DOCUMENT RESUME

ED 401 333 TM 026 132

TITLE	The Status Report of the Assessment Programs in the United States. State Student Assessment Programs Database School Year 1994-1995.
INSTITUTION	Council of Chief State School Officers, Washington, D.C.; North Central Regional Educational Lab., Oak Brook, IL.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	SSAP-AR-96
PUB DATE	May 96
CONTRACT	RJ96006301
NOTE	43p.; For a related document, see TM 026 133. Cover title varies: "The Status of State Student Assessment Programs in the United States. Annual Report, May 1996." Some tables contain filled-in print. In Appendices, "State Student Assessment Programs Database Order Form" not in the document received by ERIC and is unavailable.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Accountability; Constructed Response; Criterion Referenced Tests; *Educational Assessment; Educational Improvement; Elementary Secondary Education; Multiple Choice Tests; National Surveys; Norm Referenced Tests; *Performance Based Assessment; Program Evaluation; State Programs; Test Construction; *Testing Programs; *Test Use
IDENTIFIERS	*High Stakes Tests; Improving Americas Schools Act 1994 Title I; Large Scale Programs; Test Directors

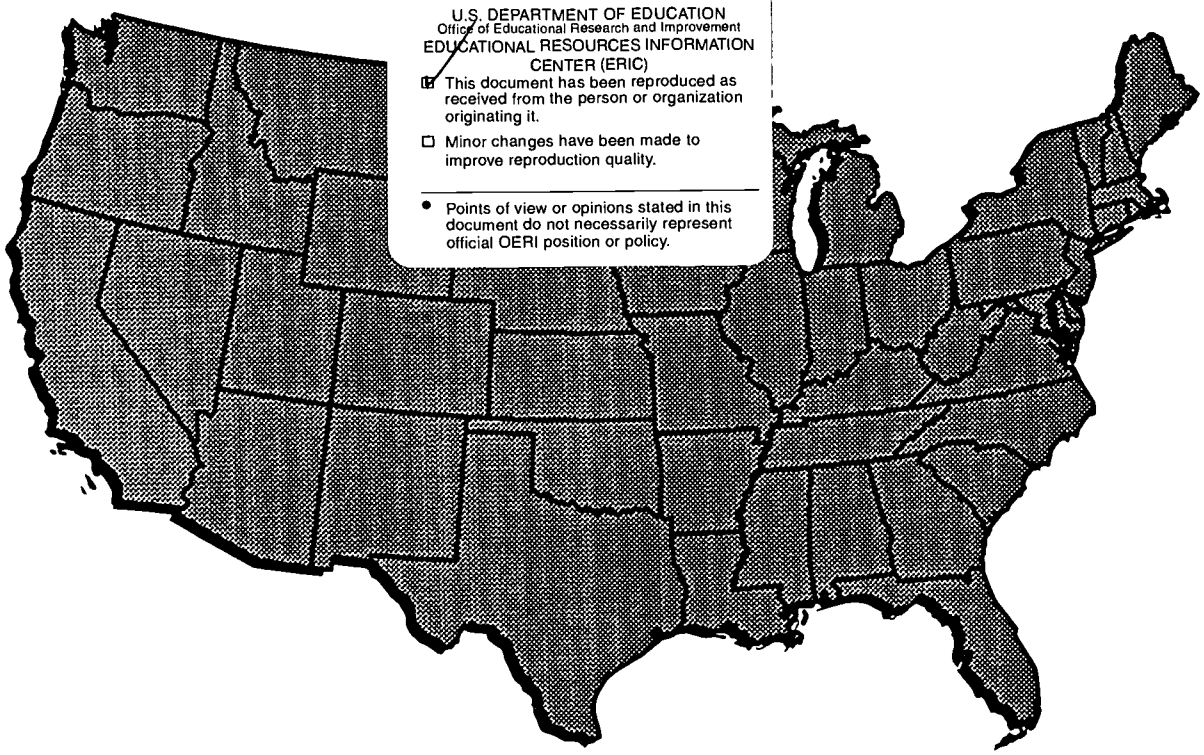
ABSTRACT

The Association of State Assessment Programs, an informal association of state assessment directors, began collecting information about large-scale assessment programs in 1977. This report is a continuation of that effort that is currently conducted by the North Central Regional Educational Laboratory and the Council of Chief State School Officers. The annual survey asked state test directors to comment on assessment programs, including nontraditional assessments and Title I assessment and evaluation. The results of the survey completed by the 50 states are presented. Statewide assessment programs are found in 45 states, and 2 others have temporarily suspended their assessment systems as they design new ones. Thirty-two states have at least two components to their programs. An approximately equal number of states report the use of multiple-choice and nonmultiple-choice assessment types, and an approximately equal number of states use norm-referenced and criterion-referenced tests and writing samples. Performance testing is used more often than constructed, open-response testing, and portfolio assessment is used in only a few states. Most states use their assessment results for two to four purposes, with improving instruction, school accountability, and program evaluation the most common. The tensions that exist when assessment is used for accountability and instructional improvement cause difficulty for those who design and implement these programs, and these tensions are exacerbated by placing negative consequences on poor performance, thus increasing the stakes for schools and students. An appendix presents a survey summary table. (Contains 26 charts, 3 tables, 4 figures, and 1 appendix table.) (SLD)

TM

The Status of State Student Assessment Programs in the United States

Annual Report, May 1996



NCREL
North Central Regional Educational Laboratory

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

One Massachusetts Ave. NW, Suite 700
Washington, D.C. 20001-1431

Gordon Ambach, Executive Director
Barbara Clements, Acting Director, State Education Assessment Center

The Council of Chief State School Officers (CCSSO) is a nonprofit organization of the 57 public officials who head departments of public education in every state, U.S. territory, and the District of Columbia. CCSSO seeks its members' consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public. Because the Council represents the chief education administrator in each state and territory, it has access to the educational and governmental establishments in each state, and the national influence that accompanies this distinct position. CCSSO forms coalitions with many other educational organizations, including those organizations that are active in assisting the states and the nation in setting high standards for their students and those that assess the performance of students against these high standards.

The State Education Assessment Center provides a central clearinghouse to improve data acquisition, monitoring, and the assessment of education. More recently, the State Collaborative on Assessment and Student Standards (SCASS) was formed to network states and other groups to develop prototype and complete assessment components for a variety of content areas. Projects are taking place in a number of areas. The goal in all of these projects is to encourage the development of higher quality student assessments at lower cost to the states. The Council also supports the Association of State Assessment Programs (ASAP), an informal technical assistance network of the assessment staffs in the states.

NORTH CENTRAL REGIONAL EDUCATIONAL LABORATORY

1900 Spring Road, Suite 300
Oak Brook, IL 60521-1480

Jeri Nowakowski, Executive Director
Bill Quinn, Director, Evaluation & Policy Information Center

The North Central Regional Educational Laboratory (NCREL) is a not-for-profit agency funded through multiple government and commercial grants and contracts. Through the application of educational research and development, NCREL works to strengthen and support schools and communities in systemic change so that all students achieve standards of educational excellence.

NCREL specializes in the area of educational technology and accomplishes its work through dissemination, policy analysis, and technical assistance, and by leveraging the power of partnerships and networks. Organizationally, NCREL carries out its work through "learning centers," one of which is the Evaluation and Policy Information Center (EPIC). EPIC develops networks of researchers and research users, monitors and tracks state and local reform efforts, conducts rapid-response inquiries that inform and influence policymaking on "hot" education issues, provides resources for state policy seminars and studies, and holds online policy seminars.

NCREL also operates the North Central Regional Technology in Education Consortium and the Midwest Consortium for Mathematics and Science Education.

The Status Report of the Assessment Programs in the United States

State Student Assessment Programs Database
School Year 1994-1995

Linda Ann Bond, Director of Assessment
David Braskamp
Evaluation and Policy Center
North Central Regional Educational Laboratory

Edward Roeber, Director of Student Assessment Programs
Council of Chief State School Officers

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RJ96006301. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any agency of the U.S. Government.

SSAP-AR-96 $9.95

Acknowledgments

Special thanks go to all the State Assessment Directors who make the State Student Assessment Programs Database possible by providing rich information about their assessment programs. Thanks also to the Council of Chief State School Officers who supported us in this effort from its inception.

We could not have managed this project without the tireless efforts of Deb Roeber, who spent numerous hours on the phone "nudging" those who were a little late returning the survey, and Dina Czocher, who followed through to see that all was progressing smoothly. Our thanks to all of the NCREL staff who pitched in to help us complete this project. Lenaya Raack and Stephanie Blaser, editors, and Melissa Chapko, graphic artist, spent many hours putting the document into its final form; and Arie van der Ploeg, database manager, worked with David Braskamp to design the tables and charts.

Table of Contents

List of Charts

List of Tables

List of Figures

Chapter One

Introduction to the State Student Assessment Programs Database

The topic of student assessment generates considerable controversy among educators and members of the public. Some view large-scale assessment programs as a critical element of the reform and change needed in American schools. Two primary reasons for this are (1) assessment can provide direction and motivation to students, parents, teachers, and others to help students learn the skills needed to succeed both in school and in life after school; and (2) assessment programs can help gauge the success of our schools. An indication of the strength of their appeal is the number of states that currently have assessment programs: 45. Of the remaining five states, Colorado and Massachusetts temporarily suspended their assessment programs while developing new ones. Nebraska is at work developing its first assessment program, to be implemented before 1998. Iowa and Wyoming are the only two states that are not presently administering or developing a statewide assessment program.

Those educators and members of the public who view many large-scale assessments with reservations, feel such programs can exert negative pressure on teachers and students. Much of the debate surrounds such issues as the content covered by the assessments, the type of assessment used, how the assessments are scored, and the uses made of the assessment results. But, however viewed, large-scale statewide assessment programs are a fact of life in the United States.

State assessment programs share some common purposes and methods, but they can also be quite different. Differences exist for various reasons—for example, the educational policy climate in the state, the technical quality issues surrounding the use of assessment to make high-stakes decisions, or the status of curricular reform in the state. We need to recognize these differences in order to understand the assessment programs that exist and the options that are available to change these programs.

In addition, we need to recognize the movement in Washington, D.C., to limit the federal role in education by shifting this role to the states. A result of this has been that states likely will have more control over the educational resources provided to their schools. Similarly, states have shifted more responsibility and control to the district and school levels. The price for increased flexibility and control has traditionally been increased accountability and, therefore, increased assessment. Historically, states were locations where lots of assessment activity and experimentation in new forms of assessment occurred. We will be keeping an eye on how these shifts in responsibility will affect state assessment and whether state assessment will continue to play a major role in educational reform.

The Association of State Assessment Programs (ASAP), an informal organization of state assessment directors, began collecting information about large-scale assessment programs at the state level in 1977. The results of the annual ASAP surveys were provided to states in the form of a written summary of each state's assessment program. In 1991 Ed Roeber, ASAP's chairperson, became director of student assessment programs for the Council of Chief State School Officers (CCSSO). A partnership with the North Central Regional Educational Laboratory (NCREL) led to the current form of the State Student Assessment Program (SSAP) database. This report is a result of the fourth year of that partnership.

As the amount of information increases over time, we are able to provide more meaningful information to states because we are able to monitor patterns of change in state assessment

programs. As data collection continues in the future, we hope to sharpen the analysis of change in statewide assessment practices.

The survey annually collects three kinds of information: Part One asks each state to describe what programs exist, who its collaborative partners are, and what it is developing. Part Two of the survey asks each state to describe its efforts in nontraditional assessment and, this year, in state curriculum frameworks and Title I assessment. Part Three asks each state to divide its assessment program into components, or groups of assessments that are used to gather a set of data used for the same assessment purposes. For each component, states explain who is tested, what subjects are tested, and what types of assessments are used. From these particulars, we can build a more detailed picture of what statewide assessment programs look like and how they are attempting to accomplish their state assessment goals. This report is a summary to provide an understanding of what the 50 states are doing and how they are doing it.

Overview of State Student Assessment Programs

This chapter provides an overview of the assessment the states conduct. A tabular overview appears in the Summary Table in the Appendix. The detailed responses for each state to the survey are available in a companion publication, *State Student Assessment Programs Database,* School Year 1994-1995.

Number of States With an Assessment Program

Statewide assessment programs are almost universal. In the 1994-1995 school year, 45 of the 50 states conducted some form of statewide assessment: mandatory, voluntary, or both. As mentioned earlier, of the remaining five states, Colorado and Massachusetts temporarily suspended their assessment programs while developing new ones. Nebraska is at work developing its first assessment program, to be implemented before 1998. Only Iowa and Wyoming report that there is no state-mandated assessment program in place or in development.

