DOCUMENT RESUME

ED 366 616 TM 020 987

AUTHOR Barton, Paul E.; Coley, Richard J.
TITLE Testing in America's Schools. Policy Information Report.
INSTITUTION Educational Testing Service, Princeton, NJ. Policy Information Center.
PUB DATE 94
NOTE 46p.
AVAILABLE FROM Policy Information Center, Mail Stop 04-R, Educational Testing Service, Rosedale Road, Princeton, NJ 08541-0001 ($7.50 prepaid).
PUB TYPE Statistical Data (110) -- Reports - Evaluative/Feasibility (142)

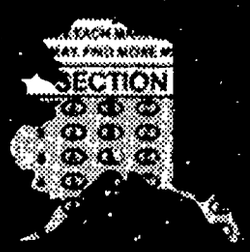
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Ability; Accountability; *Achievement Tests; Criterion Referenced Tests; Disadvantaged Youth; Educational Diagnosis; Educational Practices; *Educational Testing; Elementary Secondary Education; Essay Tests; Ethnic Groups; Instructional Improvement; Language Tests; Mathematics Tests; Multiple Choice Tests; Norm Referenced Tests; Portfolios (Background Materials); Profiles; Program Evaluation; Racial Differences; School Districts; Standardized Tests; *State Programs; Teacher Made Tests; *Testing Programs; *Test Use; Writing Tests
IDENTIFIERS *Alternative Assessment; North Central Regional Educational Laboratory; *Performance Based Evaluation; State Student Assessment Prog Database 1992 93

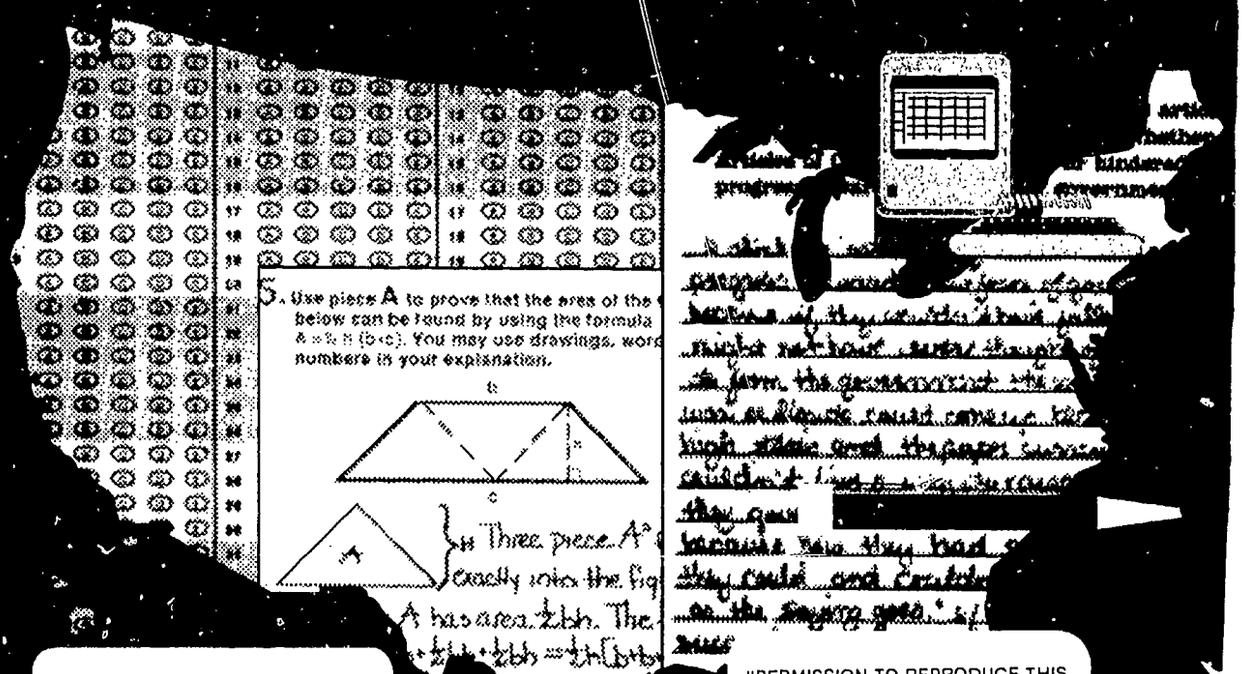
ABSTRACT

This report provides a profile of state testing programs in 1992-93, as well as a view of classroom testing practices by state, school district, school, or individual teacher. Information, taken from a variety of sources, including the National Assessment of Educational Progress and a General Accounting Office study, indicates that the multiple-choice test remains dominant. The most prevalent purposes of state programs are accountability, instructional improvement, and program evaluation. Virtually all states test in mathematics and language, and most also test in science, writing, and social studies. Thirty-eight programs include writing samples, and 34 states use norm-referenced tests while 34 use criterion-referenced tests. Seventeen use some form of performance assessment, and six collect student portfolios. At least 36 percent of all students were tested in state programs in 1992-93. Only one state uses a norm-referenced test for high school graduation purposes, while 20 use criterion-referenced tests. In the classroom, in contrast, non-multiple-choice tests appear to be the predominant mode. It is also concluded that patterns of traditional and alternative testing in the classroom are very similar for students of different races, ethnicity, ability groups, and resource adequacy. Fifteen figures and three tables present study information. Document notes refer readers to information sources. Four appendix tables amplify the information. (SLD)

