time (fossil evidence).

Example of open-ended question:

A small tree is planted in a meadow. After 20 years it has grown into a big tree, weighing 250 kg more than when it was planted.

Where do the extra 250 kg come from? Explain your answer as fully as you can.

Example of a response by a fourth-grade student in a special education reading program.



Example of a response by a sixth-grade student.

Since the tree grew, it would weigh more. It grew by taking in water, light, & minerals. It grows like any other living thing. The tree used the minerals, light, and water to make food. It then uses the food to make cells. The cells are what make up things. It continues to make cells so the tree gets bigger and heavier. After 20 years, the tree should be 250 kg. more than when it was planted because of this.

Reprinted from *Performance Standards*, draft 5.1. Pittsburgh, PA: New Standards, June 1995. Used by permission.

stand topic, main idea, and supporting detail in authentic material"?[13] How is a teacher or a school to interpret any of these standards when designing specific curricula, choosing instructional strategies, or evaluating student progress?

To the extent that content standards include examples of assessment tasks and of various levels of real student performance in response to those tasks, they can better provide teachers and the public with insights into the concrete meaning of the standards and the specific expectations for student learning. (See Figure 2.) Inclusion of examples of student work also helps to ensure that the standards are assessable, thus maintaining the focus on what students should learn. *For these reasons, the Panel recommends that, to the extent possible, connections between content standards and benchmarks for student performance be evident throughout content standards documents.*

SUMMARY OF RECOMMENDATIONS FOR CONTENT STANDARDS

The Panel makes the following recommendations concerning national content standards:

Recommendation I.1: Certification of voluntary national content standards should not be restricted to one set of standards per subject area.

Recommendation I.2: Voluntary national content standards should embody a coherent, professionally defensible conception of how a field can be framed for purposes of instruction.

Recommendation I.3: Content standards should be consistent with a research-based model of learning.

Recommendation I.4: All relevant stakeholders, especially the public and the profession, should be informed about and involved in the adoption of state and local content standards.

[13] South Carolina Foreign Language Curriculum Framework Writing Team. 1992. *The Global Challenge: A Foreign Language Framework for the 1990's*. Field Review Draft. Columbia, S.C.: South Carolina Department of Education, p. 31.

Recommendation I.5: Although content standards in each subject should be challenging, it should nonetheless be possible to teach them effectively within the constraints of the normal school day and year.

Recommendation I.6: To the extent possible, connections between content standards and benchmarks for a range of student performance should be evident throughout content standards documents.

Performance Standards

Definition

In Goals 2000, performance standards are defined as "concrete examples and explicit definitions of what students have to know and be able to do to demonstrate that such students are proficient in the skills and knowledge framed by the content standards." (Goals 2000, Section 3) For example, according to the NCTM, a specific expectation in the area of whole-number computation for grades K-4 is that students be able to "select and use computation techniques appropriate to specific problems and determine whether the results are reasonable."[14] Another example, from the Australian curriculum frameworks for English spanning grades 1-10, has three strands—speaking and listening, reading and viewing, and writing—which are further organized into four categories—texts, contextual understanding, linguistic structures and features, and strategies (See Figure 3, for the text and linguistic features expectations of the Australian curriculum). Even in its generalized statements, the Australian document makes clear the level of work expected by referencing specific texts and giving examples of what students must be able to do. In addition, the Australian curriculum provides samples of student work with features of the work highlighted to show how they meet the standards.

Performance standards may be used to define the level of accomplishment expected for a single assessment task or for an entire course of study. For example, to judge the quality of a persuasive essay, scoring criteria might require that an opinion be stated, that reasons be given to support the opinion, and that

[14]NCTM, op. cit., p. 44.

Figure 3. A sample performance standard from the Australian curriculum frameworks for English.

LEVEL 8 Reading and Viewing

TEXTS

At level 8, a student:

8.5 Analyses and criticises in a lucid way texts produced for a range of purposes and audiences, including popular texts and linguistically demanding texts which may involve varied narrative perspectives and subtle subtexts.

Evident when students, for example:

- Approach with confidence the reading and understanding of visually forbidding texts written for adult audiences (long and densely written material such as extended analyses in 'quality' newspapers, extracts from poetry and prose of previous centuries such as Homer's *Iliad*, Swift's *A Tale of a Tub* or Milton's *Areopagitica*).
- Construct sophisticated readings of popular texts such as children's adventure stories, fairytales and cartoons (examine the social values implied in Bugs Bunny cartoons or in Enid Blyton stories).
- Offer interpretations of texts from different theoretical perspectives (offer both a feminist reading and a more traditional interpretation of Chaucer's 'The Wife of Bath's Tale' or Mukherjee's *Wife*).
- Draw conclusions from synthesising material read and viewed, making generalisations based on patterns and trends (the formula to which Mills and Boon books are written; distinctions within and between texts in terms of characters, themes, inferred authorial intention).
- Make connections that show insight between their own reading experiences and the way others might react to texts in order to explore the significance of aspects of texts ('Vivian's genuine, caring approach to life wins and holds audience support. The audience finds delight in her struggle to come to terms with a world that is foreign to her, reacting with joy when her tormentors get their just desserts.').
- Explore a text through an imaginative response that implies an analysis of or a point of view about the text (composing a letter from Ophelia to Hamlet written just before her 'mad' scene drawing from material in Shakespeare's *Hamlet*.

LINGUISTIC STRUCTURES & FEATURES

At level 8, a student:

8.7 Analyses how linguistic structures and features influence interpretation of texts, especially in the construction of tone, style and point of view.

Evident when students, for example:

- Write about the effect of production design on films (the way black-and-white footage implies documented truth; the way set design subtly suggests aspects of a character's socio-cultural context, as the peacock-tail pattern in the wallpaper behind the main character in the film *My Brilliant Career* offers a perspective on Sybylla's struggle to break free from society's expectations regarding feminine beauty).
- Discuss the way interpretation is influenced by variations of rhythm and pace (by modifying established metres in poetry and by editing in films).
- Explain the specific effects of sound imagery (alliteration in poetry and headlines).
- Analyse the factors, such as certain sentence structures, that contribute to the narrative voice in imaginative texts (the factors which make the Peter Porter poem *Your attention please* sound like a government warning, or the factors which contribute to the macabre qualities of Emily Dickinson's poetry).
- Analyse the use of techniques—such as characterisation, wit, cynicism and irony—designed to position readers and viewers to take particular views ('By giving Edward vulnerable qualities, the film-maker does not allow him to alienate the audience as his associates do.').
- Describe the tone of various texts, using a widening range of adjectives such as 'indignant', 'arrogant', and justifying judgement by close reference to the text (discuss the way the writer of a letter to the editor seems to feel about the topic or describe and analyse Pope's attitude to vanity in his poem *The Rape of the Lock*).

These extracts are reproduced from *English—a curriculum profile for Australian schools*, 1994
Publisher: Curriculum Corporation, 141 Rathdowne St, Carlton Vic 3053
Australia Tel: +61 3 9639 0699 Fax: +61 3 9639 1616.

the opposite point of view be clearly explained and refuted.[15] In addition, a sample essay would serve to show what is expected as an excellent piece of writing. When performance standards are set for an entire discipline, such as language arts or history, the evaluation criteria must also include quantitative rules about how levels of performance (or points) on varied tasks and demonstrations are to be combined to reach an overall standard. If all of the subparts and skills are equally important, then a simple percentage (of all possible points) might serve as the standard. In many applications, however, not all subparts would have equal weight, and some categories of performance might require a higher level of proficiency than others. Furthermore, it must be determined whether, to meet the overall standard, it is necessary to meet the standard for each task or whether exceptional performance on one task can compensate for weak performance on another.

USES OF PERFORMANCE STANDARDS

Performance standards may be used for a variety of purposes, including exemplification of content standards, educator accountability, and certification of individual achievement. The characteristics of performance standards may need to be quite different to meet these different purposes. Providing examples of student work and showing how sample tasks would be scored would help to clarify content standards and can be accomplished without developing the scoring rules for a complete assessment. However, if assessment results are going to be used to report the number of students who have met a performance standard, then the standard must be defined in terms of points to be earned on a complete assessment not just isolated tasks.