Number of Assessment Components Per State

Table 2-1 displays the number of assessment components per state. For our purposes, we define a component as a single assessment or group of assessments that share a common purpose or set of purposes. When we inspect Table 2-1, we notice that there are 32 states that have at least two components in their assessment programs. This indicates that data are collected from a variety of assessment types, for a variety of assessment purposes and consequences, and from distinct grade levels and subjects. This variety is discussed in the rest of this chapter.

Table 2-1
Number of Assessment Components

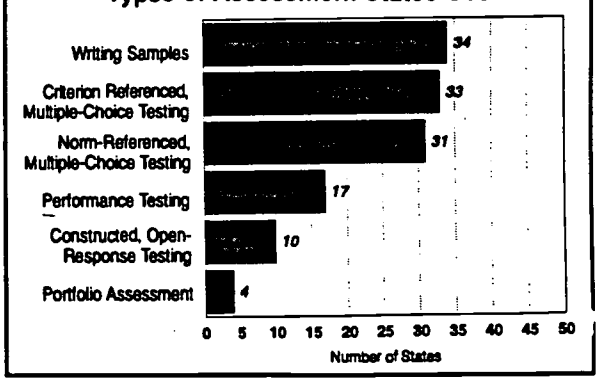
State	#	State	#	State	#	State	#	State	#
AK	2	HI	2	ME	1	NJ	2	SD	2
AL	5	IA	0	MI	2	NM	4	TN	4
AR	1	ID	2	MN	1	NV	3	TX	1
AZ	1	IL	1	MO	2	NY	5	UT	3
CA	2	IN	1	MS	3	OH	4	VA	2
CO	0*	KS	1	MT	1	OK	2	VT	2
CT	2	KY	3	NC	3	OR	2	WA	1
DE	1	LA	4	ND	1	PA	2	WI	2
FL	3	MA	0*	NE	0	RI	2	WV	3
GA	5	MD	3	NH	1	SC	2	WY	0

*CO and MA suspended their statewide assessment programs in 1994-1995.

Types of Assessment Used by States

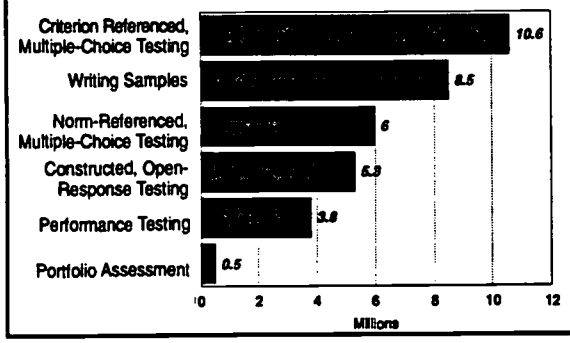
Chart 2-1a displays the number of states that report the use of multiple-choice and non-multiple-choice assessment types (refer to Glossary for assessment type definitions). An approximately equal number of states use norm-referenced (NRT) and criterion-referenced multiple-choice testing (CRT) and writing samples. Performance testing is used more often than constructed, open-response testing, and portfolio assessment is only used in a few states. It appears from this chart that NRTs, CRTs, and writing samples are the most popular types of assessments for states.

Chart 2-1a
Types of Assessment States Use



When we display the minimum[1] number of students tested by each assessment type (see Chart 2-1b), a different picture of assessment type use is presented. The minimum number of students who are assessed by CRTs is more than the number assessed by writing samples and NRTs. These results are different than the relationships of assessment type use indicated by Chart 2-1a. Chart 2-1b reveals that CRTs are the most commonly administered type of assessment, with writing samples and NRTs second and third, respectively. It also appears that constructed, open-response testing is administered to more students than performance testing. This appears to indicate that although constructed, open-response items are administered in fewer states, the student populations in those states outnumber those in which performance testing is administered.

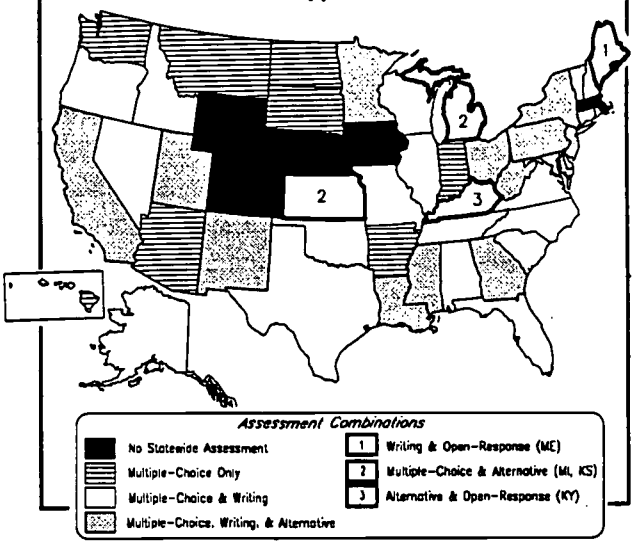
Chart 2-1b
Minimum Number of Students Tested, by Test Type



When we categorize states by assessment type combinations, we gain a more comprehensive understanding of assessment type use than we could in interpreting Chart 2-1a or Chart 2-1b in isolation. Figure 2-1 displays the 50 states categorized into seven different assessment type combinations.

[1]We use the term minimum because the states report the number of students tested per grade level by testing component. There may, therefore, be some students who participate in more than one assessment. To avoid counting students twice, we simply report the number of students tested by the largest component.

Figure 2-1
Assessment Type Combinations



In Figure 2-1, a multiple-choice testing category refers to NRTs and CRTs, and an alternative assessment category refers to performance testing and portfolio assessment. The most common combination, multiple-choice testing and writing samples, can be found in 17 states. This combination can mainly be found in the Southeast, Midwest, and Western United States. The combination of multiple-choice testing, writing samples, and an alternative assessment can be found in 16 states. This combination can be found across the country, but half of them are bunched from Ohio to Vermont. Multiple-choice testing by itself can be found in eight states as well as throughout the country, but half of the states are in the Northwest. It is clear that states more often use a variety of assessment types rather than depend on just one to accomplish different purposes.

Purposes of Statewide Assessments

Most states use each of their assessment components for two to four purposes, as may be seen in Chart 2-2. This situation may create tension for students, teachers, and schools, especially if some of the purposes are seen to be incompatible.

Chart 2-2
Number of Purposes Per Assessment Component

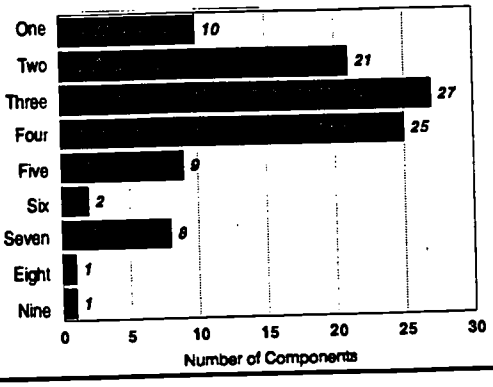
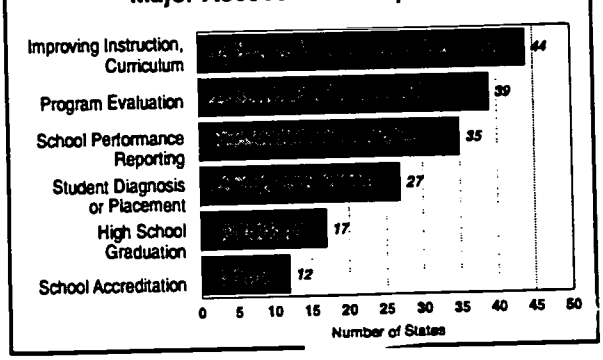


Chart 2-3 displays the six most common purposes states cite for assessing student performance. All are school and student purposes. Only Tennessee reports using one of its assessment components for teacher evaluation (New York allows districts to do so if they choose). With respect to individual student purposes, 17 states use assessments for high school graduation tests and 27 for student diagnosis. The top three overall assessment purposes—improvement of instruction and curriculum, program evaluation, and school performance monitoring (a form of school accountability)—are all school or program-based purposes.

In addition to the information revealed in the chart, we found that 31 states, approximately 70 percent of the states with assessment programs, operate at least one assessment component that has all three of these purposes. Thirty-four states, or 75 percent, have at least one component for which both accountability and instructional improvement are cited.

Chart 2-3
Major Assessment Purposes



As discussed earlier, states depend on assessments to meet many purposes, but some combinations of purposes create more tension than others. Attempting to use a state assessment program for school or student accountability *and* for instructional improvement can be especially problematic. Designing an assessment program to meet high-stakes accountability purposes typically requires standardization of content, administration, and scoring. Accuracy of scoring and standardization of procedure is paramount, particularly if a high school diploma may be denied based on a student's score. Test security is high, with results determined at a centralized scoring center and returned weeks, sometimes months, after the assessment is administered.

The very safeguards that ensure comparability and fairness limit the utility of the results for instructional decisionmaking. For an assessment to be effective as an instructional improvement tool, the results need to be made available almost immediately so teachers can adjust their instruction. Reviewing assessment results over the summer may be helpful for curriculum planning, but teachers need access to ongoing assessment information to modify instructional strategies within the classroom. A classroom-based assessment system, albeit somewhat standardized by virtue of the learning goals being assessed, requires continuous, unobtrusive collection of assessment data, flexible administration, and immediate feedback. Unfortunately, this flexibility, vital to classroom assessment, is typically seen to violate the standardization necessary for accountability purposes.

The state assessment directors acknowledge the difficulty inherent in using one assessment program for both accountability and instructional improvement purposes. However, law and regulation often require they do so. States, therefore, are designing assessment systems that try to capture both sets of purposes in ways to minimize the conflict between them. Some states, such as Illinois, are developing assessment systems with layers at the state and local levels that are aligned to the same learner goals, but used for different purposes. The state assessment serves

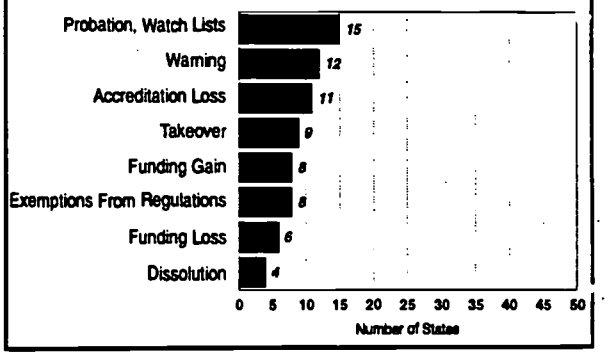
accountability purposes primarily, while the local assessments are used for instructional improvement and school improvement planning. With a new state superintendent in place, however, this system is under review, partly because those at the local level didn't have the resources or the expertise to meet this requirement. Other states, such as Vermont, are combining regionalized scoring of some student assessments with intensive teacher inservice to improve the accuracy of classroom portfolios for use as potential accountability data. The local flexibility of this approach, however, has limited the portfolio's usefulness for accountability purposes. Still others, Kentucky, for example, are auditing the results of local assessments to ensure that scoring guidelines are being applied uniformly across the state to improve comparability of scores. As of this last year, they are also planning to return multiple-choice items to the assessment in order to improve its utility for accountability purposes. Balancing the design of the assessment program to meet both accountability and instructional purposes continues to be one of the major issues facing states.

The most commonly stated goal of state assessment continues to be the improvement of instruction in order to help students meet new, challenging standards. But states seem unsure whether improved assessment content and format or increased accountability will result in the most improvement. They therefore continue to do both, a situation that limits the utility of the assessment program for either purpose.

Assessment Consequences

This year's survey asked also about the consequences of assessment results for schools, staff, and students. Chart 2-4 displays the most common consequences identified for schools. In 15 states, schools that demonstrate low performance on the state assessment are placed on probation or watch lists; in 9 states, schools can be taken over by the state; and in 6 states, they can lose state funding. Clearly, these consequences can be quite severe.

Chart 2-4
Possible Consequences of Assessment Programs for Schools



In some states, schools can suffer multiple consequences. From Table I in the Appendix, we can see that some combination of funding gains and losses, loss of accreditation status, warnings, and eventual takeover of schools are potential consequences in 23 states.

Currently, consequences for school staff are much less common, with two states, Kentucky and North Carolina, reporting financial awards, and one state, Kentucky, reporting financial penalties and probation. New York leaves decisions about any school staff consequences up to local districts.

Consequences for students remain fairly rare also. Five states—Indiana, Louisiana, New Mexico, South Carolina, and Virginia—report basing student promotion decisions on state assessments, and 12 states make student award and recognition decisions based on their assessments.

High school graduation tests, however, are another matter. Figure 2-2 shows the 18 states that conducted high school graduation tests in 1993-1994. As is indicated on the map, most of the high school graduation testing occurs in the south, going across the country from West Virginia to New Mexico[2].