POLICY INFORMATION REPORT

TESTING IN AMERICA'S SCHOOLS





Educational Testing Service
Princeton, New Jersey 08541-0001

Contents

This report was written by Paul E. Barton and Richard J. Coley of the ETS Policy Information Center.

Additional copies of this report can be ordered for $7.50 (prepaid) from:

Policy Information Center
Mail Stop 04-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541-0001
(609) 734-5694

Preface

There has been an explosion in testing in the United States over the past 20 years or so. It has almost reached the point where education reform, at least in some circles, has come to mean "more testing." Given the volume of school testing, and the variety of purposes it serves, it is hard to portray accurately both the quantity and kinds of testing occuring in today's classrooms. This report attempts to gather together new information on this topic. As the nation enters a new era of expecting more from tests, while at the same time broadening the use of various forms of performance assessment, we need a baseline from which to observe change.

Paul E. Barton
Director
Policy Information Center

Acknowledgments

Much of the data in this report is drawn from the *State Student Assessment Program Database, 1992-93*, produced by the Council of Chief State School Officers (CCSSO) and the North Central Regional Educational Laboratory (NCREL). This database contains descriptions of the state testing programs in the United States, and we are grateful to the state testing directors for supplying the information. We are also grateful to the National Assessment of Educational Progress, from which we used data.

The report was reviewed by Ed Roeber of CCSSO and by Linda Bond and Arie van der Ploeg of NCREL. At ETS, the report was reviewed by John Mazzeo, Howard Wainer, and Michael Zieky. Shilpi Niyogi was the editor, David Freund provided data analysis of NAEP, Carla Cooper supplied desktop publishing services, and Ric Bruce designed the cover.

Summary and Highlights

Testing in America's Schools provides a profile of state testing programs in 1992-1993, as well as a view of classroom testing practices, whether required by the state, school district, school, or individual teacher. Tests used at the state or district level, of course, are typically standardized, whether they are multiple-choice or alternative assessments. Teacher-constructed tests typically are non-standardized, and may include a variety of tasks, including multiple-choice, short response, essay, and performance tasks.

The nation is entering an era of change in testing and assessment. Efforts at both the national and state levels are now directed at greater use of performance assessment, constructed response questions, and portfolios based on actual student work. However, as will be seen in the following pages, the multiple-choice, machine-scored approach is still very much dominant. At the same time, however, alternative assessment approaches seem to be spreading rapidly, and have reached further into the testing system than many may be aware. The states are moving purposefully in pursuing alternative forms of assessment, but show no signs yet of abandoning traditional assessment programs.

This report tries to convey the present status of assessment in a way that can help to track changes as we head toward the year 2000. It is principally a statistical report, not an attempt to evaluate the quality of the testing programs of the U.S., or make judgments about the adequacy of our nation's testing practices. Highlights are presented below.

State Testing Programs

■ *Purposes* — The three most prevalent purposes of state testing programs are accountability, instructional improvement, and program evaluation. Next are student diagnosis/ placement and high school graduation.

■ *Subjects* — Virtually all states test students in mathematics and language, and most also test in science, writing, and social studies. A few test students in vocational subjects.

■ *Types of Tests* — Thirty-eight state programs include writing samples, 34 use norm-referenced tests and 34 use criterion-referenced tests. Seventeen use some form of performance assessment and six collect student portfolios.

■ *Volume of Testing* — At least 14.5 million students were tested in 1992-1993, representing at least 36 percent of all K-12 students. However, the testing is concentrated in certain grades, particularly at grades four and eight, where about half of the students are tested. Testing is lightest in kindergarten and grades 1, 2, and 12. Testing for exiting school is limited in the U.S., although it is standard practice in many countries.

■ *Testing for Accountability* — Of the states that test for accountability or program evaluation, *41 test every student in a grade.* Just one state uses sampling, and three use a combination of sampling and universal testing. The less intrusive sampling approach used by the National Assessment of Educational Progress has not been adopted by the states in their efforts to judge the effectiveness of schools and districts.