EXEMPLIFICATION OF CONTENT STANDARDS

Because performance standards are more specific than content standards, they help to clarify and explain the kind of instruction and learning intended by the content standards. By including real examples of student work—examples that meet the standard and examples that fall short—performance standards help teachers, students, policy makers, and members of the public understand the substance of curricular goals and the level of rigor intended. Figure 4 provides an example of a high-quality student response to an eighth-grade mathematics

[15] C. Gentile. 1992. *Exploring New Methods for Collecting Students' School-Based Writing: NAEP's 1990 Portfolio Study*. Washington, D.C.: National Center for Education Statistics.

Figure 4. **A sample assessment task and high-quality response that could serve to exemplify the problem solving and communication standards for eighth-grade mathematics.**

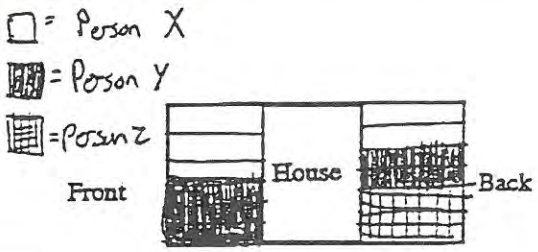
ALL IN A DAY'S WORK

Your neighbor hired you and two of your friends to rake leaves. The house has a back yard and a front yard that are about the same size. The neighbor agreed to pay the three of you $60 for the entire job.

On the day of the job you and one friend arrived to start the job at 9 am. By the time the third friend came, the front yard was finished. All three of you finished the back yard together.

How should the money be split between you? Each person must be paid based on the amount of the yard raked by that person.

Justify your solution in two ways. In one of the ways, use the following sketch.



1. Explain your first way:

I broke the yard into 12 segments, 6 in back and 6 in front. Since person X and Y did half the front yard each, they would have done 3 segments each. In the back Person X, Y, and Z all did a third of the yard, or 2 segments. I divided 12 segments into $60 so each segment raked is worth $5. Person X did 5 segments and gets $25. Person Y did the same, and also gets $25. Person Z did 2 segments and gets $10.

2. Explain your second way:

Each yard is worth a total of $30. Person X did $\frac{1}{2}$ of the work in the front and gets $15 for that. Person Y did $\frac{1}{2}$ of the work in front and gets $15 too. All three did $\frac{1}{3}$ of the work in back and get $10 each. Person X gets $25 total, Person Y gets $25 total and person Z gets $10 total

Reprinted from *Wisconsin Student Assessment System Performance Assessment Sampler*, Bulletin No. 95210. Wisconsin Department of Public Information, January 1995, pp. 9-14. Used by permission.

problem. The task itself is more challenging than traditional algebra word problems, especially because students must generate more than one solution and explain their reasoning. The selected student response illustrates the kind of reasoning and ability to communicate that can be expected of a proficient or advanced eighth-grader. Together, the task and student work sample give meaning to the more general curricular statements that students should be able to "use problem-solving approaches to investigate and understand mathematical content" and "model situations using oral, written, concrete, pictorial, graphical, and algebraic methods."[16]

Accountability and certification

In some state assessment programs, the purpose of the assessment is to report on the level of educational achievement for the state as a whole, for districts, and for schools. When performance standards are used to evaluate parts of the educational system, they serve an accountability purpose. In other states and districts, performance standards and assessments are used to determine whether individual students have reached a certain level of achievement, such as high-school graduation requirements. If results are reported for individual students, then standards and assessments serve a certification purpose.

Accountability and certification uses are distinct from the exemplification use because both require that performance standards be explicitly tied to an assessment that adequately represents the content standards. Moreover, performance standards for accountability or certification must be articulated in the form of rules for translating assessment results into student achievement categories. For example, once a complete assessment in American history has been developed reflecting both required content knowledge and historical inquiry skills, then a scoring rule or standard might be that students must earn 75 percent of the possible points, or must earn 80 percent on the core areas and 60 percent on the specialty sections. Although the accountability use does not always require that individual student scores be reported, the assessment system nonetheless must count the number of students in each achievement category (off the record) in order to report what percent of students are advanced, proficient, basic, and so forth.

---

[16]NCTM, op.cit., pp. 75, 78.

KEY ISSUES

Educators and measurement specialists have a great deal of experience setting passing scores on tests. However, to the extent that the intention of some reformers is to change curricular expectations substantially, there are new technical and practical problems to be resolved in deciding how good is good. In this section, we discuss the difficulties in setting performance requirements for yet-untried assessments, the need for benchmarks leading toward attainment of performance standards, and the possible legal distinctions between reward and punishment uses of standards.

ACCOUNTABILITY AND CERTIFICATION REQUIRES GREATER FIELD EXPERIENCE

Deciding on classification rules for individual students—How much is good enough to be advanced? How much is good enough to be proficient?—raises both substantive and technical problems that are not easily solved until new curricula and assessments are tried out in real classrooms. Some might argue that adequate models already exist for teaching and then assessing new content standards. All we have to do, they say, is extend, to a much larger segment of the student population, the curriculum and performance expectations once reserved for an elite subgroup of students. However, to the extent that content standards include new kinds of content even for the most advanced students, there are no models in place for what can reasonably be covered, with some depth of understanding, in a given school year.

Some experience with content that is both reasonable and challenging is necessary in order to answer the "how much" question; we would not, for example, arbitrarily set eighth-grade standards for fourth-graders. Still, some upgrading is under way. Obvious examples of recent changes in mathematics content are the increased attention to geometry, measurement, probability, and statistics in the early grades. Here, performance standards are not already established, and it is not known how serious attention to these areas as well as to problem solving might modify existing performance standards even in familiar areas such as operations and computation.

In the past, the most frequent experience with standard setting, at least for large-scale assessment programs, was for minimum-competency examinations. Standard setters, reviewing questions all designed to measure mastery of low-level, basic content, merely had to agree on what percentage of this content it was acceptable to miss, and very often set the standard at 70 percent or 80 per-

cent correct. When assessments are redesigned to measure more challenging content, it does not follow that a cutscore for advanced performance can be set by applying minimum-competency procedures near the top of the score scale. Most specifically, it does not follow that 70 percent or 80 percent of advanced tasks should be performed correctly for a student to be considered advanced.

Individual assessment items and tasks are no longer interchangeable in their degree of difficulty and familiarity, the way that basic-skills items have been. Many tasks should, in fact, require highly specialized and in-depth knowledge. Unless and until there is a universally accepted and comprehensive course syllabus for individual knowledge domains, it would not be reasonable to expect advanced students to know a high percentage of all advanced content. For example, the Advanced Placement (AP) examination in biology is based on a shared general content outline and a set of key themes.[17] But because each high school teacher selects his or her own readings and in-depth assignments, students are warned that the content of the test must be "so comprehensive that no student should be expected to attain a perfect or near-perfect score."[18] In AP history, the difficulty of the multiple-choice section is deliberately set so that answering about 60 percent of the questions correctly qualifies for college placement. The International Baccalaureate (IB) is another challenging program of studies and examinations. The IB program is explicitly organized to identify both those things that all students must know (the content of the subsidiary exams) and the higher level of knowledge students must demonstrate in their self-selected areas of specialization. Arriving at the performance standards that have been established for the AP and IB examinations required experience in giving the exams. In the case of the AP program, performance in levels equivalent to college credit can even be tied empirically to performance in equivalent or subsequent college courses.

The exemplification use of performance standards does not pose such difficult technical and conceptual issues as the accountability and certification uses. It only requires examples of the expected quality of performance and not a complete representation of the knowledge domain. Exemplification of content standards could occur using only narrative statements, as in the Australian English profile, without any connection to scoring or judging student responses on specific assessment tasks. Adding examples of assessment tasks and student performances or samples of class-

---

[17]College Entrance Examination Board. 1994. *Advanced Placement Course Description Biology: May 1995, May 1996*. Princeton, N.J.: Author.

[18]College Entrance Examination Board. 1994. *Advanced Placement Course Description History, United States History, European History: May 1995*. Princeton, N.J.: Author, p. 32.

room work representing excellent or proficient performance would greatly enhance the usefulness of content standards in communicating expectations, but it stops short of incurring the technical difficulties involved in specifying how much of the content domain must be mastered and how many of the assessment tasks must be answered correctly. *Given this difference in technical requirements, the Panel recommends that initial emphasis be placed on the exemplification use of performance standards, especially in the early years of implementing a standards-based curriculum.*

By beginning with performance standards that exemplify expectations, it may be possible to build the capacity of the educational system to deliver instruction consonant with the content standards at the same time that more practical real-life experience is being gained with student performance and new assessment systems. Then we may be able to answer the how-much question and set defensible performance standards for accountability or certification purposes.