[2]For more information about high school graduation testing, please read the NCREL paper, *State High School Graduation Testing: Status and Recommendations.* (Bond & King, 1995).

Figure 2-2
High School Graduation Testing

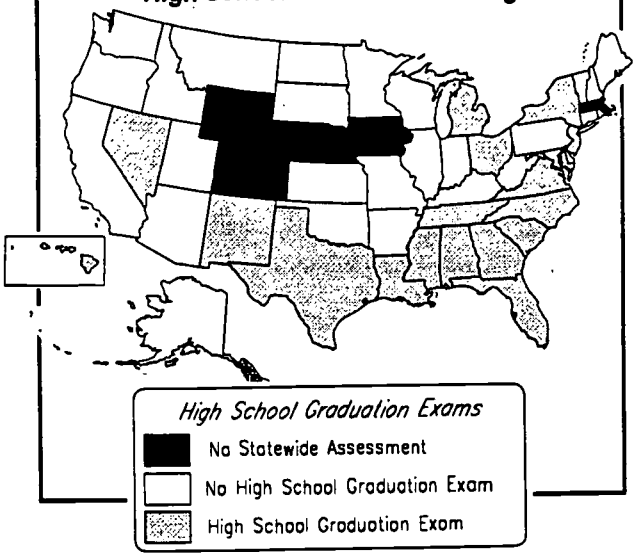


Table 2-2 categorizes the states by the requirements they place on students to graduate from high school, to receive an endorsement on their diploma, or to receive an honors diploma. These tests are the ones that most often end up in court (Mehrens, 1992; Mehrens, 1995). In order to successfully defend against a lawsuit, careful attention must be paid to the content of the test (it must match what has been taught), the timing of the notice (students need to know approximately three years ahead of time that passing the exam will be a requirement for graduation), and the technical quality of the exam (the test must be reliable, valid, and fair) (Phillips, 1993).

Table 2-2
States with a
High School Graduation Examination

Exit Examination Requirement

Alabama	Louisiana	New Jersey	Ohio
Florida	Maryland	New Mexico	South Carolina
Georgia	Mississippi	Nevada	Tennessee
Hawaii	North Carolina	New York	Texas
Virginia			

Endorsed Diploma Examination

Michigan	New York	Tennessee

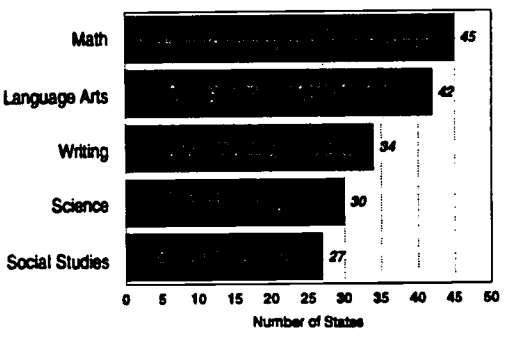
Honors Diploma Examination

New York	Ohio	Tennessee

Subject Areas Assessed

Five subjects are likely to be assessed by states no matter what assessment is used (see Chart 2-5). All the states with assessment programs assess mathematics; language arts (including reading) is assessed in every state but three. Writing is assessed in 34 states, down from 36 last year. There was also a drop in science (down from 34 states in 1991-92 to 30 states in 1994-95) and social studies (down from 29 states to 27). These decreases may be the result of a number of factors: (a) state department of education budgets are decreasing; (b) federal Title I assessment and evaluation legislation require states to assess mathematics and language arts and encourage the other subjects; and (c) some state programs, such as California and Arizona, have had significant cuts in the amount of assessment being conducted, which also has an impact on the number of states assessing each subject.

Chart 2-5
Major Subjects Assessment



Other subjects, such as music, foreign languages, health, vocational education, visual arts, and physical education, are assessed by fewer than five states apiece.

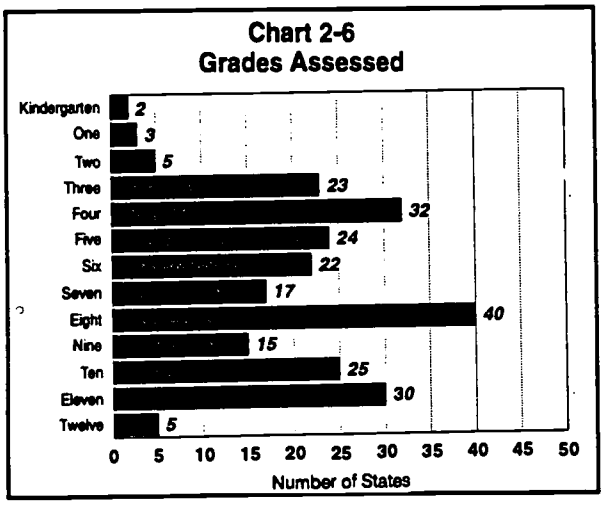
Subjects appear not to be assessed separately for purposes of accountability and improvement of instruction. Assessment in these five subjects most often follows the pattern of multiple purposes; in each subject area, almost all assessments are used for both accountability and instructional improvement.

Grade Levels Assessed

Which grades and how many grades are assessed varies widely among statewide assessment programs and components. Some patterns are worth mentioning, however. States are least likely to assess students in the early primary grades. States are most likely to assess students in grades 4, 8, and 11, as shown in Chart 2-6. All forms of assessment tend to be administered at these benchmark grades. Forty of the 45 states with assessment programs assess in the 8th grade, and 32 and 30 assess at the 4th and 11th grade levels, respectively.

In reviewing additional data from the Database, we were able to look at the relationship between assessment types and grade level. We found that generally:

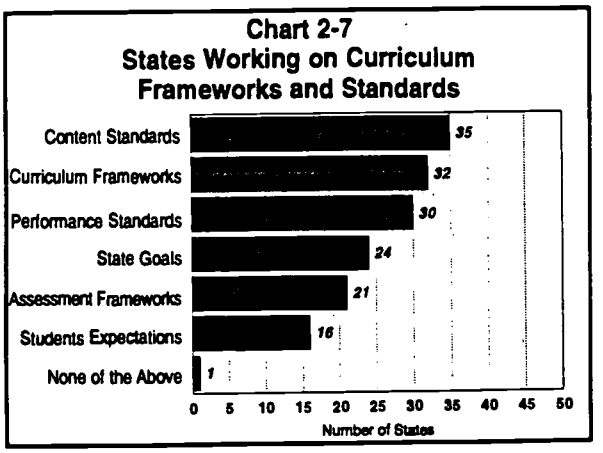
- Norm-referenced assessments clearly peak at benchmark grades 4, 8, and 11.
- Criterion-referenced assessments also peak at these benchmark grades, but are also frequently given at the grade levels between.
- Writing samples also occur most at the benchmark grades, but with a particularly strong peak at grade 8.
- Performance assessments show a similar grade-level pattern as NRTs.
- Portfolios are given in too few states to detect a pattern.



Curriculum Frameworks and Standards

Interest in curriculum frameworks and student standards continues to increase. The 1996 National Education Summit Policy Statement states, "We believe that efforts to set clear, common, state and/or community-based academic standards for students in a given school district or state are necessary to improve student performance" (National Governors Association, 1996, p. 2).

States are involved not only in the revision and/or development of assessments, but also in the redefinition of curriculum frameworks and student standards. This year, we devoted an entire section of the survey to this topic. When asked if they had state goals, student expectations, curriculum frameworks, performance standards, content standards, and/or assessment frameworks, most of the 35 states that responded reported that they had three to five of the above (see Chart 2-7). The most commonly checked terms were content standards (35), curriculum frameworks (32 states), and performance standards (30). One thing we discovered is that states use these terms differently and some states use the terms interchangeably. Even though we defined these terms in the Glossary and asked states to refer to it when completing this section of the survey, they relied instead upon the terms and definitions they use within their states, making comparability across states very difficult. Many other terms were also used by states in describing their standards. For example, Washington State refers to "essential academic learning requirements," Texas calls them "essential elements," Wisconsin refers to "goals and learner outcomes," and Oklahoma refers to "priority academic student skills." Still, Chart 2-7 makes it clear that there is a lot of activity surrounding standards development in the states. These frameworks and standards are not the same as those put in place during the earlier reform movement in the 1980s. Instead, almost all of the states that reported having completed curriculum frameworks, state goals, or student standards, had completed them by 1992. In many cases, state assessments are being developed or revised to match these "new" standards.



Another major issue states confront is whether or not their assessments match, or are aligned with, their standards. Since schools and students are being held accountable for demonstrating mastery of the standards through performance on the state assessment, it is important that the assessment match the standards. Our findings suggest that in reading, mathematics, and writing, more than half the states report alignment (25, 25, and 24 states, respectively), while in science and social studies, about half report alignment (19 and 20, respectively). In states that do not report alignment, most are in the progress of doing so or are planning to do so.

States appear to be working independently of one another in developing their standards, relying primarily upon educators and curriculum organizations within their state. Some report having their business community and "public" review the standards. A few states mention that they are working with contractors (e.g., Riverside in Washington State), universities (e.g., Florida State University; Wisconsin Center on Educational Research), or regional laboratories (e.g., Mid-continent Regional Educational Laboratory in Florida and Wyoming). One might surmise that a lot of reinventing the wheel is going on, although a review of state standards would need to be conducted to assess their comparability. In the area of mathematics, where the NCTM standards have been out since 1989, there may be more comparability. In language arts and social studies, where national standards have not been as readily accepted, the similarities among state standards are less likely. The New Standards Project and the CCSSO State Collaboratives on Assessment and Student Standards are also helping states to work together in the development of standards and assessments. Comparability among standards and assessments will need to be addressed as states use their assessments to evaluate the effectiveness of Title I programs.

Summary

Over the last four years, certain findings of the survey have been consistent. State assessment remains a significant tool for educational reform in 45 states. In general, students are assessed most often at grades 4, 8, and 11 for the purposes of improvement of instruction, school accountability or school performance reporting, and program evaluation. Approximately one-third of the states with assessment programs require students to pass an exam to graduate from high school. Students are most often assessed with a combination of multiple-choice testing and writing samples, with a combination of multiple-choice testing, writing, and alternative assessment running a very close second. Only a few states rely on multiple-choice testing or alternative assessments exclusively. The use of alternative assessments or constructed, open-response testing in conjunction with multiple-choice testing continues to grow, but the exclusive use of one or the other form of assessment is lessening.

The tensions that exist when assessment is used for both school or student accountability and instructional improvement continue to cause difficulty for those who design and implement these programs. Unfortunately, most state legislatures require these conflicting purposes in their assessment programs. The tensions are often further complicated by placing negative consequences on poor performance, thus increasing the stakes for schools and students.

Most states have recently revised their standards or are in the process of doing so. Assessment development and revision are also taking place to ensure alignment between the assessments and the standards. A lot of work remains to be completed in this area, however, and it looks like most states are working independently in this endeavor.

One word that describes state assessment activity over the past four years is change, and that change seems to be occurring at an even greater pace. Examination and revision of standards is driving a lot of that change.

Newer Forms of Statewide Assessment

Traditional multiple-choice assessments continue to be the most popular form of assessment in state assessment programs. In fact, 9 states rely exclusively on norm-referenced and/or criterion-referenced multiple-choice assessments, and 43 of the 45 states with a statewide assessment program administer at least one multiple-choice test.

Prompted by a growing concern that the kinds of skills needed for success in the 21st century go beyond those that are typically taught and assessed in traditional educational settings, states have been revising their student learning goals, their curricula, and the forms of assessment they use to measure mastery of those student goals. As a major part of this educational reform effort, states have explored alternative[3] forms of assessment that require students to produce answers rather than simply select correct answers. Most states have added these alternatives to their existing forms of assessment. Moreover, a small, but highly publicized group of states embraced alternative forms of assessment as their primary or exclusive means of measuring student success. Over the last two years, a few of these states have hit some major roadblocks.

The Pendulum Swings Again

Three of the states that were farthest along in their use of alternative assessments as a primary assessment strategy have hit major detours due to technical problems, cost, and public criticism of content. They are California, Kentucky, and Arizona. In California, the state's major assessment program, the California Learning and Assessment System (CLAS), which relied heavily upon performance assessments and constructed-response items, has been discontinued. In its place will be a statewide basic and applied academic skills assessment at key grade levels, and a voluntary Pupil Incentive Testing Program. The highlights of the Pupil Incentive Testing Program include:

1. Districts will receive $5.00 per student to select a published achievement test.
2. Students need to be assessed in reading, spelling, written expression, and mathematics by a standardized test from a state-approved test list.
3. Districts must administer the tests to all eligible students from grades 2 through 10.
4. Districts must report the results annually to their students, teachers, parents, and governing boards.