■ State testing programs use a variety of tests for different purposes:

• *For accountability and program evaluation purposes*, most state programs use norm- or criterion-referenced tests (36 and 40 programs, respectively). But 22 are using performance tests, eight are using portfolios, and nearly all the state programs use writing samples.

• *For diagnosis and placement purposes*, most state programs use traditional kinds of tests. Twenty-one and 24 programs, respectively, are using norm- and criterion-referenced tests; and 23 use writing samples. Ten use performance tests, and two use portfolios.

• *For student promotion purposes*, three and nine state programs, respectively, are using norm- and criterion-referenced tests; three use a performance test, two use portfolios, and six use writing samples.

• *For high school graduation purposes*, just one state program uses a norm-referenced test, and 20 use criterion-referenced tests. Two are using performance tests, 13 use writing

samples, and no state program uses portfolios for this purpose.

■ *Test Design* — Seventy percent of tests given in statewide systems are multiple-choice, 12 percent are writing samples, and 18 percent are multiple-subject performance tests.

■ *Kinds of Tests* — Eighty percent of the state systemwide tests given are achievement tests, 8 percent are aptitude tests, 6 percent assess vocational interest, 3 percent school readiness, and 3 percent assess other areas.

■ *Alternative Assessment* — States are busy developing and implementing alternative forms of assessment. Some are ready to use, some are in pilot stage, some are in the beginning stage, and some have been funded but not yet started. Fourteen states are involved with enhanced multiple-choice items, 18 with short-answer open-ended questions, 22 with extended-response open-ended items, 14 with individual performance assessments, seven with group performance assessments, nine with portfolios or learning records, five with projects or demonstrations, three with interviews, and four with observations.

■ These alternative item types are being used in the subjects generally tested for accountability purposes.

Testing in the Classroom

■ While the debate continues about switching more to alternative forms of assessment in system-wide testing, *non-multiple-choice testing appears to be the predominant mode in the classroom.*

• *Multiple-choice Tests.* At the fourth grade, 6 percent of students have teachers who give multiple-choice tests once or twice a week, 43 percent once or twice a month, and 51 percent yearly or never. The comparable percentages for eighth graders are 4, 30 and 66.

• *Problem Sets.* About half of the fourth graders have teachers who use problem sets once or twice a week, 39 percent once or twice a month, and 9 percent yearly or never. The comparable percentages for grade 8 are 58, 32, and 10.

• *Written Responses.* Forty-four percent of fourth graders are given tests requiring written responses at least monthly, 16 percent once or twice a year, and 40 percent never or hardly ever. For eighth graders the comparable percentages are 44, 22, and 33.

• *Projects, Portfolios, Presentations.* For grade 4 students, 20 percent are given these forms of assessment at least monthly, 25 percent once or twice a year, and 54 percent are never or hardly ever given these assessment types. For grade 8 the comparable percentages are 21, 32, and 47.

■ *Student reports on the frequency of math tests*

• Nine percent of students in fourth grade report taking math tests almost every day and 30 percent at least once a week.

• At grade 8, 6 percent of students report taking math tests almost every day and 55 percent at least once a week.

• Four percent of all twelfth graders report taking math tests almost every day and 50 percent at least once a week.

• Of twelfth graders taking math courses, 4 percent report taking math tests almost every day, and 57 percent report taking them at least once a week.

■ *Teacher reports on the frequency of math tests*

• Among the 44 jurisdictions participating in the 1992 NAEP math assessment, the percentages of students taking multiple-choice tests at least once a week ranged from zero in Guam and Nebraska to 22 percent in the District of Columbia. The average was 4 percent.

• Nationally, about 60 percent of teachers give their own math tests to students at least once a week. The percentages vary from 88 percent in Louisiana to 40 percent in Oregon.

• *Testing Equity* — The patterns of student exposure to traditional and alternative tests in the classroom were very similar for students of different race/ethnicity, students in classes of different ability groupings, and students in classes with differing resource adequacy as reported by their teachers.

The nation has launched a serious effort to improve education, and testing is one of the levers being used to raise achievement standards. In the 1970s, we had an explosion in state standardized testing programs. The momentum carried into the 1980s under state leadership in the era of the Excellence Movement. As this gave way to a new wave of reform in the late 1980s, culminating in the Charlottesville Education Summit, calls were made for broadening educational assessment to include performance assessment, portfolios, and constructed responses.

The purpose of this Policy Information Report is to provide a current summary of testing in America's schools. Data are drawn from several sources, but the survey of state assessment programs conducted by the Council of Chief State School Officers and the North Central Regional Educational Laboratory (CCSSO/NCREL) is used as the major source of information on state programs.* Data from the National Assessment of Educational Progress (NAEP) and information from a U.S. General Accounting Office study are also used in this report. Those interested in the details and intricacies of particular state testing programs will need to obtain the CCSSO/NCREL survey (see page 38 for ordering information).