THE VALUE OF MULTIPLE BENCHMARKS

The most important reason for standards-based reformers to affirm a high-level performance standard for all students is to prevent the self-fulfilling consequences of setting lower expectations for previously low-achieving groups of students. While we agree that instructional efforts should be organized to move all students toward the same content and performance standards, it is likely that some students, especially those with uncorrectable handicaps affecting cognitive functioning, will not be able to meet the new high standards.

What should happen when students don't measure up? In the sections on opportunity-to-learn and assessment, we discuss protections for individual students when the educational system is unable to provide adequate instructional support. But it is in just such situations, where the performance level of a substantial number of students is below the intended standard, that the accountability use of assessment results should focus on progress being made at the lower end of the performance continuum.

If the only reporting metric is the percent of students reaching a high proficiency standard, then school personnel may focus their efforts on students just below the cutoff and ignore students they consider to be out of reach. Multiple assessment levels, such as Title I of the 1994 Improving America's Schools Act's partially proficient, proficient, and advanced categories, or other benchmarks along the performance continuum, make it possible to check on progress for students at all levels of achievement. Attention can then be given to helping

students move up to the next level of proficiency wherever they are on the scale. Furthermore, if lower-level benchmarks are developed substantively and empirically in the same way that a single high proficiency standard should be developed, then they can serve an important pedagogical function by describing how students are progressing along a developmental continuum. Benchmarks are essentially the same as multiple performance standards located along the performance scale.

Some people fear multiple levels of reporting as a capitulation to present inequalities in performance. The Panel recognizes this danger but concludes that reporting accurately for students at the low end of the performance continuum is necessary to ensure their inclusion in the accountability system and thus focus efforts at instructional improvement for these students as well. *Accordingly, the Panel recommends that performance standards include multiple benchmarks along each scale of proficiency in order to demonstrate progress and ensure instructional attention for students at all levels of achievement.*

The defensibility of rewards vs. punishments

Many argue that for high standards to have their desired effect on the educational system and on student learning, there must be visible rewards for meeting the standards. Others insist that individual students should not be punished if the system can be partly to blame for their failure. Goals 2000 intended to provide such a safeguard, in fact, by forbidding the certification of state assessments if they were to be used for graduation decisions or grade promotion and retention within the five-year period following passage of the legislation.

When performance standards and assessments are used for school accountability purposes, individual students are not at risk. In fact, students are likely to benefit if instructional inadequacies are identified and remedied in response to assessment results. Students also benefit when performance standards are used by classroom teachers as a regular part of instruction. Individual grades or scores are determined by the teacher, but individual assessments usually do not result in a major, potentially career-altering event such as failing to graduate. The use of performance standards and assessments to make certain critical decisions about individual students, especially grade promotion and graduation, has however, a special status which sets it apart from all other uses.

There is a substantial legal history regarding the use of tests to make major decisions such as awarding of the high school diploma. In the landmark Debra P. case involving a minimum-competency examination in Florida, the court required

that students receive adequate notice of the intended standards and that the state demonstrate "curricular validity." Curricular validity invokes the same requirement as opportunity-to-learn; in essence, there must be some demonstration that students have been exposed to instruction aimed at the content and skills needed to pass the test.

Members of the Panel are not legal experts; hence the Panel recommends that a national organization or a consortium of states support state and local districts in their standard-setting efforts by sponsoring more in-depth study of these issues. Panel members are qualified, however, to comment on features of the educational context that distinguish current standards-based reforms from minimum-competency testing. First, to the extent that new standards and assessments are true to the rhetoric of high, rather than minimal, standards, it will be difficult for educational systems to prove that adequate instruction is in place for all students. Second, the adequacy of new forms of assessment for certifying individual student performance will also be much more subject to challenge than when assessment is used for instruction or for school- or state-level accountability.

Third, and perhaps most significant, the intended high level of academic standards may raise new fairness and legal issues when tied to high school graduation. If holding a diploma has economic consequences, is it defensible to withhold it from students who possess minimum academic skills but do not meet standards traditionally met only by college-bound students? In some local reform examples, the standards set in mathematics, art, English, and foreign languages have been so high that some teachers themselves could not meet the standards outside their own department. Could such a system survive legal challenge? Would it be fair in a larger moral sense? It is important that educators receive sound advice about how stringency of standards may interact with the use of assessment. For example, does awarding an "honors" diploma have different legal standing from withholding a regular diploma? We believe that giving awards or certificates to acknowledge advanced performance, rather than using standards to determine graduation eligibility, may actually help to preserve the integrity of high-level standards.

Summary of Recommendations for Performance Standards

The Panel's recommendations concerning the development of performance standards are summarized as follows:

Recommendation II.1: Since the technical requirements for performance standards used in accountability and certification are more difficult to meet, initial emphasis should be placed on the exemplification use of performance standards, especially in the early years of implementing a standards-based curriculum.

Recommendation II.2: Performance standards should include multiple benchmarks along each proficiency scale so as to demonstrate progress and ensure instructional attention for students at all levels of achievement.

Recommendation II.3: A study group should be established, by a national organization or a consortium of states, to explore the technical, legal, and fairness issues with respect to the use of performance standards and assessments for student certification purposes.

Opportunity-to-Learn Standards

Definition

Opportunity-to-learn (OTL) standards define the level and availability of programs, staff and other resources sufficient to enable all students to meet challenging content and performance standards. "Opportunity" comprises such things as teachers who are well prepared in their content area, instructional materials and resources adequate to instructional goals, a safe school environment, and courses and instructional activities consistent with more demanding standards of content and performance.

OTL standards had their origin in the "school delivery standards" and "system delivery standards" proposed by the National Council on Education Standards and Testing in January 1992. The concept of OTL standards arose largely in response to concern that it was unfair to hold students accountable for what and how well they were learning in school unless they were provided with the opportunity to learn. Given the existing vast inequalities in resources available to American youth—both inside and outside school—educators, reform advocates, and others worried that higher standards, particularly if combined with high-stakes assessments, would further disadvantage large numbers of already

underserved students. Opportunity-to-learn standards were thus conceived as a way to protect students from being further penalized for inequities in the system. Some proponents also hoped that opportunity-to-learn standards could leverage new spending and equalization of resources among schools.

USES OF OPPORTUNITY-TO-LEARN STANDARDS

In a sense, opportunity-to-learn standards imply an entire set of expectations and measures addressing quality of education that parallel the system of content standards, performance standards, and assessments of student learning. Thus, the term opportunity to learn is used sometimes to mean the broader concept of high-quality instruction and school environment and at other times to refer to statistical indicators (assessments) intended to measure aspects of opportunity to learn. In the paragraphs that follow we discuss several different purposes for developing and then assessing OTL. Professional standards of practice—the kind of substantive guidelines that professionals use to improve their own practice or that other professionals might use on a site visit to a school—are more important for some purposes than quantifiable, statistical indicators.

PROVIDE VISION

One use of OTL standards is to provide a concrete vision for teachers and others not just of good practice, but of all the elements of a supportive educational environment that would enable students to attain achievement standards outlined by the reforms. Most schools and teachers would like to provide high-quality instruction, but many lack relevant experience from which they can draw, particularly for identifying the ways in which their practice needs to change.

OTL standards grounded in research on effective instructional practices and school strategies might provide much-needed guideposts, and help educators reshape practice by making expectations explicit. One example is NCTM's Professional Standards for Teaching Mathematics. This document stops short of a plan of action that teachers can follow, but it illustrates with examples and narrative the types of activities teachers might use to promote student mastery of the NCTM content standards.

GUIDE SCHOOL REVIEW.

A second possible use of OTL standards is to guide school-quality reviews. These reviews might be of the type being piloted in New York and California, where

educators and constituent groups are working on a system of "indicators," or of the type used by accreditation and professional associations. For example, the National Association for the Education of Young Children has developed professional standards of practice for evaluating and accrediting developmentally appropriate early-childhood programs. Professional standards or statistical indicators should address questions such as, What types of schools should provide what types of education to what types of students? Do students from poor families receive the same higher-order-thinking and problem-solving instruction as do students from affluent families? Are instructional resources as available to students in one geographic location as they are to students in another?