Another state that was moving away from multiple-choice items and toward the exclusive use of performance assessments and portfolios has faced similar problems. In Kentucky, multiple-choice items will be returned to the assessment program and a traditional, standardized test will be added. In relying on performance assessments and portfolios exclusively, Kentucky found that they needed more information per student, and it needed to be collected in a cost-effective and technically sound manner. Arizona is yet another state that had a major nontraditional assessment program suspended in the school year 1994-1995. For now, it is only administering a norm-referenced, multiple-choice test.

Two other states that were moving toward a heavier reliance on performance assessment have had the funding for their programs withdrawn. In Wisconsin, the performance-assessment component of the state assessment system lost its funding after a three-year developmental period was nearly completed. Full implementation in language arts and mathematics had been planned for next year. Indiana similarly lost its funding

[3]Throughout this chapter, alternative assessment and nontraditional assessment refer to non-multiple-choice assessment.

after developing and piloting a new assessment program that included a move away from norm-referenced testing to criterion-referenced testing, and the inclusion of a substantial number of open-ended items and performance tasks. The new legislation calls for the continuation of the norm-referenced test with its criterion-referenced supplement, and one open-ended mathematics task and one writing sample at benchmark grade levels. Even this minimal inclusion of alternative assessment was challenged in a lawsuit claiming the test invaded the privacy of children. Indiana won the suit.

The first three states discussed (California, Kentucky, and Arizona) were among the leaders in the alternative assessment movement. Kentucky is the only one of the three that remains in the forefront; California's program is defunct and Arizona's is on hold. The other two states discussed, Wisconsin and Indiana, may have been caught in the flak that resulted from the very public attacks against the first three states' programs. Political battles, concern over so-called "nonobjective" and "intrusive" forms of assessment, high costs, and technical difficulties seem to be at the heart of much of the retreat from alternative assessment activity. Some of these concerns will be discussed more fully later in this chapter.

A Blend of the Most Common Assessments

Nontraditional assessment items previously have been defined in the SSAP Survey as writing samples, performance events, and portfolios. This year we have included the category of constructed, open-response assessment since a number of states use this terminology to describe open-ended assessment strategies that are not as "involved" as performance events. Thirty-seven states report the use of nontraditional assessment items, with 17 of these having a writing sample as their only alternative form. Nine other states report being in the very earliest stages of development or having plans to develop alternatives. Seventeen states are using performance events; ten are using constructed, open-response items; and five are using both to enhance traditional, multiple-choice assessments. Five states report the use of portfolios, but two of these programs are voluntary, one is locally determined, and one is not "scored" (see Table I in Appendix). While incorporating alternative assessments into state assessment programs will probably continue, their exclusive use is not likely.

The most common pattern of assessment types this year is some combination of multiple-choice testing and a writing sample (17) or multiple-choice testing, a writing sample, and an alternative form of assessment (16 states). Table 3-1 includes a summary of this information. Figure 2-1 (see page 4) shows which states administer which combinations of assessment types. Table 1 in the Appendix indicates that 19 states report the use of some performance measure and/or portfolio, with 14 reporting the use of performance events, 1 reporting the use of portfolios, and 4 reporting the use of both. Only 13 of the 19 states using either performance events or portfolios require that they be used with all the students. The others have a voluntary program or use a statewide sample of students.

Table 3-1
Combination of Assessment Types Used by the States

Combination	*Number of States*
Multiple-Choice (NRT or CRT) only	8
Multiple-Choice and Writing Sample	17
Writing and Constructed, Open-Response	1
Multiple-Choice and Alternative (Performance events and/or Portfolios)	2
Multiple-Choice, Writing, and Alternative	16
Alternative and Constructed, Open-Response	1
No Statewide Student Assessment Program	5

The movement toward the exclusive or primary use of alternative forms of assessment in state programs is slowing down. However, states continue to explore alternatives to multiple-choice assessment as a supplement to their traditional assessments. Figures 3-1 and 3-2 demonstrate the amount of growth in nontraditional assessment since we began collecting state student assessment data in 1991-1992 and then again in 1994-1995. Interestingly, the majority of the growth has been on the East Coast with a noticeable lack of activity in the Midwest and Northwest (Kentucky and Minnesota are the exceptions to this).

Figure 3-1
Performance and Open-Response Item
1991-1992

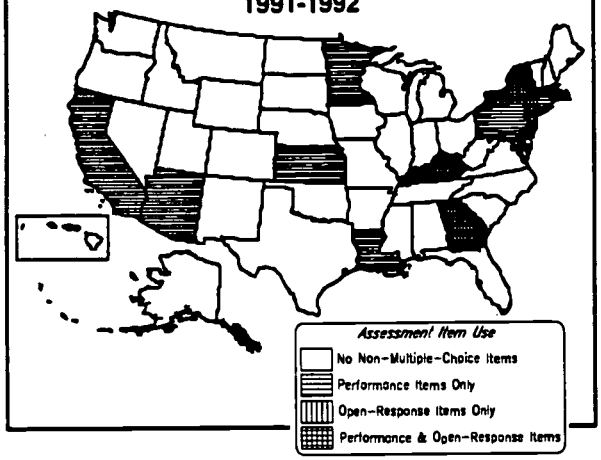
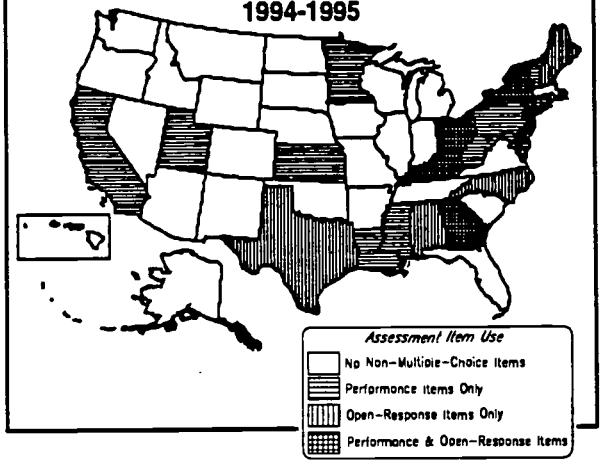


Figure 3-2
Performance and Open-Response Item
1994-1995



In spite of all of the pressures away from the exclusive use of alternative forms of assessment, some states are still moving full-steam-ahead in their implementation of assessments based in whole, or in large measure, on alternative assessments. Maine continues to use constructed, open-response items and writing assessments, and has just this year completed the move to an "all alternative assessment" system. Two other states continue to rely heavily on nontraditional assessments. Maryland retains a traditional seventh-grade functional literacy test and a norm-referenced test as part of its assessment program, but its major assessment component continues to rely upon performance assessments and writing samples. Maryland also has plans to move away from its multiple-choice functional literacy test toward a more performance-based model. It administers its norm-referenced test to only a sample of its students. Vermont primarily uses mathematics and writing portfolios but also administers uniform tests in mathematics (a short, criterion-referenced test) and a uniform assessment in writing (a writing sample). However, its program continues to be challenged by technical quality issues. Whether these states will continue to move forward or whether they too will be forced to slow down will be something we will watch over the next couple of years.

There are also some new players in the alternative assessment movement as well. Kansas reports a change in focus from content and knowledge toward process and product, which calls for the inclusion of a performance-based format in all subject areas. While multiple-choice items continue to be a necessity, these questions are now focused on cognitive processes and greater care is given to measuring problem solving and critical thinking. Pennsylvania added performance tasks to the state's reading and mathematics assessments to help encourage performance assessments at the local level. Georgia also reports the use of performance events and constructed, open-response items as part of its assessment program. North Carolina is working with Grant Wiggins at the Center on Learning, Assessment and School Structure (CLASS) to create a different kind of assessment system that will use alternative assessments and teacher involvement in new ways. However, the state is

also moving forward on a more traditional accountability measure that will be used in conjunction with the alternative form. Again, blended approaches seem to be the norm. Whether or not the "blends" give a better overall picture of student learning in a state, or a disjointed picture based on the lack of alignment between different assessments, is an empirical question that needs to be addressed.

Why a Blended Assessment Approach?

The fact that states are moving toward the use of multiple types of assessment makes sense. After all, no single form of assessment is appropriate for all purposes. There are trade-offs involved in the use of any assessment strategy.

Alternative forms of assessment are being explored for many reasons. First, there is a national movement to clearly define student standards, that is, what students should know and be able to do. Along with this standards movement comes a desire to accurately describe what students now know and can do vis-a-vis the standards. Alternative forms of assessment are being designed to make these determinations, particularly with standards that cannot be assessed with a paper-and-pencil test. In addition, many of the standards are different from what has traditionally been taught in schools. Changes in the workplace and in the skills needed for life in an information age suggest that students need knowledge and skills that will enable them to solve increasingly complex problems. Some of these skills cannot be assessed using traditional, multiple-choice assessment, and this is causing many states to explore alternatives.

Multiple-choice assessments require students to select a "right" answer from among several "wrong" answers. These assessments are useful for assessing knowledge and the straight-forward application of that knowledge. On the other hand, open-ended assessments that require students to generate their own solutions to assessment problems or tasks are becoming increasingly necessary to assess new learner outcomes that call for more complex applications of knowledge and skill. Many states are concerned that relying exclusively on traditional multiple-choice, basic skills assessments results in a narrowed curriculum that produces students who memorize a lot of facts and skills, but have little ability to apply them to real-life situations. However, these assessments are easy to administer, fairly inexpensive, and yield a broad sample of student performance in a relatively short period of time. They simply can't be used to assess more complex applications of student knowledge, and they offer few clues to the teacher about why the student gave a correct or incorrect answer.

This is why states are adding alternative forms of assessment. One of the major benefits of nontraditional assessment is that, in addition to judging the correctness of the student's answer, the appropriateness of the procedure that the student employed is also considered. This gives teachers more information for diagnostic purposes because the teacher can determine where the student is having difficulty. But nontraditional assessments also have their trade-offs—most notably, the increased cost and time associated with their development, administration, and scoring. Ensuring the reliability of these assessment results has also proven costly and difficult, although the benefits in improved assessment of complex skills and the modeling of good instruction is worthwhile to some states. Another difficulty of nontraditional assessments is generalizability. Different performance tasks evoke different levels of skill from the same students. This limits the likelihood that a given performance on a small sample of tasks will be strongly indicative of the student's overall ability.

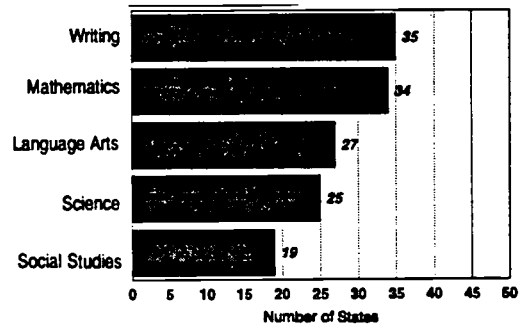
For these reasons, most states are combining traditional assessment programs with nontraditional assessments (see Figure 2-1 in Chapter 2 and Table 3-1 in Chapter 3). They are also examining their traditional programs, which are getting a face-lift with new content and standards.

Nontraditional Exercise Development in the 1994-1995 School Year

The number of states with nontraditional exercises in all subjects is depicted in Chart 3-1. As was the case last year, mathematics and writing are the most common subjects assessed with nontraditional exercises.

Comparing this with last year's findings in the 1995 Annual Report, we see that nontraditional assessment activity is down from last year in all subjects except for science. Some, but not all of the decline can be explained by the elimination of California's program and the suspension of Arizona's program. Most of the ongoing developmental work is apparent in writing, mathematics, other language arts (including reading), science, and social studies. These are the subjects most commonly assessed with traditional forms of assessment as well. As reported in Chapter 2, most of this activity is being conducted as part of a blended assessment program, one that includes both traditional and nontraditional assessment.

Chart 3-1
Major Subjects Assessed by States With Nontraditional Items



Types of Nontraditional or Alternative Items

Chart 3-2a shows the most commonly used types of nontraditional exercises in language arts and writing. Extended-response, open-ended items are by far the favorite means of assessing writing. Language arts is assessed most often with short-answer, open-ended items; extended response open-ended items; and interviews.