These data sources enable us to view testing in the schools through several different windows, and it is hard to see the whole house through any one of them. The report starts by examining statewide testing programs, where the development and deployment of broader forms of assessment are largely taking place. But that is not all of the standardized testing going on in the schools — individual districts and schools have their own testing programs, beyond what comes from the state capitol. And all this is only a fraction of the testing that occurs in the American classroom. For it is the individual teacher who conducts most of the testing and constructs most of the tests. So we also provide data from teachers on what kinds of tests they use, and how frequently they give them.

The report format presents data graphically on the right-hand page and provides a narrative description of the data, along with information about data sources, on the left-hand page.

*In the 1992-1993 school year, 46 states had statewide testing programs. Iowa, Nebraska, New Hampshire, and Wyoming were the four states without a statewide program. Most states have more than one "program." California's testing program, for example, has three components. *The Career-Technical Assessment Project* is used to determine a student's readiness to enter the workforce or enter post-secondary study; *Golden State Exams* are used to qualify students for awards; and *Performance Assessment, Grades 4, 5, 8, and 10* provide on-demand assessment used to support the state curriculum frameworks, to facilitate good instruction, and to demonstrate accountability.

The Purposes of State Testing Programs

States test their students for a variety of reasons.

- to demonstrate accountability to tax payers
- to use the results to improve instruction and evaluate the effectiveness of education programs
- to diagnose student strengths and weaknesses and help in student course placement
- to certify that students are ready for the next grade or to graduate from high school, or to certify their competence to employers
- to accredit or approve schools or school districts
- to identify students and schools for rewards
- to certify that children are ready for school

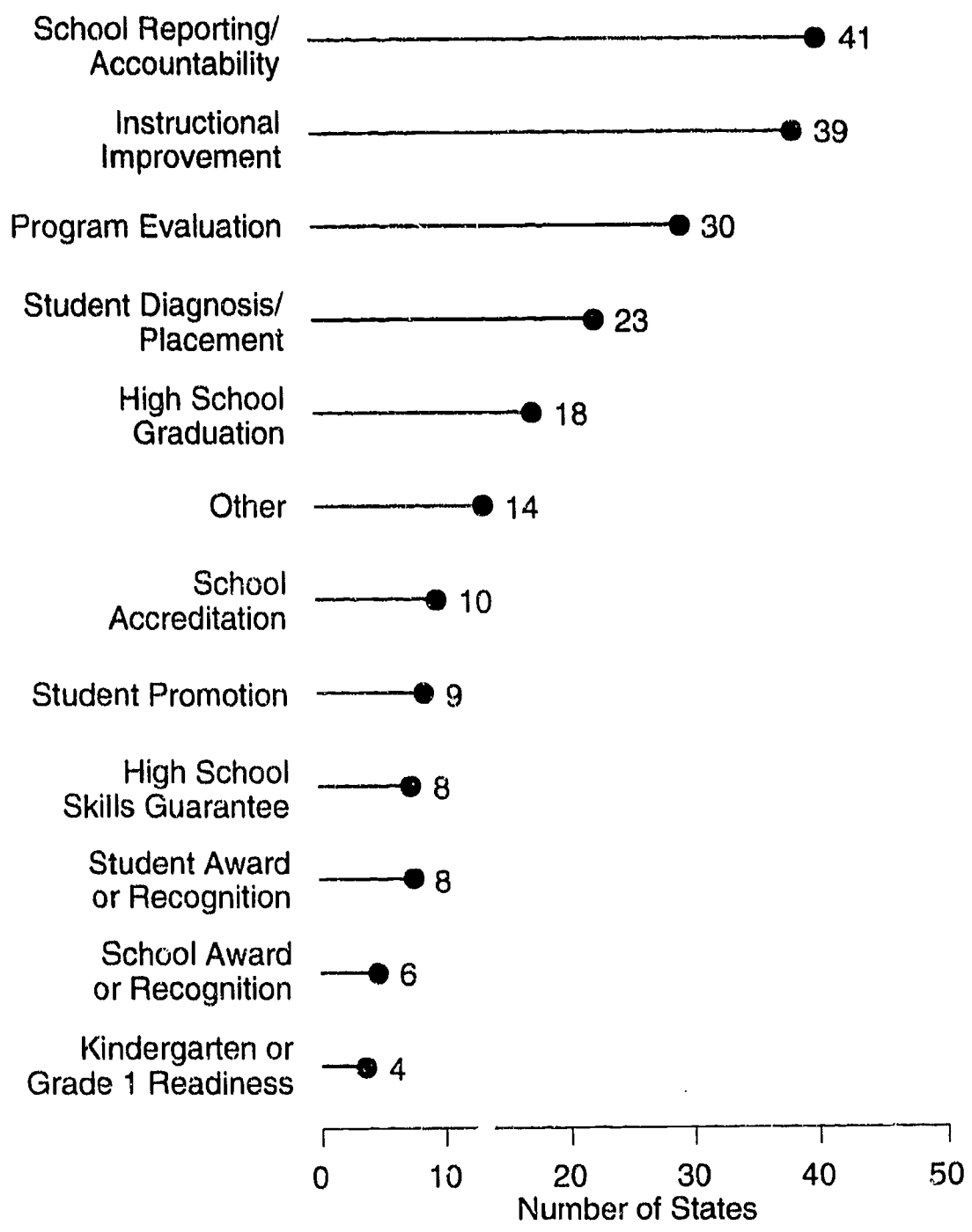
Figure 1 provides a count.* School performance reporting/accountability, instructional improvement, and program evaluation are the major purposes of state testing programs.