As one source of evidence regarding the quality of schooling, student assessments themselves can serve as an indicator of opportunity to learn. Assessment results can help to identify schools that either lack the capacity or the will to deliver high-quality instruction to all their students. These schools can then be targeted for further review and follow-up improvement efforts. This is true with respect to the performance of the school's student body in the aggregate as well as with regard to particular subgroups of students within the school. For example, large numbers of students, particularly those with limited proficiency in English, are currently exempted from state and local assessments used for accountability. Leaving these students out of the accountability system provides us with no indication of the kind or quality of instruction they receive or of its effect. Incorporating these students into the assessment system would provide some of the information needed to help ensure that they do have the opportunity the U.S. has always promised its residents, both immigrant and native born.

Guide system accountability

The third possible use of OTL standards is as a safeguard for students and teachers, and as a basis for school and system accountability. OTL standards could protect students from the negative consequences of high-stakes tests until the opportunity to learn the content of those assessments had been demonstrated. Likewise, OTL standards could protect teachers and other school staff from penalties for failure to meet more rigorous performance standards they were ill equipped to address. For example, it would be unfair and arbitrary to hold a social studies teacher working without a book for every student or up-to-date materials accountable for the same student outcomes as a colleague with full access to appropriate instructional resources.

KEY ISSUES

This seemingly straightforward and reasonable account of the relationship between accountability and opportunity to learn masks enormous conceptual complexity and controversy about the nature and use of OTL standards. The old system of input indicators is inadequate to measure opportunity to learn because those indicators correlate only weakly with student learning, and derive from a research base that used low-level basic skills as outcome measures rather than the more demanding performance standards reformers envision. Fundamental conceptual questions about opportunity-to-learn standards must still be resolved: What is meant concretely by "opportunity to learn"? Who should specify and define OTL standards? What system of indicators can reliably link opportunity to learn and performance standards?

DEFINING WHAT REALLY MATTERS

We lack well-grounded research to identify clearly the most influential OTL elements. To date, research and policy have tended to focus on elements that are easily quantifiable, but that are considerably distant from the actual processes of instruction. Traditional input indicators such as course offerings, teachers' years of experience, and per pupil expenditures show only marginal effects at most on student achievement, when other factors are controlled. Such indicators are only proxies for what really matters in the classroom, and distant from the ways in which teachers and students actually relate together around content.

Small-scale research has documented aspects of OTL that affect the quality of student learning, such as the ways in which learning environments are constructed and enacted, the pre-service and professional development opportunities available for teachers, uses of time and space in school, and so on. These elements of an educational environment that affect teaching and learning capture less readily observed aspects of school. Consider, for example, how to define opportunity to learn with respect to the NCTM standard for mathematics as communication. Never before have we expected elementary-school children to be able to discuss or write about mathematical ideas. Indeed, mathematics educators are still defining what that communication standard means and how to measure it in assessments of student performance. Given this, how can we define and provide a school or classroom environment where all students have the opportunity to achieve this standard? In a classroom where most of the students

are poor and half are immigrants who speak one of five different non-English languages, what additional capabilities (or school resources) must the teacher have to provide opportunity for all her students to achieve this mathematics standard? If opportunity-to-learn standards are to be meaningful for large numbers of American students, they must acknowledge and attend to the complexities both of the learning goals themselves and of the instructional context. *These concerns underlie the Panel's recommendation that OTL standards should focus on those elements of schooling that are directly related to student achievement (rather than relying solely on traditional "input" indicators such as per pupil expenditure or teacher salary), and, to the extent possible, on the enacted curriculum rather than on the official or reported curriculum.*

Measuring OTL

Many of the elements of an educational environment that, in small-scale studies, are shown to affect teaching and learning present serious measurement challenges because they involve aspects of school life that are less readily observable and are often situation-specific. While it is possible to identify extremes of OTL—most especially, instances where opportunities to learn are obviously absent—definition of OTL for the bulk of American classrooms will be exceedingly difficult. Effective instruction for different students in different settings should allow for a range of instructional features. Further, research also shows that the same opportunities seem to operate differently in different settings.

A related issue concerns the problem of measuring curriculum and instruction as actually practiced in the classroom. Schools and teachers have enormous control over the content and opportunities they provide students. The same program component or curriculum can be translated into substantively different practices and occasions for students' learning depending on teachers' beliefs and expertise, or school policies such as ability grouping or course assignment.

While there is evidence that certain pedagogical strategies used in delivering content, as well as curriculum-embedded resources such as laboratories for science instruction, improve results for students, the difficulty for developing OTL standards lies in more specifically defining and measuring elusive and situational elements of instruction. Surrogate measures such as the types of courses offered or the availability of texts, computers, and other instructional resources may provide some information, but they do not reveal how these resources are used or by whom. And they say nothing about the quality of that use.

There are, however, some more direct measures and techniques being developed through research and practice that might lay the basis for reliable quality indicators. School-based self-studies, like the Program Quality Review in California or the self-study and external inspection system in New York State, get closer to this goal. Teacher logs and less labor-intensive surveys, developed for research but being adapted for practice, may prove useful. Since the information obtained by these techniques comes through self-report, and is much more corruptible than student achievement measures, it is unlikely that they will be deemed sufficiently reliable for high-stakes accountability purposes. Nonetheless, they may provide OTL indicators that can be useful for school improvement.

Whatever the validity and reliability problems involved in measuring opportunity to learn, however, it is an inescapable fact that many students, most of whom are poor or are students of color, do not have access to the very basics of education necessary for success in school. Many advocates of reform point out that we do not need to wait for sophisticated measures of opportunity to learn to know that if students do not have books, if teachers are ill prepared, if necessary courses are not even offered, students will not perform to the desired levels. Moreover, there are schools that have been failing their students, and districts or states that have been failing their schools, for decades with virtual impunity. It is not in the best interest of students in these circumstances to postpone all use of OTL standards until current measurement problems are resolved. *Thus, the Panel recommends that tentative OTL indicators should be tried out and evaluated to determine if they are reasonably connected to student achievement. The Panel also recommends that student assessment should not be postponed until OTL standards have been met, but that high-stakes consequences should not be allowed until OTL has been addressed.*

Objections to OTL standards

Critics have raised several key concerns about the use of OTL standards for school or system accountability. One concern is that used in this way, OTL standards will shift the reform focus away from the student and what the student is learning and back to educational inputs. Moreover, they will do so without a strong research base showing what are the most important and effective resources and conditions. Critics fear that, as a result, the reforms are less likely to lead to the desired improvements in student learning.

A related concern is that, to be applied reliably from school to school, OTL standards are likely to become much more prescriptive than is desirable for promoting good professional practice. The more prescriptive the standards are, the

less discretion is available to teachers and other school personnel to design the most effective instructional program based on the needs of their students and on their own professional expertise. Such prescriptiveness is again predicted to hinder improvements in student learning and to signal interference with local determination of best practice.

There is also a fear that OTL standards will simply bog the system down further in a morass of bureaucratic rules and structures, fostering a compliance mentality and detracting efforts from more productive ways to support schools and teachers. A related concern, that OTL standards signal federal intrusion into state and local practices and unfunded federal mandates, has largely been assuaged as the authority for establishing OTL standards has appropriately devolved to the states.

The Panel's recognition of these complex implementation and instrumentation issues leads to a recommendation that the U.S. Department of Education or other national group, as well as state agencies and consortia, support efforts to investigate issues and approaches to ensuring OTL.

SUMMARY OF RECOMMENDATIONS FOR OPPORTUNITY-TO-LEARN STANDARDS

Recommendation III.1: Opportunity-to-learn standards should focus on those elements of schooling that are directly related to student achievement, rather than relying solely on traditional "input" indicators.

Recommendation III.2: To the extent possible, OTL standards should focus on the enacted curriculum rather than the reported or official curriculum.

Recommendation III.3: In the absence of proven reliable and trustworthy technology for measuring OTL, tentative indicators should be tried out and evaluated to determine if they are reasonably connected to student achievement.

Recommendation III.4: Student assessment should not be postponed until OTL standards have been met, but there should be no high-stakes consequences based on assessment outcomes until OTL has been addressed.

Recommendation III.5: The U.S. Department of Education as well as state agencies and consortia should support efforts to investigate issues and approaches to ensuring opportunity to learn.