Chart 3-2a
Nontraditional Exercise Use: Languages Arts and Writing

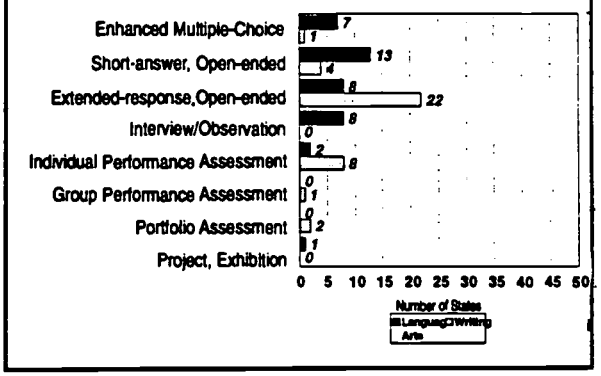
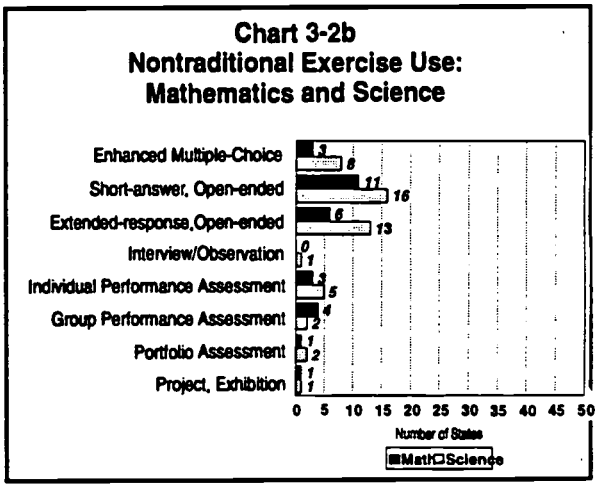


Chart 3-2b shows the most common exercise types for mathematics and science. Short-answer, open-ended exercises are used most commonly with mathematics, with extended-response, individual performance assessment, and enhanced multiple-choice exercises following. Science shows a similar pattern.

Chart 3-2b
Nontraditional Exercise Use: Mathematics and Science



There is a noticeable decline from last year's data in the number of states (approximately three to four) using nontraditional exercises in every subject area and every type of nontraditional exercise except for interview and observation in language arts. This same decline is evident when we compare Chart 3-3a and 3-3b with last year's data. Again, approximately four to six fewer states are developing, and one to four fewer states have completed development of nontraditional items in mathematics and writing, the two most common subjects for nontraditional

assessment. A similar pattern can be observed in Chart 3-3b for science and language arts, although the biggest drop in nontraditional items is in language arts: 4 states compared to 18 states last year. Although one year's data cannot be used to detect a trend, the fact that this activity has been increasing every year until now is significant.

Chart 3-3a
Development of Nontraditional Items: Mathematics and Writing

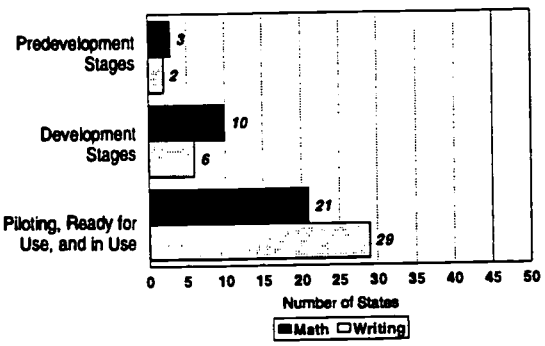
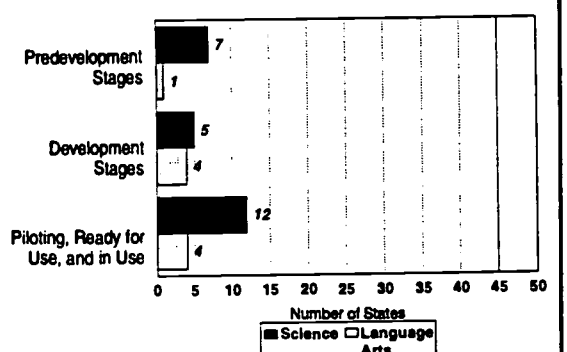


Chart 3-3b
Development of Nontraditional Items: Science and Language Arts



Constraints on Developing Nontraditional Assessments

While the changes in assessment programs, and the criticism in use of nontraditional assessment programs in particular, have been in the news, the survey responses to questions about the kinds of constraints states faced as they implemented alternative assessments do not reflect the difficulty states are facing. In response to question 3.13, "If this component included nontraditional items or assessments, did your state encounter major difficulties in developing them?" only 6 of the 21 states responding said yes. The six states included Kentucky, Maine, and Vermont, which have major investments in nontraditional assessment. However, states such as California, Arizona, Indiana, and Wisconsin, all of which lost their nontraditional assessment funding, did not respond. This may be because we asked them about "existing" assessment programs, and by the time the survey was completed, their programs were no longer in existence.

Purposes and Consequences Make a Difference

Three states reported that time was a major constraint, two indicated cost was the limiting factor, one reported insufficient evidence of technical quality, and three reported resistance to change to nontraditional measures. Their responses pointed to the following issues, among others:

Time. There are two time constraints. The first is the time to develop a test. This constraint is compounded by a sense of urgency: Several states reported legislative mandates to put their programs into place before the tests were ready. The second constraint is the time to administer an alternative assessment in the classroom. In the time it would take a student to complete one or two performance tasks, that same student could have completed 200 items on a multiple-choice test.

Cost. Again, there are several issues. Since the technologies are new, the procedures to develop items or tasks are not nearly as well established as they are for multiple-choice assessments. It takes more people more time to develop and test such items. The time required for classroom testing also adds to the cost of alternative assessment. Alternative assessment items are more expensive to score than multiple-

choice tests. Alternative assessments require teachers or other professionals to record observational data or make judgments about extended artifacts of student performance. This requires the skill and time of individuals if the work of many students is to be assessed.

Professional development is also a considerable expense for alternative assessment: Staff need to understand the changes, need training in the consistent conduct and use of alternative assessment items, and need support in using and reporting the results of alternative assessment. However, the professional development benefits derived from teachers who design, implement, and/or score the nontraditional assessments is a benefit many states cite as a major reason to continue this work.

Technical Quality. Because nontraditional items are a new technology, it is far from easy to obtain uniform results. While some technical concerns are not unique to nontraditional items and may in fact pose less of a threat—for example, the issue of validity (are we assessing important learning?)—they remain real. Others, such as reliability (student results are an accurate reflection of the student's performance rather than a result of extraneous influences such as who does the scoring) or generalizability (scores on this assessment would be similar to scores on similar assessments), continue to be daunting. There is so much more flexibility with nontraditional assessment that maintaining uniformity of administration, scoring, and interpretation is more difficult.

Resistance to Change to Nontraditional Methods. This resistance comes mostly from students, teachers, and parents. All three are more familiar with standardized tests where minimal preparation and administration time are required, and reports are straight-forward and support a norm-referenced grading system (A, B, C, D, F). Organized groups of parents have also fought the new assessments in a number of states due to concerns that the open-ended nature of performance assessments will allow students to be judged on the basis of the personal values they include in their responses rather than their academic performance.

In reviewing the data on nontraditional assessment activities this year, it would appear that where states have implemented performance assessment as a slow and deliberate process without much fanfare, their programs have been spared. Connecticut was one of the first states to proceed with performance assessment, but did so through a series of research grants and only implemented the assessments once they had been thoroughly researched. What the results are used for also seems to make a difference. Most of the states that report a lack of major difficulties in implementing nontraditional assessments tend to use their assessments as end-of-course exams (e.g., Alabama's Math End-of-Course Test and California's Golden State Exams), for early-childhood screening (Georgia's Kindergarten Assessment Program), for career/employability skills assessment (California's Career-Technical Assessment Program), as instructional planning tools (Connecticut's Academic Performance Test), or when the alternative assessment is a writing sample (Idaho's Writing Assessment, Rhode Island's Writing Assessment, and Vermont's Uniform Test in Writing). All of these are fairly low-stakes purposes, meaning that consequences of poor performance are not severe for students, schools, and/or teachers. State assessments seem to come under attack most often when the use of the test results is high-stakes—student graduation, school accreditation, school takeover, and so on. Of course, these assessments also receive the most press attention and public appraisal. Most programs have flaws, but when severe consequences are dependent upon the results, any flaw becomes more pronounced.

Summary

In summary, it has been a "challenging" year for states that are moving to incorporate nontraditional forms of assessment into their assessment systems. A number of highly publicized programs, such as those in California, Kentucky, and Arizona, have come under attack, with California losing its program, Kentucky losing some of its funding and receiving a mandate to add more traditional forms of assessment to the program, and Arizona having its program suspended for further investigation. Other states, such as Indiana and Wisconsin, have lost funding after a number of years of developmental work, and still others are finding themselves moving more slowly and cautiously in their development and use of alternative assessment. Concerns about cost, technical quality, possibility of values-laden content, and time have been the major points of contention. Interestingly, there is still considerable activity in states to design and implement alternative forms of assessment, but it is possible that the recent criticisms may slow down these efforts as well.

It would appear that states that moved "full-speed ahead" and were the greatest alternative assessment advocates are the ones that have incurred the most attack. As is often the case with any innovation, the risk-takers are oftentimes prodded to do more and more at a faster and faster pace, running the very real risk of making mistakes or getting caught in the bright light of publicity before they are ready. Most of the programs that have failed or are in trouble admit that they have not done a sufficient job of bringing their publics along with them. They have been so busy designing, pilot testing, and refining, that they simply haven't spent enough time explaining the need for the change and the safeguards that have been taken against potential problems.

Perhaps this roadblock will not become a dead-end for nontraditional assessment and will instead give those who are leaders in the area a chance to study the benefits of nontraditional assessment, improve upon its shortcomings, and allow states to implement it at a slower and more reasonable pace and for the purposes for which it is most useful (i.e., student diagnosis and instructional planning). It would be a shame if we once again throw out the baby with the bath water. Major benefits can be derived from understanding why students respond as they do and how they use their thinking processes to work through a problem, understandings that can only be derived from alternative forms of assessment. Perhaps the very common approach of adding nontraditional assessment to traditional state programs will continue to be the trend for the next few years.

Chapter Four

Special Topics: Part I: Assessment of Students With Disabilities and Limited-English-Proficient Students

When the 103rd Congress overhauled the Elementary and Secondary Education Act, its new Title I legislation, the Improving America's Schools Act, called upon states to hold all students to the same high expectations and to ensure they have equal educational opportunities (Phillips, 1995). The definition of those high expectations and the design of the assessment system used to determine whether or not students have achieved those high expectations are left to individual states and local school districts. This has spurred a growing debate over which students should be tested and how that testing should be conducted. A major concern surrounds the inclusion of students with disabilities and Limited-English-Proficient (LEP) students in statewide assessment programs.

Two questions are of paramount importance in understanding the current practice of assessing these students. How many students with disabilities and LEP students currently participate in statewide assessment programs, and what kinds of special testing conditions or accommodations are allowed to enable them to participate? These questions were included in the fall 1995 edition of the Association of State Assessment Programs (ASAP) survey. Additional information about students with disabilities is provided by the National Center on Educational Outcomes (NCEO), a group committed to assisting states in implementing activities to improve outcomes for these students and to document states' efforts in doing so (Ysseldyke, 1996). The author also relied heavily upon an article written for NCREL by Susan Phillips, an attorney and measurement professor at Michigan State University, entitled "All Students, Same Test, Same Standards: What the New Title I Legislation Will Mean for the Educational Assessment of Special Education Students" (1995).

Participation of Students with Disabilities and Limited-English-Proficient Students in Statewide Assessment

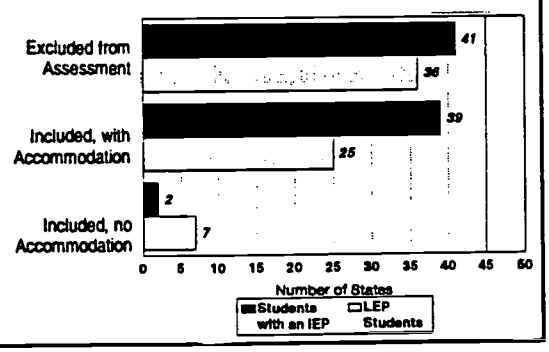
Forty-one states have written guidelines about the participation of students with disabilities in their statewide assessment programs. Of the 133 different assessments employed by states, participation rates can be estimated by state special education directors for only 49 (Ysseldyke, 1996). When participation rates for students with disabilities are offered, they range from 6 to 14 percent of the total tested elementary population and 5 to 10 percent of the total tested high school population. The accuracy of these participation rates are questioned by both state testing directors and special education directors because the data are not collected systematically in many places. In fact, similar participation rates are not available for LEP students. Better and more precise information will need to be collected to have an accurate estimate of the participation rates of these students.

Chart 4-1 shows that 41 states allow students with disabilities to be excluded from the state assessment program, while 36 states allow for the exclusion of LEP students. In many states, schools are allowed to exclude these students if the assessment is not appropriate for them (for example, the content is not included in the student's Individualized Education Plan or the student does not know enough English to successfully

complete the exam). Few states collect data regarding the numbers of students with disabilities or LEP students who are excluded. Most states can estimate the percentage of the tested population who are students with disabilities or LEP students, but few can determine what percentage of the total population of students with disabilities or LEP students are excluded from assessment. We are working with the National Center on Educational Outcomes to improve the collection of this information next year.