*While a few states have only one testing program, or component, most states have several different ones, each designed for a different purpose or purposes. For example, Louisiana uses the Kindergarten Development Readiness Screening Program to screen all entering kindergarten students; employs the Louisiana Education Assessment Program to test grade-appropriate state curriculum skills in grade promotion decisions and to allocate state funded remediation; and uses the Louisiana Statewide Norm-Referenced Testing Program to provide comparisons with other states, demonstrate school accountability, and provide a basis for program evaluation.

Source for Figure 1: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

Accountability, instructional improvement, and program evaluation are major purposes of state programs.

Figure 1: The Purposes of State Testing Programs, 1992-93



Subjects Tested in State Testing Programs

Virtually all of the states that have testing programs assess students in mathematics and language. Most include science, writing, and social studies as well. A few states test students in vocational subjects, as well as to determine readiness and aptitude. Figure 2 shows the number of states reporting assessment by subject.

Types of Tests Used in State Testing Programs

Thirty-eight states use a student writing sample as part of their assessment program. Norm-referenced tests and criterion-referenced tests* are each used by 34 states. Seventeen states use some form of performance assessment and six states collect student portfolios. These counts are shown in Figure 3.

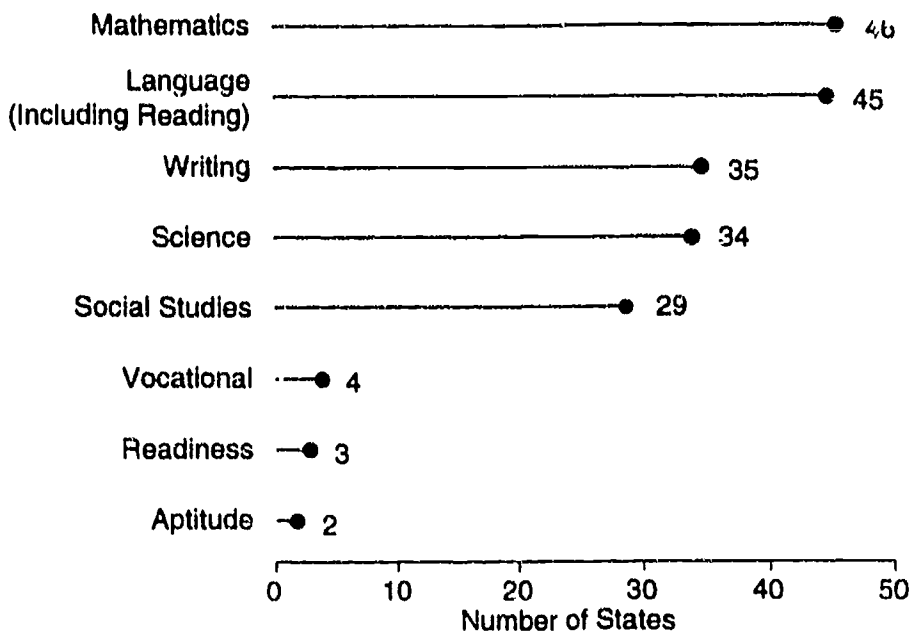
While the multiple-choice tests in this survey have been categorized as norm-referenced or criterion-referenced, this is not the case for the other types of tests reported, which could be either norm- or criterion-referenced.

*In a norm-referenced test, student performance is compared to that of other students taking the same test; in a criterion-referenced test, student performance is compared to a predetermined performance criteria.