ASSESSMENT

DEFINITION

Assessment refers to the measurement of what students know and can do. Although assessment, measurement, and testing have the same meaning—i.e., to appraise performance in relation to a criterion—assessment is the preferred term because it connotes a move away from exclusive use of traditional multiple-choice, basic-skills tests. Educators generally acknowledge that new forms of assessment are needed to address challenging content standards.

USES OF ASSESSMENT

In the discussion of performance standards, we distinguished between exemplification uses of performance standards, which might only involve sample assessment tasks, and educator accountability and student certification uses, which require complete assessments. Here, when addressing requirements for assessments, it is important to examine further the distinction between aggregate, system-level uses of assessment and various individual-student uses. The intended use of assessment data has a profound effect on both the technical and practical features of the assessment program. Assessments well designed for one purpose may not be appropriate for other purposes.

SYSTEM-LEVEL MONITORING OF EDUCATIONAL ACHIEVEMENT

The National Assessment of Educational Progress is an example of a large-scale program used to monitor trends in achievement for the nation as a whole and for regions and states. Because there is no single national curriculum, the assessment content must be comprehensive and inclusive of various curricular emphases found in local settings. Individual student scores are not reported; therefore, a broad array of assessment tasks can be devised with each examinee taking only a small portion of the total assessment. Validity of the assessment is judged in

terms of the fidelity of assessment content to national content frameworks, although specific assessment tasks might be drawn from a variety of local curricula. Because individual scores are not reported, individual portions of the exam do not have to be equally difficult for or familiar to all participants. For example, some students might be asked to respond to content not covered in their school, which would raise an issue of fairness in an individual assessment.

System-level monitoring for accountability

State and district assessment programs are also used to report achievement trends for the educational system as a whole. System-level assessments of this type, like national assessments, do not have to yield reliable and valid individual student scores. In contrast to national assessments, however, state and district programs are more likely to be tied to specific curriculum frameworks or to state content standards. This closer alignment between assessment content and instruction allows an increased emphasis on accountability for assessment results. Reporting of results by district, by school, or even by classroom can not only provide information but also hold educators in those units accountable. If assessment results have serious consequences for schools or for teachers, however, the accountability system must have validity for that purpose. For example, if schools are threatened with loss of funding because of poor student performance, then assessment results should be linked demonstrably to poor teaching rather than to, for example, high mobility of the student population.

Student certification

As previously described, student certification requires both a complete assessment, covering the intended content domain, and performance standards defined in terms of what the student must demonstrate on the assessment. Examples of certification uses include high school graduation examinations and Advanced Placement or honors certificates. Obviously, certification requires that assessment results be reported for each student, and because the consequences are potentially much more serious than the use of individual assessments in the classroom, certification measures must meet the most stringent standards for reliability, validity, and fairness.

Instructional benefits

We mention instructional uses of assessment because policy makers often assume that the same large-scale assessments can be used for classroom pur-

poses as well as for accountability purposes or to monitor achievement trends. This is rarely the case. For one thing, the content of large-scale assessments is not tied closely enough to the specific content of classroom instruction. For another, the once-per-year assessment results would not be available at the time each topic was being covered in the classroom.

Ideally, classroom-level assessment with results for individual students is an integral part of everyday instruction. Many assessment tasks—math problems and explanations, a critical essay, devising a controlled experiment, contrasting two historical periods, building a model of a molecule—are indistinguishable from good instructional activities. If developed properly, classroom assessments can be used not only to determine grades at the end of a term but to support learning as it occurs. As we have mentioned before, an advantage of clearly defined performance criteria and benchmarks for progress is that teachers and students can evaluate what has been accomplished and what still needs to be learned. Classroom-level student assessments clearly should reflect the content and performance standards that have been established, but they do not have to meet the same standards for technical accuracy as certification measures because the consequences of errors are not so serious. For example, if a teacher places a student in the wrong reading group, listening to the student read in subsequent weeks will provide new information and the mistake can be corrected.

Compatible or Contradictory Purposes of Assessment

For all purposes, assessments must do a good job of representing content standards. However, many of the substantive, technical, and practical features of assessments can and should vary depending on the intended use of a particular assessment. For example, external assessments used for school and system-level accountability purposes must be comparable across very different school contexts. Traditionally in the United States, this has meant that test materials are kept secure from teachers and that teachers cannot score their own students' work without some kind of external check, because the teachers themselves are being evaluated. In contrast, classroom assessments do not have to meet rigorous technical standards for reliability and validity, tests are not kept secure from teachers—in fact, assessment-like tasks are a regular part of instruction—and teachers and students are intimately involved in scoring and self-evaluation. Moreover, instructional uses of assessment require complete adaptation to local contexts and levels of student learning. For accountability or certification purposes, an examination must be comprehensive even if it means testing content well beyond the students' reach. In contrast, the most appropriate classroom

assessment should focus directly on the zone of the student's learning, to inform the teacher about what the student does or doesn't know.

Key issues

The intention of standards-based reform is to set higher standards for all students, an approach that is fundamentally different from current instructional practice. New kinds of assessments reflecting these new standards are seen as instrumental in effecting the reform. Yet, when students are being evaluated for course grades or high-school graduation, long-standing conceptions of test validity and fairness mandate that students not be assessed on content they have not had a chance to learn. Here we discuss basic issues of assessment validity and fairness including alignment with content standards, the need for multiple modes of assessment, special issues in assessing limited-English-proficient students, and a means for addressing the discrepancy between higher standards and current practice.

Aligning assessments with content standards

At the heart of standards-based reform is the insistence that assessments must faithfully reflect important learning goals. In the past, traditional forms of measurement were thought to be good "indicators" of achievement because they correlated well with independent evidence of student learning. However, they were not synonymous or interchangeable with a comprehensive curriculum. To the extent that traditional assessments measured only a limited portion of desired student knowledge, distortions could occur both in the meaning of the test results and in subsequent efforts to improve learning.

Various terms, such as *authentic, direct,* and *performance-based* assessment, are used to emphasize the idea that assessments must be closely aligned with the intended content standards. Individual assessment tasks should elicit the kinds of demonstrations and applications of knowledge ultimately expected of students; and the complete assessment should represent the extent and range of knowledge expected. A close correspondence between an assessment and the content standards is essential to measurement validity. *This requirement leads the Panel to recommend that assessments should be compatible with and exemplify the content standards.* In addition, assessments that do a good job of portraying the quality and level of performance standards serve to guide and direct educational effort for teachers, students, and their parents.

Providing evidence of validity, reliability, and fairness

It goes without saying that assessments must meet technical standards for validity, reliability, and fairness, and the Panel makes a recommendation that such evidence is essential. Valid measurement means that when schools or students are judged to be proficient or deficient on the basis of assessments, such conclusions are verifiable by independent measures of the same performance criteria. Faithful representation of the intended content standards is a component of validity. In addition, validity requires that the mode of measurement—the format of questions and the particular tasks selected—must allow all examinees to show what they know without introducing any irrelevant sources of difficulty. Fairness and safeguards against bias are closely related to the larger concept of validity. For example, an assessment used to measure mathematical problem solving might involve the use of calculators. However, such an assessment would be unfair to students for consequential uses—even though content standards require calculators for complex problems—if they had not had prior experience working with calculators. Because there are large discrepancies between instruction implied by national content standards and current practice, such instances of unfairness to individual students could arise often. This is one reason that the Panel recommends moving ahead with some uses of new assessments while postponing uses that have serious consequences for individual students.

Reliability refers to the dependability or consistency of assessment results. In the context of performance assessment the most salient issues involve the scoring of open-ended problems and essays and the generalizability of conclusions across assessment tasks. For example, it is well known that a student's apparent writing proficiency can vary to a large degree depending on the specific writing task. Therefore, an assessment with only a few writing tasks may not produce reliable results for individual students.

Given that many new forms of assessment are in the process of being developed and tried out in states and local districts, an analogy has been coined in the measurement field—developing new assessment programs is like building an airplane while trying to fly it. The Panel supports these development efforts and concurs that assessment programs cannot be sequestered in the laboratory until extensive validity studies have been completed. In fact, the most important validity studies can only be conducted in real classrooms and in the context of concomitant changes in instruction. However, as we have emphasized, validity and other technical requirements depend on use.