When students with disabilities or LEP students are included in statewide assessment, the extent to which testing accommodations[4] are allowed for these student varies from state to state. Only 2 states include students with disabilities in the state assessment program without accommodations, and 39 include them but allow accommodations. Seven states include LEP students without accommodations, and 25 with accommodations. For most of these decisions, if the assessment is deemed inappropriate, that is, the student is not expected to master the content of the assessment as part of his or her instructional plan, a decision may be made to exclude him or her from the assessment. If the assessment is seen as appropriate as is, the student is included without accommodation. If the assessment is seen as appropriate, but only with special accommodations, the student is allowed those accommodations. The decision is never as clear-cut as this sounds. A great deal of local flexibility is allowed in most states, and local districts interpret the broad state policies in varied ways.

Chart 4-1
IEP and LEP Students:
Inclusion/Exclusion and Accommodation



Determination of Which Students With Disabilities and Which LEP Students Participate

The survey asked state testing directors to describe the policies their states use when determining whether or not students with disabilities and LEP students should participate in the state assessment program. For most states, a special education student is included or excluded from the state assessment based on the recommendations included in the student's Individualized Education Plan (IEP). For LEP students, the level of English proficiency and/or the number of years the student has been in English-as-a-Second-Language classes are the determining factors.

In a few states, the determining factor for inclusion of students with disabilities is whether or not the student is reading at grade level. A number of states, including California, Idaho, Michigan, and Utah, use the 50 percent rule (if the student spends 50 percent or more of his or her time in regular education classes in the tested subject, the student is included in the state assessment). Even in these states, however, the IEP may override the 50 percent rule.

Even when special education students participate in the statewide assessment program, their scores may not be included in the state, district, and school averages. Many states offer schools this option, partly because they are interested in having as many special education students tested as possible, and partly because the special circumstances under which some special education students take the test make the results less comparable to those of other students. Although an exact number is unavailable, many states report that the assessment results of students with disabilities or LEP students may be eliminated from state, district, and school assessment summaries.

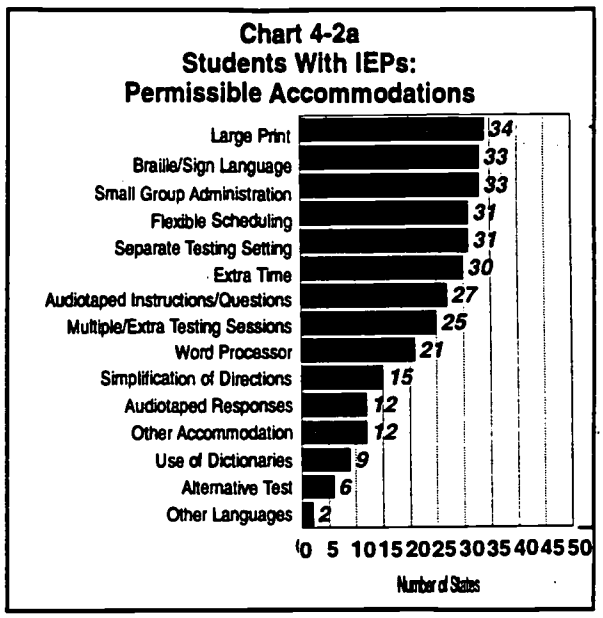
[4]Testing accommodations refer to special conditions or supports that minimize the impact of the student's disability on his or her performance. Examples of testing accommodations include Braille and large-print versions of the test for vision-impaired students, scribes for students who are physically incapable of writing, smaller or separate testing settings for students whose disabilities cause them to be easily distracted, and so on.

Special Testing Accommodations for Special Education Students

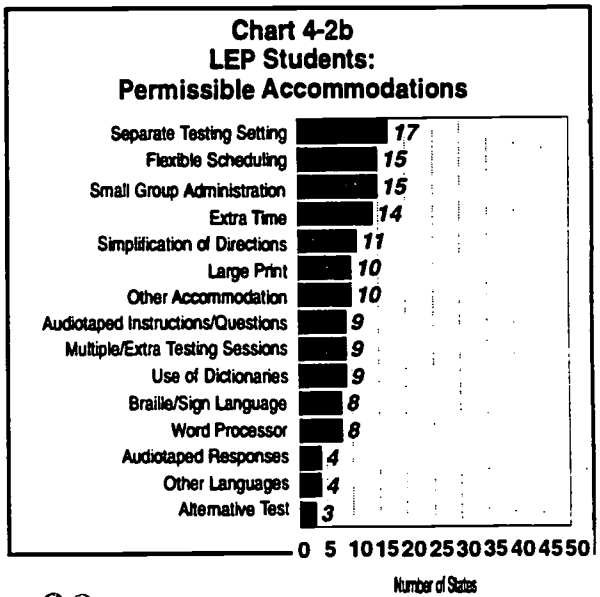
In order to give students with disabilities an "even" chance to pass the state test, many states allow special testing conditions or "accommodations." Most states have little problem allowing testing accommodations that allow a physically handicapped child the tools he or she needs to "take" the test. No one disagrees that a blind student should be allowed a Braille version of the test or that a student with muscular dystrophy should be allowed a scribe—someone who will write down the student's answers.

The problem arises when the disability is cognitive in nature. Some accommodations for cognitive problems provide students with "extra help" in the subject being tested. The score of a student with dyslexia who is read the reading test is not an accurate or valid measure of that student's reading ability. On the other hand, if that same student is read the questions on a mathematics or social studies test, the accommodation is not as closely related to the skill being assessed. The student's mathematics score or social studies score therefore would be a more reasonable estimate of his or her mathematics or social studies knowledge than the reading score would be of his or her reading ability (Phillips, 1995).

Chart 4-2a reports the testing accommodations states allow for students with disabilities. Of the 37 states reporting the use of special testing accommodations for special education students, nearly all reported allowing the use of Braille and large-print versions of the test, small group administrations, and flexible scheduling. Most allowed extra time and separate test administrations. Some states, such as Maryland and Hawaii, provide numerous accommodations, including reading and/or transcribing the test, extended time periods, small group administration, audiotaped versions, signed versions for the hearing impaired, use of calculators and/or word processors, large print, and Braille. A number of states mentioned that decisions concerning special accommodations depended on their impact on the validity or interpretability of the results (for example, reading a reading test to a student would not be allowed).



A much smaller number of accommodations are allowed for Limited-English-Proficient students. Chart 4-2b shows the responses of 17 states to the question, "What kinds of testing accommodations do you allow for LEP students?" Nearly all of these states reported allowing the use of separate scheduling and testing settings, small group administrations, and extra time. Approximately half of the states who responded allow audiotaped instructions, multiple/extra testing sessions, simplification of directions, and use of dictionaries. Only four states reported that they allowed other languages to be used with LEP students, and only three states administered an alternative form of the exam.



What Next?

While the field of measurement has contributed a set of rules concerning reliability and validity of results that help govern the inclusion and/or accommodation of students with disabilities and LEP students, little actual research exists that demonstrates the impact of accommodations on test validity. Several studies are under way, a number of them sponsored by the National Center for Educational Statistics (Phillips, 1996), to address this question empirically. In addition, special education and assessment representatives from 30 states met on January 10, 1996, at a CCSSO Special Education State Collaborative on Assessment and Student Standards (SCASS) to discuss these and other related questions concerning the assessment of special education students. Other studies are needed to assess the impact of inclusion or exclusion of special education and LEP students on the educational opportunities they receive as a result of that decision. The studies now under way should provide additional guidance to those who are concerned for the right of these students to be assessed and to be provided the opportunity to reach the same high standards as their nondisabled or English-speaking peers.

Part II: State Title I Assessment and Evaluation Plans

A separate section of the fall 1995 Association of State Assessment Programs Survey was dedicated to states' assessment and evaluation plans for Title I. While only a handful of states returned complete descriptions of their Title I assessment and evaluation plans, a number of interesting findings were noted:

1. States that had existing standards and assessment-based reform programs in place were in good shape for responding to the new Title I assessment and evaluation requirements. For example, Kansas, Kentucky, Maryland, New York, Ohio, Tennessee, and Texas reported very complete plans with minor concerns/changes expected in the final assessment plan.
2. Those states that were "between reforms" or in the very early stages of implementation of reform reported having a difficult time responding to the Title I requirements. Although the final set of standards and assessments do not need to be in place until the year 2000-001, a final consolidated (or Title I) plan is due to the USDOE this May. A number of these states made comments such as "political changes make direction uncertain."
3. Eleven states simply didn't respond to this section of the survey, and one gave a minimal response—"We're working on it."
4. A few states mentioned that the person who filled out this section of the survey was not the same as the assessment director. In a number of states, Title I directors are trying to design the Title I assessment plan without input from state assessment directors. This is not always intentional—with state education agency downsizing so common, state assessment directors and Title I directors are so busy putting out fires, they are not available to work together on Title I assessment issues. The result may be two separate standards and assessment systems within the state, leading to confusion and mixed messages to schools about the definition of "quality."
5. States that are well along in developing Title I Evaluation and Assessment Plans tended to report specific problems/concerns such as how to define "adequate yearly progress", how to set performance standards, and how to determine appropriate inclusion criteria and accommodations criteria for special education and LEP students.

6. At least five states reported that they planned to use the state's norm-referenced test as the primary assessment tool for Title I. Only Colorado mentioned that the use of an NRT would not be allowed.

7. States that do not have statewide assessment programs, and do not plan to implement them (e.g., Iowa and Wyoming), reported that the evaluation of Title I programs would be largely up to local school districts. Iowa reported that they will allow use of the voluntary Iowa Test of Basic Skills program and will also provide districts with a number of assessment models and models of best practice for districts to use on a voluntary basis.

It would appear that states are either not as far along with their Title I Assessment and Evaluation Plans as predicted, or that they had difficulty responding to this part of the survey. From the authors' experiences with other national Title I Assessment and Evaluation planning efforts, we believe states are genuinely struggling with the "flexibility" provided by the new Improving America's Schools Act legislation. It might be helpful to provide states with a number of models that could be collected from those states that are farthest along. Efforts such as the Council of Chief State School Officers' Title I State Collaborative on Assessment and Student Standards, which give states the opportunity to work together on some of the crucial assessment issues, also hold promise.

Statewide Assessment History and Trends

Introduction

This is the fourth year in which the information about statewide large-scale assessment programs has been collected systematically and made available by CCSSO and NCREL. With data being collected for four years, it is possible to see trends in the information. These trends are further supported by information collected informally from state testing directors throughout the history of the Association of State Assessment Programs. While we feel fairly comfortable reporting these trends, readers are asked to interpret them cautiously since changes in student assessment programs take several years to conceptualize and implement.

The purpose of the following sections is to comment on some of the changes that have occurred in the past 15 years. In addition, several issues that may imply future changes in assessment are mentioned.

Criterion-Referenced Assessment and Minimum Competency Tests

When the Association of State Assessment Programs was formed as an organization representing the assessment programs at the state and national level in 1977, two strong innovations had occurred and were being spread throughout the states. First, states such as Michigan had adopted a new form of measurement called "criterion-referenced tests" in the early 1970s. Rather than comparing student (or school or district) scores to national norms, scores were reported as pass-fail for individual objectives and for the proportion of the objectives passed. Second, other states were using tests to determine whether students had learned enough to receive a high school diploma. This use of minimum competency testing for high school graduation was exemplified by the landmark program in Florida.

The Association was formed for states to help one another in developing quality assessment programs with a minimum of wasted effort or controversy. Early ASAP meetings were filled with discussions about the procedures for developing criterion-referenced tests, as well as surviving the inevitable legal challenges to the minimum competency tests, since the landmark legal case Debra P. v. Turlington was occurring at that time.

The predominant form of large-scale assessment at that time was norm-referenced tests. Interest in criterion-referenced tests was pushed along not only by the states that had adopted them as a form of assessment, but also by the National Assessment of Educational Progress (NAEP) in its early years. At that time, several states (such as California, Connecticut, Minnesota, and Wyoming) gave the early NAEP assessments in "piggyback" style in order to obtain state and national data on their students. Not only did this practice introduce these states to criterion-referenced testing, it also served as an introduction to the concept of the state NAEP assessment program.

Advent of Writing Assessment

In the 1970s, assessment was limited usually to mathematics and reading, with performance assessments just beginning in the area of writing. The NAEP assessments of writing in the early 1970s had encouraged the belief that having all students at one or more grade levels actually write essays would be feasible. Although more expensive than the much more prevalent multiple-choice tests of "writing," essay tests were thought to be more content valid, and it was believed that they would lead to better teaching of writing. However, strong debates about this concept occurred during this time.