Source for Figures 2 and 3: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

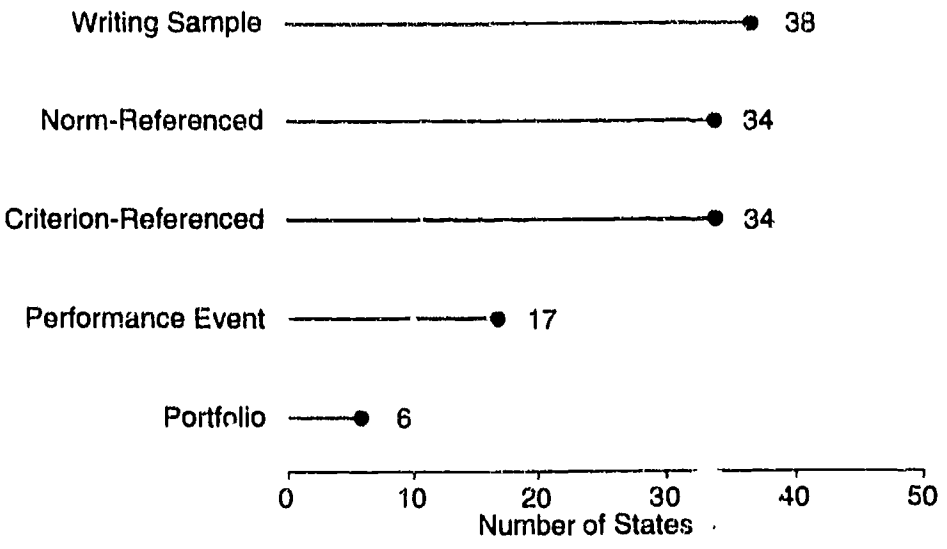
State testing programs assess students in core curriculum subjects.

Figure 2: Subjects Tested in State Testing Programs, 1992-93



States use a variety of test types in their testing programs.

Figure 3: Types of Tests Used in State Testing Programs, 1992-93



Number of Students Tested in State Testing Programs

During the 1992-1993 school year, the CCSSO/NCREL survey revealed that at least 14.5 million students were tested in state programs (this does not include students tested for district or classroom purposes).* Figure 4 allows a look at the grade levels of the tested students. While relatively few students are tested at the beginning and end of their school years, large numbers of students are tested between grades 3 and 11, with the highest numbers reported in grades 4 and 8.

The survey also reports that the 14.5 million students tested represented at least 36 percent of all students. Nationally, at grades 4 and 8, about half of all students were tested, on average. These numbers vary considerably, however, by state.

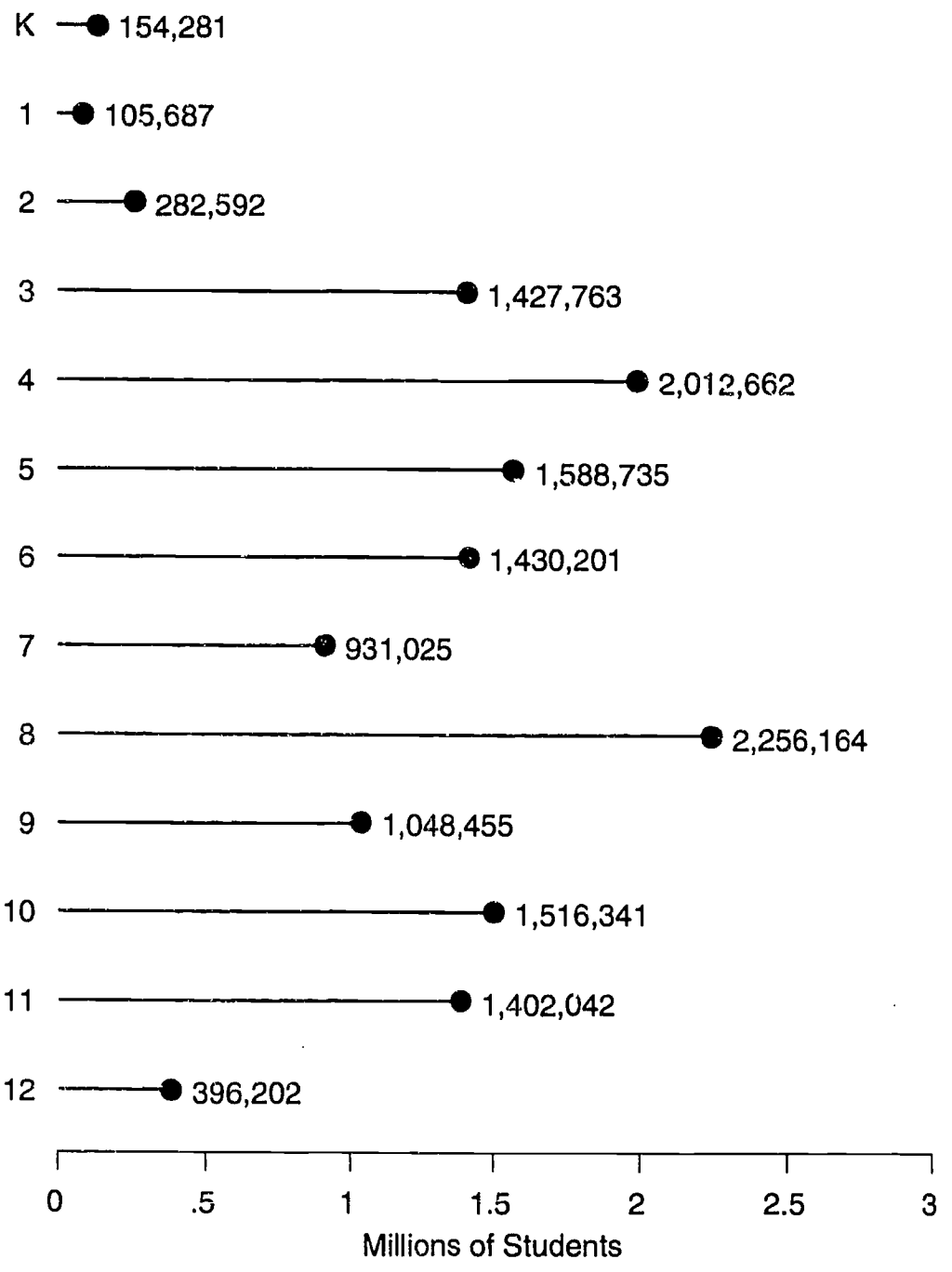
While it is common in many countries for graduating secondary school students to be tested, relatively few seniors are tested at all in the U.S., compared to most other grades.