Depending upon the purpose and potential consequences of a given assessment, only some uses should be pursued without prior and rigorous evidence of validity. System-level monitoring of educational achievement and classroom-level instructional uses are relatively safe applications of new assessment technology; for these purposes it is reasonable to try out assessments and revise them based on experience and evaluation data. Other applications with serious consequences for individual students, or in some cases for schools and teachers, should only be pursued if the assessments have been shown to meet rigorous standards for reliability, validity, and fairness.

Allowing for Demonstration of Proficiency by Multiple Methods

Validity and fairness require that a student not appear competent or incompetent because of the particular mode of assessment. Consider the following examples. One, a young child exhibits poor reading comprehension based on responses to isolated recall questions. Is it valid to conclude that the child is a poor reader if he or she can give an oral retelling of the story? Two, performance assessments in mathematics involve much more extensive written responses. If a student performs poorly, is it because of lack of mathematical knowledge or poor writing skills? Three, in many existing examination systems, in the U.S. and internationally, there is a consistent gender difference in performance on multiple-choice and essay portions of tests. Given equal content knowledge, women do relatively better on essays and men do relatively better on multiple-choice questions. How can assessments be structured so that the result of an assessment is not determined by its format?

New forms of assessment tend to require completion of complex tasks and therefore involve greater integration of skills. This makes it essential that developers of assessments be clear about what is being measured and decide which features of the assessment are relevant or irrelevant to the intended construct. For example, communicating mathematically is an important content standard and should be an integral part of assessment in mathematics. However, if writing becomes so pervasive a requirement of math assessment tasks that students are no longer able to show strength in mathematical reasoning except through writing, then scores may not validly reflect student proficiency.

Multiple methods of measurement—which may include open-ended and extended-performance tasks, portfolios of classroom work, and analytic, multiple-choice questions—ensure that errors or distortions in one mode are detected and offset by other sources of evidence. In large scale measurements like the National Assessment of Educational Progress, where individual scores are not reported,

multiple modes of assessment nonetheless help to ensure that the content domain is represented fully. *For important individual uses of assessment results, the Panel recommends the use of multiple methods and multiple opportunities to demonstrate proficiency as essential for fair and valid measurement.*

Assessing limited-English-proficient (LEP) students

Issues affecting assessment of limited-English-proficient (LEP) students are myriad and are confounded by lack of availability of adequate bilingual programs or instruction in English as a second language (ESL). Any attempt to develop guidelines about the proper treatment of language in assessments must acknowledge that so-called LEP students differ dramatically from one another in their primary-language fluency, exposure to formal literacy instruction in their primary language, degree of English fluency, and exposure to instruction in mathematics and other subjects in either their first or second language.

Because true bilingual programs—where non-native-English-speaking students would learn mathematics, science, and U.S. history in their primary language—are few in number, most LEP students receive content instruction in English. For many students, this creates a hopeless dilemma for fair assessment. Using mathematics as an example, assessing students in their native language would put them at a disadvantage because all of the vocabulary and subject-matter experiences they have acquired are in English. A third-grader who learned how to find the "perimeter" or "area" would not necessarily know what was being asked for in Spanish or Chinese. At the same time, because students' development of English is incomplete, giving them a standard English assessment is likely to underestimate their grasp of the content.

An overarching principle should be that lack of English proficiency should never be mistaken for lack of subject-matter knowledge. Because most LEP students receive content instruction in English, and because one's demonstration of achievement is closely tied to vocabulary and key curricular concepts, LEP students should be assessed in the language of instruction. But special assistance should be provided by bilingual adults to ensure that LEP students understand what is being asked. In some cases students should also be able to answer and demonstrate what they know in their primary language.

The extent of special accommodations would depend on the intended use of the assessment. In large-scale assessment programs intended to monitor achievement trends, it would be possible to provide special second-language-supported

Figure 5. A sample task from a Spanish version of the California Learning Assessment System Grade 4 Mathematics Assessment.

La máquina de ventas automáticas

María quiere comprar un antojito que cuesta 75 centavos de la máquina de ventas automáticas. La máquina acepta solamente monedas de 5¢, 10¢, y 25¢. Mária tiene 7 monedas de 5¢, 5 monedas 10¢ y 2 monedas de 25¢.

Parte 1
Muestra todas las maneras diferentes en que ella puede pagar por el antojito. Puedes utilizar palabras, diagramas, tablas o cuadros.

Parte 2
¿Cuál de las maneras diferentes de pagar utiliza el menor número de monedas? Explica por qué es cierto.

Part 1

5¢	10¢	25¢	total
○○○○○ ○○	○○○○○	○○	$1.35
1	2	2	.75¢
5	5		.75¢
7	4		.75¢
6	2	1	.75¢
3	1	2	.75¢
2	4	1	.75¢
5		2	.75¢
4	3	1	.75¢
	5	1	.75¢

Esto es una grafica de combinaciones que puede usar María para comprar un antojito. Tambien ise unas monedas sueltas que son las monedas que tiene María y las conte y el total de todo fue $1.35.

Part 2

La combinacion menor de las combinaciones que ise en la pagina 4 es la numero uno porque todas las demas son mayores que cinco y la combinacion que yo ise es 1 cinco 2 dieses 2 pesitos y las monedas que acabo de numerar me dieron un resultado de cinco monedas y esa era la menor cantidad de monedas todas eran 6, 7, 10, 11 y muchas mas.

Reprinted from *Students, Standards, and Success.* California Learning Assessment System, Communications Assistance Packet. Sacramento: California Department of Education, 1993. Used by permission.

Figure 6. A sample task from a California Learning Assessment System Grade 4 Mathematics Assessment.

The Vending Machine

Maria wants to buy a 75-cent snack from a vending machine. The machine takes only nickels, dimes, and quarters. Maria has 7 nickels, 5 dimes, and 2 quarters.

Part 1
Show all of the different ways she could pay for the snack. You may use words, diagrams, or charts.

Part 2
Which of your ways uses the fewest number of coins? Explain why this is true.

Part 1

1. 2 quarters
   2 dimes
   1 nickel
2. 2 quarters
   1 dime
   3 nickels
3. 2 quarters
   5 nickels
4. 1 quarter
   5 dimes
5. 1 quarter
   4 dimes
   2 nickels
6. 1 quarter
   3 dimes
   4 nickels
7. 1 quarter
   2 dimes
   6 nickels
8. 5 dimes
   5 nickels
9. 4 dimes
   7 nickels

Combination of Coins

Part 2

The first way uses the fewest number of coins. Ways 2-9 use 6-11 coins, way number 1 uses 5 coins. To me it is weird how 5 coins and 11 coins could equal the same amount of money. After I looked over it, I understand

Reprinted from *Students, Standards, and Success.* California Learning Assessment System, Communications Assistance Packet. Sacramento: California Department of Education, 1993. Used by permission.

administrations of the assessment to a carefully selected sample of LEP students. Data from these students could then be used to estimate the performance of the larger population. For classroom purposes, bilingual aides could be trained to help monolingual teachers assess students' comprehension of key ideas and concepts. The most intractable problems occur, of course, for high-stakes accountability and certification uses of assessment, where standardization must be assured and the consequences of misrepresenting what students know are potentially very serious.

A separate issue from the assessment of content knowledge is the need to assess the language and literacy development of LEP students in both their first and second languages. Although in many instances this may not be feasible, because of the many languages involved and the cost of developing special tests and providing examiners, it should be a goal in the same way that English language arts and foreign-language proficiency are assessed for native English speakers. In some special cases, parallel assessments could be used for both purposes. For example, at more advanced stages of English language development non-native speakers could be expected to participate in English language arts assessment and could be assessed in their native language literacy by taking the same examination designed as a measure of foreign-language proficiency for native English speakers. Before this kind of assessment can be developed, however, there must be clearly articulated content standards for ESL. Without a well-defined continuum of developing proficiency and performance standards (i.e., what degree of language proficiency should be expected at each stage of development), it is not possible to create new, or to know when it is appropriate to use existing, English and foreign-language measures.

In sum, the Panel recommends that LEP students should be assessed in mathematics, science, and history in the language of instruction, and that their language and literacy development should be assessed in both languages.

Making assessments count while guarding against unfairness

Teaching to achieve new, higher content standards requires that teachers develop new strategies and classroom activities, develop or adopt new materials and new assessments, and, in some cases, acquire new content knowledge themselves. For example, in schools where elementary-grade science has been either nonexistent or focused primarily on learning vocabulary, teachers would have to learn how to introduce young children to hands-on science experiments. The Panel's position

on the development of assessments makes it clear that, for some purposes, it is legitimate to begin using new assessments before all technical problems are resolved and before reforms in curriculum and instruction have been fully implemented.