Expansion to Other Subject Areas

In the 1980s, additional states adopted large-scale assessment programs as a tool for school reform and improvement. Each year at the ASAP meetings, one or two states new to large-scale assessment efforts would attend. In addition, states were beginning to add other subject areas to their assessments. They began to develop assessments in areas such as science, social studies (or one or more of its components, such as history or geography), health education, physical education, the arts, and vocational education. Interest also grew in the sharing of assessment items or tasks among the states, since so many new states were now interested in large-scale assessment. Attempts were made to create item banks among the states, but these generally proved to be unsuccessful since each state clung to its own set of student expectations, making sharing of corresponding items challenging at best.

Performance Assessment

For most of the history of state assessment, multiple-choice tests were (and still are) the major form of assessment used in most states, with the exception of states that used a writing sample. However, strong criticism of multiple-choice tests in the late 1980s led to the exploration of performance assessment by states. From this early exploration until now, it appeared that more states were implementing performance assessment each year. During the last few years, however, a couple of trends have started to emerge. First, a small group of states (Maryland, Arizona, and California—joined later by Maine) were the first to entirely or mostly rely upon performance assessment to collect student data. Other states are considering developing such programs, including Massachusetts and Delaware. These states have demonstrated that it is feasible to administer alternative forms of assessment in a relatively cost-effective manner, but parents, legislators and teachers haven't necessarily agreed with the alternatives. For example, concerns about test content and technical quality caused the innovative assessment programs in California and Arizona to be shelved last year.

Second, a number of states are working on or piloting alternative forms of assessment. This innovative work includes performance assessments that are given to individuals or small groups of students; curriculum-embedded tasks in which assessment is intricately interwoven within teaching and assessment information is collected over several weeks or months; the use of portfolios to collect examples of student work for later scoring; and other innovative forms of assessment. As the SSAP survey indicates, few states have actually implemented these innovative alternative forms of assessment, but given the number of states reporting such work, it is logical to assume that these numbers might increase. It is likely that, given the costs of alternative assessment in money and time, most states will move toward the concept of an assessment system, with different forms of assessment being used at different levels. For example, large-scale, standardized assessments with some alternative approaches might be used for state-level reporting, while more extensive programs of performance and/or portfolio assessment might be used to meet school or classroom assessment needs. Hence, several states report that such innovative performance assessments are being developed for use by local educators.

A very real challenge to states thinking about innovative approaches to assessment are the costs (both financial and instructional time) involved in using such measurement strategies, as well as very real technical concerns about these new approaches to assessment. Although they have a strong advantage of illustrating better approaches to learning and teaching, alternative assessments may be less reliable for reporting individual student or school results and certainly are more expensive. Therefore, in recent years, several states have considered the use of a "mixed" assessment model in which students are assessed with a combination of multiple-choice

and open-ended exercises. This approach has the advantage of allowing states to assess more content but at lower cost than an entirely open-ended assessment. Kentucky has and will be using this approach and Massachusetts is considering it.

Another approach to broader content coverage is the use of every-pupil matrix sampling designs. This approach is useful where school and district information is more important than individual student results. Kentucky has used this approach for several years.

Professional Development on Assessment

Attention to the forms of assessment used at both the state and local levels has encouraged another trend at the state level. As state-level educators have debated the form(s) of assessment appropriate for the state to use, increasing attention has been paid to the training of classroom teachers to collect and use information that might be gathered from such innovative approaches to assessment within their classrooms. This trend is actually the convergence of several trends, including changes in student standards to emphasize thinking and problem-solving skills (while deemphasizing memorization of content knowledge), plus support for alternative approaches to assessment, such as projects, exhibitions, demonstrations, and the use of portfolios. The result is that many local districts and some state agencies are now providing classroom teachers with assessment learning experiences that they can apply in their classrooms. This attention to professional development on assessment for classroom teachers is particularly important given that few, if any, teachers receive much in the way of preservice training on assessment, and that the understanding of appropriate uses and interpretation of assessment information is critical to the improvement of learning.

Norm-Referenced Tests

When the ASAP group began meeting in 1977, the most commonly used assessments were commercially available (off-the-shelf) norm-referenced tests. Despite the attention to forms of measurement such as criterion-referenced assessments, which are more widespread today than 20 years ago, it is interesting to note that norm-referenced tests are still the predominant form of large-scale assessment in the United States. In 1993, 31 states used norm-referenced tests; in 1994, 30; and in 1995, 31.

There had been an expectation that the number of states using NRTS would decrease in 1995 given the deemphasis on norm-referenced assessments in the Improving America's Schools Act (IASA), the reauthorization of the Elementary and Secondary Education Act. States are no longer required to use such assessments for the evaluation of Title I compensatory education programs nor the monitoring of individual Title I student improvement. This was a major change in the legislation, which advocacy groups and others fought for and won. In place of such tests, states are required to develop and operate "comprehensive assessment systems" capable of reporting whether individual students and school programs are making "adequate yearly progress."

Two events conspired to confound this prediction. First, the November 1994 election brought to power chief state school officers, state board of education members, legislators, and governors with strongly held ideas about student standards and assessment that were oftentimes contrary to the spirit of using new forms of assessment to raise standards. Given problems in some of the assessment efforts first implemented (in Arizona, California, Georgia, and Maine, to name a few), policymakers pushed to set aside innovative approaches to assessment and to return to commercially available norm-referenced tests. While such debates and changes are still taking place in some states, they bear watching in the future.

Second, the changes implemented in the IASA legislation have proven to be less far-reaching than originally thought. Due to political changes in Washington, D.C., states will be required to change their statewide assessments substantially less than they had anticipated. States, for example, have five to six years to develop permanent comprehensive assessment systems (and those in only mathematics and reading, not in all of the national goal areas, unless they do so for all students). In the interim, transitional assessments of any type (norm-referenced, criterion-referenced, or performance assessments) can be used at state choice so long as they are deemed to "measure challenging state content standards," which is left poorly defined in the federal legislation.

For these reasons, as well as because many policymakers desire to have comparative data using test instruments developed outside of the state, it is likely that norm-referenced tests will continue to be a major type of assessment used in states. To satisfy this desire for normative information, but using measures of higher-level standards, some states (such as Kentucky and North Carolina) have administered the NAEP assessments to samples of students taking their statewide assessments in order to provide NAEP-like scores to buildings and districts (as well as the state). This recent innovation in providing normative information has the promise of allowing states to pursue new forms of assessment while still providing external referents for scores on the statewide assessments. It is even possible that some form of individual student NAEP tests might be made available as well. It will be interesting to monitor the success of these efforts and to determine if this becomes a trend for the future.

National Efforts at Joint Development

Another trend is worth noting. Until 1990, most assessment development was carried out by individual states working alone or with the assistance of a contractor. Since then, two innovations in collaboration among the states have taken place. The first is the New Standards project, codirected by the University of Pittsburgh and the National Center for Education and the Economy, which has been working with a number of states and local districts to design and develop an innovative assessment system that will encourage thoughtful student learning in areas such as mathematics, language arts, and science. The second is the Council of Chief State School Officers' State Collaborative on Assessment and Student Standards (SCASS), which currently has 11 projects in which states work together to develop innovative student assessments. Both of these activities mark a first for collaboration among the states. The states are actively working together to develop assessments from which states share and use the products rather than simply exchanging information about innovative assessment approaches, as has been the case in the past.

Future Issues and Their Impact on State Assessment

Overall, an examination of the changes in large-scale assessment programs during the past 20 years shows a substantial change in the number of states with such programs, the subject areas assessed, and the types of assessment measures used, as well as the types of assessment measures being developed (and the manner in which this development is proceeding). These changes have only increased in the past few years with the considerable public attention paid to the quality of schools. Not surprisingly, these changes have led a number of states to re-examine assessment program designs that were adopted in years past. A number of states are examining whether their current assessment design is still adequate and are looking at how such recent programs as NAEP, the New Standards project, and SCASS fit within their overall assessment design. Given the number of states that are conducting such examinations, further changes in the nation's large-scale assessment programs are likely. Of course, it may take several years for these changes to be implemented.

Several trends appear at the state and local levels that may have a long-term impact on the shape of large-scale assessment programs at the state level. Certainly, the current emphasis on performance or alternative assessments is not going to disappear. Although there have been some successes (such as in Maryland and Kentucky), the set-backs in California, Arizona, Indiana, and elsewhere indicate that widespread acceptance of performance assessment is certainly not automatic. Technical issues need to be addressed in a sound manner, and policymakers and the public need to understand the reasons for such measures, the student standards that they measure, and the reasons why both innovative standards and assessments are needed. States and others interested in innovative forms of assessment will need to make sure important parties are "on board" before engaging in this new development work.

Certainly, there will be some impact from the drive now under way in some states to "deregulate" public education and return control of it to local school districts. While this drive is taking several forms, it would not be unexpected for these pressures to affect the extent and types of student assessment in the future. In some states, this trend may mean less attention to statewide student expectations and measures, while in other places, it may mean just the opposite.

The pressure to provide appropriate assessment training and experiences to classroom teachers is also not likely to abate. The collaborative work across states is likely to spread innovative approaches to assessment more quickly than it has in the past. In addition, the outside political pressures to use assessment as a tool for reform of schools is not likely to lessen. Changes brought about by federal legislation such as Goals 2000 and IASA will occur as well, but perhaps at a slower pace than once thought. In addition, it is uncertain how the battles between chief state school officers and governors shaping up over control of education funds in federal block grant programs will affect large-scale student assessment programs.

Finally, the reauthorization of the NAEP program brought several changes that also may affect states. In recent years, NAEP has offered the trial state NAEP programs, but, unfortunately, recent appropriations for the program plans have not permitted a full-scale state NAEP program to be offered. If the program is funded at a higher level, it might affect the number of states that administer norm-referenced tests to students at one or more grade levels, since the NAEP data provide the types of national comparisons that states desire that are more current, less expensive, and more technically sound. This year, the National Assessment Governing Board, the policymaking board for NAEP, has suggested a number of changes to the programs. It is uncertain at this point how many of these changes will be implemented, what the shape of the program will be in the future, nor how the NAEP of the future will affect states.

Many swirling, cross-cutting trends at the state level are affecting large-scale assessment programs, and it is likely that these trends will affect the nature of statewide assessments in the future. With the State Student Assessment Program Database, it should be easier to track the course of changes in large-scale assessment programs at the state level. Future editions of this report will begin to indicate more precisely just how such changes are occurring.

Resources

Bond, L. A., & King, D. (1995). *State high school graduation testing: Status and recommendations.* Oak Brook, IL: North Central Regional Educational Laboratory.

Mehrens, W. (1992). *Issues and recommendations regarding implementation of high school graduation tests.* Oak Brook, IL: North Central Regional Educational Laboratory.

Mehrens, W. (1995). *Issues to consider in moving beyond a minimal competency high school graduation test.* Greensboro, NC: SouthEastern Regional Vision for Education.

National Governors Association. (1996). *1996 National Education Summit policy statement.* Washington, DC: Author.

Phillips, G. (1996, January). *The National Center on Educational Statistics: Development work and research on adaptations and alternative assessments for students with disabilities.* A presentation given at a Council of Chief State School Officers SCASS Special Education Assessment meeting in Tampa, Florida.

Phillips, S. E. (1993). *Legal implications of high-stakes assessment: What states should know.* Oak Brook, IL: North Central Regional Educational Laboratory.

Phillips, S. E. (1995). *All students, same test, same standards: What the new Title I legislation will mean for the educational assessment of special education students.* Oak Brook, IL: North Central Regional Educational Laboratory.

Ysseldyke, J. (1996, January). *Overview of the issues of participation and inclusion of students with disabilities in large-scale assessment programs.* A presentation at a Council of Chief State School Officers SCASS Special Education Assessment meeting in Tampa, Florida.

Appendices

SSAP Summary Table

This table summarizes a significant amount of information for the SSAP database and is somewhat complex. Please keep the following in mind when reading the table.

Most states conduct several assessment programs side by side (labeled #COM, for components). This table aggregates across these components. It should be read, emphasizing the term "at least," in the following sense: Alaska conducts at least one program assessing all fourth or sixth or eighth graders in language arts or math or writing; it also assesses at least some fifth and tenth graders in language arts or math or writing. Alaska makes use of a norm-referenced multiple-choice test and a writing sample. These assessments are conducted to diagnose or place students, to improve instruction, to evaluate programs, and to generate reports on school performance.