*Since numbers were missing for several states, the counts presented here represent a minimum.

Source for Figure 4: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

Testing peaks in grades 4 and 8, and is lowest in grades K, 1, 2, and 12.

Figure 4: Number of Students Tested in State Testing Programs, by Grade, 1992-93



Universe and Sampling Approaches

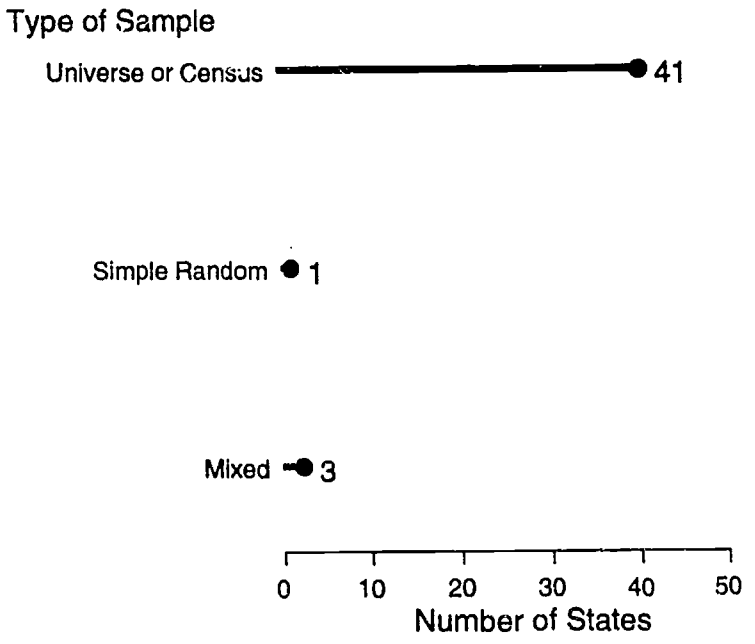
States were asked to specify the sampling frame they used for accountability and program evaluation assessment — the largest categories of state testing. The results are shown in Figure 5. Forty-one of the states indicated that they conduct "universal or census" testing, which means that they test all students within a chosen grade(s).

Only one state, Minnesota, uses a simple random sample of students. Three states use a mixture — Kansas samples students on math and reading, Kentucky samples students on a performance assessment, and Vermont uses sampling in its mathematics assessment. These three states conduct census testing for other program components, however.

Source for Figure 5: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

Nearly all of the states that test for accountability purposes test all students at the targeted grades.

Figure 5: Type of Student Samples Used for State Accountability and Program Evaluation Testing, 1992-1993



Types of Tests Used for Selected State Testing Purposes

As shown earlier, some important purposes of state testing programs are student diagnosis and placement, student promotion, program evaluation and accountability, and high school graduation. Figure 6 shows the types of tests used by state programs* for each of these purposes.

For student diagnosis and placement, traditional tests are typically used — norm- and criterion-referenced tests and writing samples. This pattern also holds for accountability and program evaluation purposes.

Alternative forms of assessment are not frequently used for student diagnosis and placement, although 10 state programs are now using performance assessment for this purpose. For accountability and program evaluation, however, it's interesting to note that 22 state programs use a performance test and eight use portfolios.