Although certain accountability uses of assessment results have potential for improving instructional practices, other uses with serious consequences for individuals are not warranted while the validity of the assessments is still being evaluated, and while it remains unclear how much the system and the individual alone bear or together share the responsibility for poor performance. Historically, students who achieve poorly because of inadequate instruction and limited opportunities have been further penalized—as a result of poor test performance—by being assigned to even more limited and dead-end learning environments.

Given this negative history, we find it unacceptable to use new forms of assessment to make significant educational decisions for individual students if the assessments have uncertain validity, if the system itself cannot yet deliver effective instruction, or if proposed remedial interventions have not been investigated for effectiveness. Similarly, other uses with very serious consequences for the system, such as placing schools in receivership or denying merit pay to teachers, should not be implemented at least until some evidence of validity has been gathered. This includes evidence of reliability and fairness in the way achievement is measured; evidence that assessment results genuinely reflect the quality of instruction provided by individual teachers and schools; and evidence that any imposed intervention—receivership or a program of merit pay, for example—has the intended positive effects.

The technical adequacy of new assessments and their relationship to instruction should be established first in low-stakes contexts. Then, while it may be necessary to try out a particular high-stakes use before its effects can be evaluated, there should be a specific plan for evaluating effects as part of provisional implementation. In addition, a thorough-going evaluation, while not excusing poor performance, should consider fairness issues such as the extent to which school and teacher results are associated with inadequate funding or lack of support for retraining to meet the standards.

The Panel envisions a continuum of potential consequences, positive and negative, that should determine both validity requirements and consideration of opportunity to learn. In traditional measurement practice, the use of a test

affected the technical demands for the test's reliability and validity. Placing a child in special education—a decision with great potential for serious negative side effects—required multiple indicators and the highest degree of measurement validity. Validity standards for special-placement decisions included a requirement for evidence of placement effectiveness in the form of "aptitude-treatment-interaction" data showing that students were better off for being in the special placement.

Classroom tests used by teachers to make day-to-day instructional decisions have never had to meet strict standards for technical validity because the consequences of mismeasurement were not so high. In the same vein, in the context of implementing performance assessments, different uses of examinations even for assessing individuals have different consequences, which should be taken into account in deciding what degree of validity evidence is needed. For example, failing to earn college credit on an Advanced Placement examination has negative consequences for an individual, but they are in no way comparable to denying him or her a high-school diploma. Differential requirements for validity and opportunity-to-learn evidence depending on consequences explain why the Panel recommends going ahead with some assessment applications and not others.

In making its recommendations regarding the use of standards and assessments for high-stakes individual decisions, the Panel has weighed the potential benefits of high-stakes consequences against the potential harm. Advocates for high stakes believe that students will not be motivated to study hard and perform well unless the results of examinations matter. Although research evidence to evaluate this claim is sparse, one case in point would be findings from follow-up studies of the effects of the NCAA's Proposition 48.[19] Initially, fewer African-American students received athletic scholarships as a result of Proposition 48 test score and academic course requirements, but eventually more African-American students received scholarships and a higher percentage graduated from college than before Proposition 48. Thus, the presence of the "standard" appears to have had the effect of improving individual performance.

Any evidence showing positive effects from high-stakes uses of standards for individuals must be weighed, however, against the more substantial body of evidence, already described, that documents the negative effects of high-stakes testing programs, which disproportionately affect poor and minority students. The Panel's position, therefore, is that important improvements in teaching and

[19] S.P. Klein and R.M. Bell, "How will the NCAA's new standards affect minority student athletes?" *Change*, in press.

learning can and should be made through system-level accountability uses of standards and assessments, while individual uses with serious consequences should be postponed until opportunity to learn can be addressed. As stated in the discussion of performance standards, we believe that this will protect the integrity of the standards as well as safeguard individual students.

That said, the Panel has acceded to consequential uses of such assessments as Advanced Placement tests and college entrance examinations, which we acknowledge may still unfairly affect the life chances of students. For example, a recent study of 55 high-achieving Latino students from a large sample of diverse secondary schools found that none of the students had been advised that they should prepare for college entrance exams, or that these exams held important consequences for their academic futures.[20] In contrast, test preparation is an implicit part of the curriculum in affluent high schools. Given documented gross inequities in educational resources, some might argue that there should be no consequential uses of assessment whatsoever until OTL has been proven.

The Panel's recommendations—making assessments count for school accountability but postponing the most serious individual consequences, making requirements for validity and OTL evidence contingent on the degree of consequences, favoring some individual uses of assessments while arguing strongly against others—reflect our attempts to strike a balance between the leverage that can be gained by imposing high-stakes consequences and the need to protect students who have been underserved and unfairly treated by the educational system.

Summary of Recommendations for Assessments

Recommendation IV.1: Assessments should be compatible with and exemplify the content standards.

Recommendation IV.2: Assessments should be accompanied by evidence of validity, reliability, and fairness.

Recommendation IV.3: Assessments should allow for demonstration of proficiency by multiple methods.

[20] P. Gándara, "SAT Scores and Latino Students: How Much Do They Matter?" In preparation.

Recommendation IV.4: Limited-English-proficient (LEP) students should be assessed in challenging subject matter in the language of instruction, with special assistance in their first language to ensure they are able to show what they know. The language and literacy development of LEP students should be assessed in both their first and second languages.

Recommendation IV.5: Assessments should be visible indicators of school effectiveness. However, external assessments should not be used to make decisions with severe negative consequences for individual students or teachers without evidence of validity and consideration of opportunity to learn.

PART IV

CONCLUSION

Standards-based education reform is still very much a work in progress. It confronts many complex technical and political implementation problems, solutions to which may only be reached through the process of continued development, rather than before. The Panel acknowledges this less-than-ideal situation, but nonetheless recommends proceeding with standards-based reform, acting on the practical and research-based experience we do have, and remaining aware of both existing and emergent key implementation problems and pitfalls.

Here we summarize the central implementation issues and suggest strategies for addressing them. We reiterate several of our specific recommendations that follow from a "proceed with caution" stance.

IMPLEMENTATION CONCERNS

While promising examples of standards-based curricula currently exist to guide the reform (in mathematics and some areas of science, for example), a fully developed array of content and performance standards within and across knowledge domains does not.

In some content areas, such as English language arts and history/social studies, professionals continue to struggle with the difficult task of agreeing on core intellectual skills and knowledge. Further, some educators emphasize broad concepts and abilities, while others want to see factual knowledge specified in some detail. Work has barely begun on content and performance standards for new interdisci-

plinary content areas. Although social studies teachers in the middle grades have had seven or eight decades of experience with interdisciplinary teaching, it is not known whether interdisciplinary curricula at the high-school level can be pursued without jeopardizing discipline-specific knowledge that many educators consider crucial.

Links between content, performance standards, and assessments need to be established for most subject areas. The kinds of student performance and assessments that are consistent with the visions of learning that underlie new standards are still in the process of development.

Questions and concerns abound about what the new standards will look like in practice. Will they be feasible within the context of an instructional day or a school year? Since subject-area standards are being developed in isolation from one another, will they collectively make sense from the perspective of an elementary-school teacher who teaches all subjects? Little thought has been given to the political and professional difficulties inherent in trying to coordinate standards-development efforts.

Another critical implementation issue concerns the capacity of states, districts, schools, and teachers to develop and implement more challenging and meaningful standards. Educators need both knowledge and expertise to proceed, as well as substantial resources for putting into practice the changes demanded by the reforms.

Some teachers and schools have made significant and important progress toward constructing educational environments in which all students learn at high levels. Research provides powerful examples showing that students traditionally unsuccessful and unengaged in school can become good problem solvers, writers, and learners when given support to do so. These examples underscore the benefits many see in standards-based reform and can inform efforts to build capacity throughout the education system.

Getting Started with Caution

None of the above issues can be resolved easily or quickly. Nevertheless, this Panel advises educators to "get started," and to tackle the critical issues of equity and teachers' professional ownership of the reforms.