Summary Table: Statewide Assessment Programs, School Year 1994-1995

Grades

ST	# COM	Kg	1	2	3	4	5	6	7	8	9	10	11	12
AK	2	-	-	-	-	A	V	A	-	A	-	V	-	-
AL	6	-	-	-	A	A	A	A	A	A	A	A	A	-
AR	1	-	-	-	-	A	-	-	A	-	-	A	-	-
AZ	1	-	-	-	-	A	-	-	A	-	-	A	-	-
CA	2	-	-	-	-	-	-	-	-	-	V	V	V	V
CO	0	-	-	-	-	-	-	-	-	-	-	-	-	-
CT	2	-	-	-	-	A	-	A	-	A	-	A	-	-
DE	1	-	-	-	A	-	A	-	-	A	-	A	-	-
FL	3	-	-	-	-	A	-	-	-	A	-	A	A	-
GA	5	A	-	-	A	-	A	-	-	A	-	-	A	-
HI	3	-	-	-	A	-	-	A	-	A	-	A	-	-
IA	0	-	-	-	-	-	-	-	-	-	-	-	-	-
ID	2	-	-	-	-	A	-	-	-	A	-	-	A	-
IL	1	-	-	-	A	A	-	A	A	A	-	A	A	-
IN	1	-	-	A	A	-	-	A	-	A	A	-	-	-
KS	1	-	-	-	A	A	A	-	A	A	-	A	A	-
KY	3	-	-	-	-	A	A	-	-	A	-	-	A	A
LA	4	A	-	-	A	A	A	A	A	-	-	A	A	-
MA	0	-	-	-	-	-	-	-	-	-	-	-	-	-
MD	3	-	-	-	A	-	A	-	A	A	-	-	-	-
ME	1	-	-	-	-	A	-	-	-	A	-	-	A	-
MI	2	-	-	-	-	A	A	-	A	A	-	A	A	-
MN	1	-	-	-	-	V	V	V	-	V	V	-	V	-
MO	2	-	-	-	S	-	S	S	-	S	-	S	S	-
MS	3	-	-	-	-	A	A	A	A	A	A	-	A	-
MT	1	-	-	-	-	A	-	-	-	A	-	-	A	-
NC	3	-	-	-	A	A	A	A	A	A	A	A	A	-
ND	1	-	-	-	A	-	-	A	-	A	-	-	A	-
NE	0	-	-	-	-	-	-	-	-	-	-	-	-	-
NH	1	-	-	-	A	-	-	-	-	-	-	-	-	-
NJ	2	-	-	-	-	-	-	-	-	A	-	-	A	-
NM	4	-	A	A	A	A	A	A	-	A	-	A	-	-
NV	3	-	-	-	-	A	-	-	-	A	-	-	A	-
NY	6	-	-	-	A	A	A	A	-	A	A	A	A	V
OH	4	-	-	-	-	A	-	A	-	A	A	-	-	A
OK	2	-	-	-	A	-	A	-	A	A	-	-	A	-
OR	2	-	-	-	A	-	A	-	-	A	-	-	A	-
PA	2	-	-	-	-	-	A	V	-	A	V	-	A	-
RI	2	-	-	-	-	A	-	-	-	A	-	A	-	-
SC	2	-	-	-	A	A	A	A	A	A	A	A	A	-
SD	2	-	-	-	-	A	-	-	-	A	A	-	A	-
TN	4	-	-	A	A	A	A	A	A	A	A	A	A	-
TX	1	-	-	-	A	A	A	A	A	A	-	A	-	-
UT	3	-	V	V	V	V	A	V	V	A	V	V	A	V
VA	2	-	-	-	-	A	-	A	A	A	A	A	A	-
VT	2	-	-	-	-	V	V	-	-	V	-	-	-	-
WA	1	-	-	-	-	A	-	-	-	A	-	-	A	-
WI	2	-	-	-	A	A	-	-	-	A	-	A	-	-
WV	3	-	A	A	A	A	A	A	A	A	A	A	A	-
WY	0	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	45	2	3	5	23	32	24	22	17	40	15	25	30	5

Subjects and Test Type

ST	LA	R	W	M	Sc	SS	Voc	Ap	NRT	CRT	WS	Per	Prt	COR
AK	A	A	V	A	-	-	-	-	A	-	V	-	-	-
AL	A	A	A	A	A	A	-	A	A	A	A	-	-	A
AR	A	A	-	A	A	A	-	-	A	-	-	-	-	-
AZ	A	A	-	A	-	-	-	-	A	-	-	-	-	-
CA	-	-	-	V	V	V	V	-	V	V	V	V	V	-
CO	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CT	A	A	A	A	A	-	-	-	-	A	A	A	-	-
DE	-	A	A	A	-	-	-	-	A	-	A	A	-	-
FL	-	A	A	A	-	-	-	-	A	A	A	-	-	-
GA	A	A	A	A	A	A	-	-	A	A	A	A	-	A
HI	A	A	-	A	-	-	-	-	A	A	-	-	-	-
IA	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ID	A	A	A	A	-	-	-	-	A	-	A	-	-	-
IL	-	A	A	A	A	A	-	-	A	A	A	-	-	-
IN	A	-	-	A	A	A	-	-	A	A	-	-	-	-
KS	-	A	-	A	A	A	-	-	-	A	-	A	-	-
KY	-	A	A	A	A	A	A	-	-	-	-	S	A	A
LA	A	A	A	A	A	A	-	-	A	A	A	A	-	-
MA	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MD	A	A	A	A	A	A	-	-	S	A	A	A	-	-
ME	-	A	A	A	A	A	-	-	-	-	A	-	-	A
MI	-	A	-	A	A	-	S	-	-	A	-	-	S	-
MN	V	V	V	V	V	V	-	-	-	V	V	V	-	-
MO	S	-	S	S	S	S	-	-	-	S	S	-	-	-
MS	A	A	A	A	-	-	-	-	A	A	A	A	-	-
MT	A	A	-	A	A	A	-	-	A	-	-	-	-	-
NC	A	A	A	A	A	A	-	-	S	A	A	-	-	A
ND	A	A	-	A	A	A	-	-	A	-	-	-	-	-
NE	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NH	A	-	A	A	-	-	-	-	-	A	A	-	-	A
NJ	-	A	A	A	-	-	-	-	-	A	A	A	-	A
NM	A	A	A	A	A	A	-	-	A	A	A	L	L	-
NV	A	A	A	A	-	-	-	-	A	-	A	-	-	-
NY	-	A	A	A	A	A	V	-	-	A	A	A	-	A
OH	A	A	A	A	-	-	-	-	A	A	A	A	-	A
OK	A	A	A	A	A	A	-	-	A	A	A	-	-	-
OR	-	A	A	A	-	-	-	-	-	A	A	-	-	-
PA	-	A	V	A	-	-	-	-	-	A	V	A	-	-
RI	-	A	A	A	-	-	-	-	A	-	A	-	-	-
SC	A	A	A	A	A	-	-	-	A	A	A	-	-	-
SD	A	-	-	A	A	A	-	A	A	-	-	-	-	-
TN	A	A	A	A	A	A	-	-	A	A	A	-	-	-
TX	-	A	A	A	A	A	-	-	-	A	A	-	-	A
UT	A	A	V	A	A	A	-	-	A	V	V	V	-	-
VA	A	A	A	A	A	A	-	-	A	A	A	-	-	-
VT	-	-	V	V	-	-	-	-	-	V	V	V	V	-
WA	A	A	-	A	A	A	-	-	A	A	-	-	-	-
WI	A	A	A	A	A	A	-	-	A	A	A	-	-	-
WV	A	A	A	A	A	A	-	-	A	A	A	A	-	-
WY	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	30	39	34	45	30	27	4	2	31	33	34	17	4	10

Testing Program Purposes and Testing Program Consequences

ST	StD	Imp	PE	StA	StP	HD	ED	HSG	ScA	SPR	HSk	Acr	FGR	ER	War	PWL	FL	ACL	T	D	FR	CSGr	PSr	CSL	FP
AK	A	A	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AL	A	A	A	-	-	-	-	A	-	A	-	-	-	-	A	A	-	A	A	-	-	A	-	-	-
AR	A	A	A	-	-	-	-	-	-	-	-	-	-	-	A	A	-	-	-	-	-	-	-	-	-
AZ	A	A	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CA	-	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CT	A	A	A	A	-	-	-	-	-	A	A	-	A	-	-	-	-	-	-	-	-	-	-	-	-
DE	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FL	-	A	A	A	-	-	-	A	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GA	A	A	A	-	-	-	-	A	A	A	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-
HI	A	A	A	-	-	-	-	A	-	A	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-
IA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ID	-	A	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
IL	-	-	-	-	-	-	-	-	-	A	-	A	-	A	-	A	A	A	-	-	-	-	-	-	-
IN	A	A	A	-	A	-	-	-	A	A	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-
KS	-	A	A	-	-	-	-	-	-	A	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-
KY	-	A	A	-	-	-	-	-	A	-	-	-	-	A	A	A	A	-	A	A	A	-	S	-	S
LA	A	A	A	A	A	-	-	A	-	A	A	-	A	-	-	-	A	-	-	-	-	-	-	-	-
MA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MD	A	A	A	-	-	-	-	A	A	A	A	A	A	-	A	A	-	A	A	-	-	-	-	-	-
ME	-	A	A	A	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MI	-	A	A	A	-	-	A	-	A	A	-	A	-	-	A	A	A	A	A	A	-	-	-	-	-
MN	-	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MO	S	S	S	S	-	-	-	-	-	-	-	S	-	-	-	-	-	S	-	-	-	-	-	-	-
MS	A	A	A	-	-	-	-	A	-	-	-	A	-	A	A	A	-	A	A	-	-	-	-	-	-
MT	-	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NC	A	A	A	-	-	-	-	A	-	A	-	A	A	A	A	A	-	A	-	-	A	-	-	-	-
ND	A	A	A	A	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-
NE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NH	-	A	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NJ	A	A	A	-	-	-	-	A	-	A	-	A	-	-	-	A	A	A	A	-	-	-	-	-	-
NM	A	A	A	-	A	-	-	A	-	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NV	A	A	-	-	-	-	-	A	-	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NY	A	A	A	V	-	V	V	V	-	A	-	-	-	-	-	A	-	-	-	-	L	L	L	L	L
OH	-	A	A	-	-	A	-	A	-	A	-	-	-	-	A	A	-	-	-	-	-	-	-	-	-
OK	A	A	A	-	-	-	-	-	-	A	-	-	-	A	A	A	-	A	A	A	-	-	-	-	-
OR	-	A	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PA	-	A	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RI	A	A	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SC	A	A	A	-	A	-	-	A	A	A	A	-	A	A	A	A	A	-	-	-	-	-	-	-	-
SD	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TN	A	A	A	-	-	A	A	A	A	A	-	-	A	-	A	A	-	-	A	-	-	-	-	-	-
TX	A	A	A	A	-	-	-	A	A	A	A	A	-	A	A	A	-	A	A	A	-	-	-	-	-
UT	A	A	A	A	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VA	A	A	A	A	A	-	-	A	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VT	-	V	V	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WA	A	A	A	-	-	-	-	-	-	A	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-
WI	A	A	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WV	-	A	A	-	-	-	-	-	-	A	A	A	-	-	-	-	-	A	-	-	-	-	-	-	-
WY	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	27	44	39	12	5	3	3	17	8	35	8	12	8	8	12	15	6	11	9	4	2	1	1	0	1

Note: CO and MA suspended their statewide assessment programs in 1994-1995. CA suspended the mandatory part of their statewide assessment program in 1994-1995.

Summary Table Legend

Notes: Totals are computed by adding up the number of As, Ss, and Vs in each column. In any cell, only one description is included. If multiple component descriptions were present in the same cell, then this cell was coded to match the component testing the most students. See Part III responses of the SSAP database for component names and component-level descriptions.

Cell Contents
(indicates at least one component where)
A = All students are tested
S = Students are sampled
V = Inclusion is voluntary for students, schools, or school districts
L = Decision is made at local level

#COM = Number of Assessment Components

Subjects
LA = Language Arts
R = Reading
W = Writing
M = Math
Sc = Science
SS = Social Studies
Voc= Vocation Ed
Ap = Aptitudes

Test Type

Multiple-Choice Testing
NRT=Norm-Referenced Testing
CRT=Criterion-Referenced Testing

Writing Assessment
WS =Writing Samples

Alternative Assessments
Per=Performance Testing
Prt=Portfolio Assessment
COR=Constructed, Open-Response Items

Testing Program Purposes

Instructional Process
StD=Student diagnosis or placement
Imp=Improvement of instruction, curriculum
PE =Program evaluation

Student Accountability
StA=Student awards or recognition
StP=Student promotion
HD =Honors diploma
ED =Endorsed diploma
HSG=High school graduation (exit requirement)

School Accountability
ScA=School awards or recognition
SPR=School performance reporting
HSk=High school skills guarantee
Acr=School accreditation

Testing Program Consequences

For Schools
FG =Funding gain
ER =Exemption from regulations
War=Warnings
PWL=Probation, watch lists
FL =Funding loss
AcL=Accredition loss
T =Takeover
D =Dissolution

For School Staff
FR =Financial rewards
CSG=Certification status gain
Pr =Probation
CSL=Certification status loss
FP =Financial penalties

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").