A cautious beginning on standards-based reform has both strategic and substantive rationale. To wait until fully developed standards and assessment instruments are available in all knowledge domains, and capacity to implement them exists in all areas of the country, as opposed to acting on what we do know, would cheat many American students out of beneficial changes standards-based reforms could foster in their schools and classrooms.

Moreover, we do have knowledge from which to proceed. We know how to provide more challenging instruction and effective educational settings—if not for all students in all schools, at least for some students in some schools. Many teachers have grappled effectively with new concepts of content, pedagogy, and classroom relations implied by the reform. And some knowledge domains have developed assessments that support the vision of learning conveyed by standards-based reform. The Panel believes much can be learned from states and localities already trying out any one set of standards or a combination of standards, grappling with capacity-building questions, and confronting the politics of standards.

Furthermore, waiting until all practical and technical problems are resolved would deny the necessarily developmental and evolutionary nature of the standards-development process. Content and performance standards, as imagined by reformers, will never be "finished" as an intact product to be "installed."

Several of the Panel's specific recommendations follow from our conclusion that standards should be implemented before critical conceptual and technical issues are resolved. For example, were there to be a national advisory or certifying agency, the Panel recommends that multiple content standards should be allowed, representing different organizing conceptualizations of a subject area, so long as each set of standards meets criteria of rigor and professional defensibility. Each endorsed approach should, if it is used, be evaluated in terms of student learning consistent with its animating purpose.

The Panel's recommendation for multiple benchmarks on a continuum of achievement also reflects our belief that high standards for all students cannot be instituted overnight. Therefore, it is essential that students who fall short of the highest standards not be left out of the reporting system or lumped together as unsuccessful. In fact, data collected to assess the effects of the reform should be gathered from students all along the performance continuum.

The Panel's recommendations also acknowledge that certain uses of standards are "safer" and are warranted even without fully developed technical knowledge,

e.g., performance standards used for exemplification purposes and for added endorsements or honors certificates. More high-stakes applications, with potentially serious consequences for students, should be prohibited until major technical problems are resolved and reforms of educational opportunity are more fully in place. We also argue that the seriousness of negative consequences should be weighed to determine the extent of validity and opportunity-to-learn evidence needed to support the use of assessments.

Building Capacity

Professional development

Moving forward with standards-based education reform requires both more and different opportunities for educators' professional development. Much of current classroom practice flows from two concepts: content as a body of facts and teaching as the delivery of information. Staff development in the past has been oriented primarily toward presenting teachers with new techniques in one-day in-service sessions. These sessions assume that teachers will assimilate new techniques into an existing system of ideas about pedagogy and subject-matter knowledge. For most teachers, accomplishing more significant changes will not happen simply by adding new techniques to their current ideas about teaching and learning. More and different opportunities are needed for teachers and administrators to learn the skills and perspectives fundamental to the success of standards-based reform, which considers knowledge as a dynamic product of teachers and students working together.

Research and evaluation

Educators have incomplete knowledge of and experience in the most effective ways to develop standards and assessments and to change the educational system to ensure that all students have the opportunity to reach the standards. Efforts under way at local, state, and national levels can provide important formative information on effective practices, disappointments, and promising developments, but only if accompanied by sound evaluation. Well-designed evaluations will allow educators to address critical implementation concerns and to identify common strategies for proceeding effectively with standards-based reform.

COMMUNICATION AND FEEDBACK

Building capacity requires sharing information, as well as gathering it. Communication about standards-based reform can increase both the effectiveness of professional staff as well as the awareness of the general public about the role and nature of standards.

Proceeding with standards-based reform can be a process that educates participants about such important issues as what it takes to get teachers and students engaged in more rigorous subject matter, how standards can support practices that will equip students to achieve deep understanding of subject matter, and how to provide adequate opportunity to learn for all students.

Standard setters can learn both from critical feedback about their own activities as well as from the efforts under way elsewhere. Networks for communicating about developments around the country are essential to moving forward effectively. Strategies are also needed to encourage conversations among educators within the same and across different subject areas. Such conversations will exemplify the dual benefits of standards: generating agreement and stimulating debate about what students should know and be able to do.

A NATIONAL ORGANIZATION

Although a federally sponsored agency seems neither feasible nor wise in the current political context, a national body established to inform the standards-based efforts across different disciplines and locations could assist with both implementation and capacity-building issues. Such a structure could take a variety of forms. It could be a consortium of states, or an amalgam of public and private interests; it could sit entirely in the private sector.

A national organization to support standards-based reform could serve a number of functions. As a clearinghouse for examples of standards under development or in use, in the United States or elsewhere, it could stimulate and help organize a national conversation focused on many of the issues outlined above. It could also provide tools and feedback to those at state and local levels who are more directly involved in standards development, adoption, or implementation. A national body could be a rich source of data on standards and standards-based reform that could promote system-wide learning on the very issues this Panel has identified as most troubling. An organized sharing of experience and an iterative

process of review, evaluation, and feedback on a national level may help to structure and promote the reflection needed to enable states and districts to learn from one another as they struggle with similar technical, institutional, and political problems.

Refrain: The Promise and Challenges of Standards-Based Reform

The mission of standards-based reform is to improve the educational achievement of all students by defining more demanding academic standards for what and how students learn. Although the Panel does not endorse the idea that setting standards will in and of itself improve public education, we agree that the process of developing standards and continued attention to their attainment can be an important means of focusing the efforts and resources of the educational system and its participants—teachers, students, parents, administrators, policy makers, and members of the community.

Critics have been apprehensive that national standards would nationalize the curriculum and undermine professional and local responsibility for student learning. Their greatest fear—which persists whether or not the development of standards and assessments is federally sponsored—is that lack of support in reaching high standards will further victimize students already harmed by gross inequalities in the educational system.

The Panel recommends a cautious, learn-by-doing approach to the implementation of standards and accompanying assessments. Individual students and teachers should be protected from imperfect and incomplete reforms at the same time that evaluation data are used to revise the standards and assessments, and to improve curriculum and learning resources. Ultimately the goal of high standards is to continually improve the educational system—to learn what opportunity to learn means and how to provide it—so that more and more students can master challenging curricula and lead productive lives.

GLOSSARY

ASSESSMENT. Measurement of what students know and can do.

BENCHMARK. A concrete statement of skills and knowledge to be demonstrated at a specific performance level. Benchmarks are located on a performance continuum and are used as checkpoints to monitor progress from one level to the next. Some benchmarks are synonymous with grade-level performance standards, which can monitor progress from grade to grade. Benchmarks also can be used within a grade level to measure progress toward a high-level proficiency standard.

CONTENT STANDARDS. Broad descriptions of the knowledge, skills, and understandings that schools should teach and students should acquire in a particular subject area.

GOALS 2000. The Educate America Act enacted by Congress in 1994, which provided a list of ambitious goals intended to improve education for all students. Goals were established for school readiness and student mastery of challenging subject matter—including a goal that American students be first in the world in math and science by the year 2000.

HIGH-STAKES TESTS. Assessments that carry serious consequences for students or for educators. Their outcomes determine such important things as promotion to the next grade, graduation, or college admission, and, often, teacher or school "report cards."

NATIONAL COUNCIL ON EDUCATION STANDARDS AND TESTING (NCEST). Created by Congress in 1991 to consider whether there should be national standards and examinations.

NATIONAL EDUCATION GOALS PANEL. Created originally by President Bush and the National Governors' Association to report on progress toward the nation's education goals, and enacted into law by Goals 2000: Educate America Act.

National Education Standards and Improvement Council (NESIC). Authorized by Goals 2000: Educate America Act to develop criteria and to review national content and performance standards as well as state assessment systems and opportunity-to-learn standards submitted on a voluntary basis. Because of recent changes in Congress, appointments to NESIC have never been made.

Opportunity-to-learn standards. Definitions of the programs, staff, and other resources sufficient to enable students to meet challenging content and performance standards.

Performance standards. Concrete examples and explicit definitions of what students should know and be able to do to demonstrate proficiency in the skills, knowledge, and understanding framed by the content standards.

Standards-based reform. Setting standards of performance in academic subject areas as a means of improving the substance of school curricula and increasing the motivation and effort of students, teachers, and school systems and thereby improving student achievement. The reform assumes high standards for all students and has two components: challenging standards that set out what students should know and be able to do, and an accompanying agenda for educational equity.

Validity. The accuracy of a test or assessment, that is, the extent to which interpretations and decisions based on test scores are warranted and supported by independent evidence.