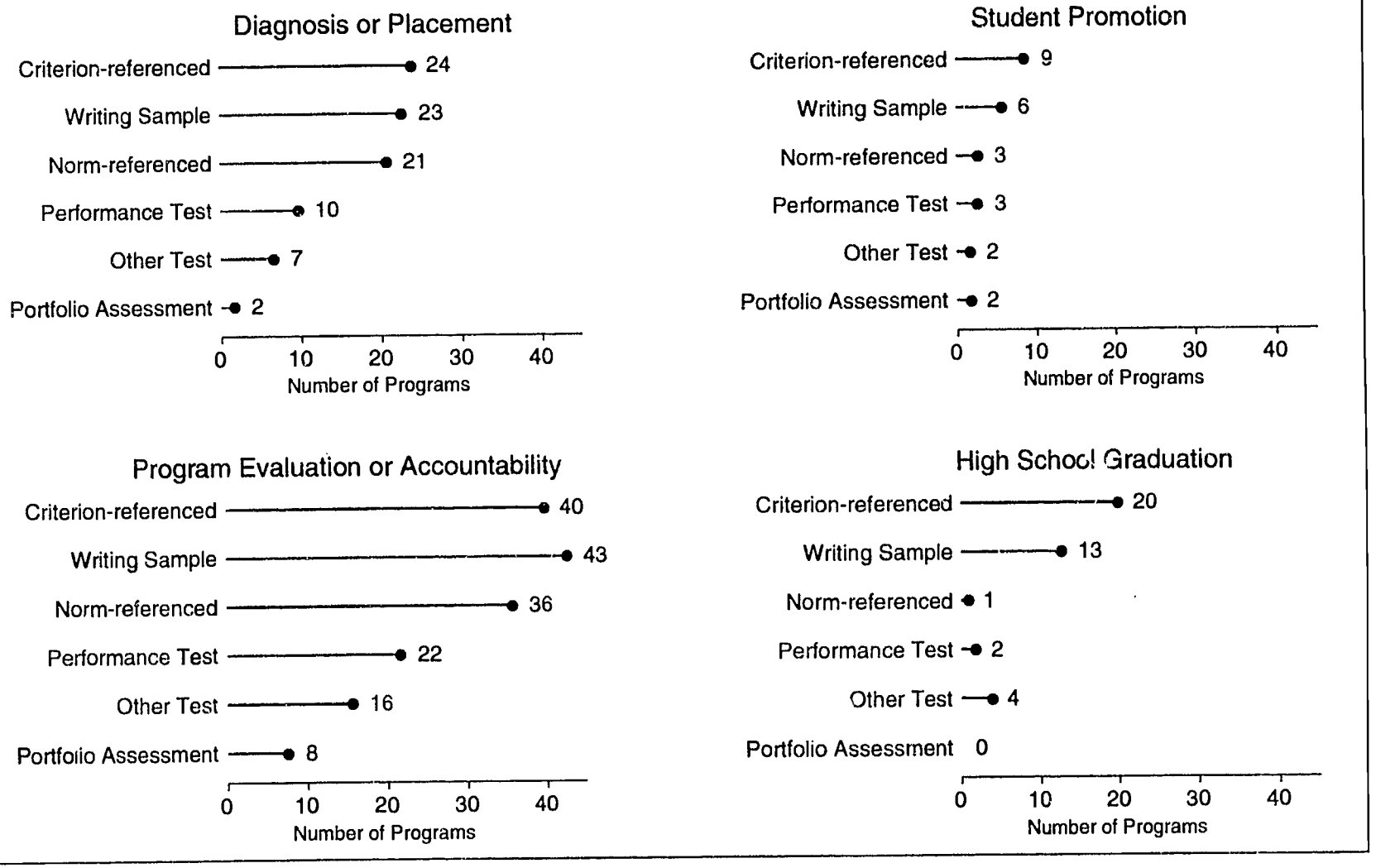
For gatekeeping purposes — student promotion and high school graduation — criterion-referenced tests prevail. However, only nine states are using criterion-referenced tests for promotion. In tests for promotion, students are assessed on what they are expected to know to advance within the educational system. Thirteen states also require students to complete a writing sample in order to earn a high school diploma.

*Note that numbers refer to the number of state programs, not the number of states. Some states use more than one testing program for a particular purpose.

Source for Figure 6: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

Alternative forms of assessment are now used in many state programs, particularly for program evaluation and accountability.

Figure 6: Types of Tests Used for Major State Testing Purposes, 1992-93



Test Design of Systemwide Tests

In its report to Congress on testing in the U.S., the General Accounting Office found while performance and criterion-referenced tests were more popular among educators, actual testing practice differed (Figure 7).

In its 1990-1991 survey of testing, the GAO found that multiple-choice tests constituted 70 percent of all tests given. Writing samples or multiple-choice tests with writing samples comprised 12 percent; multiple-subject performance tests constituted 18 percent.

Types of Systemwide Tests

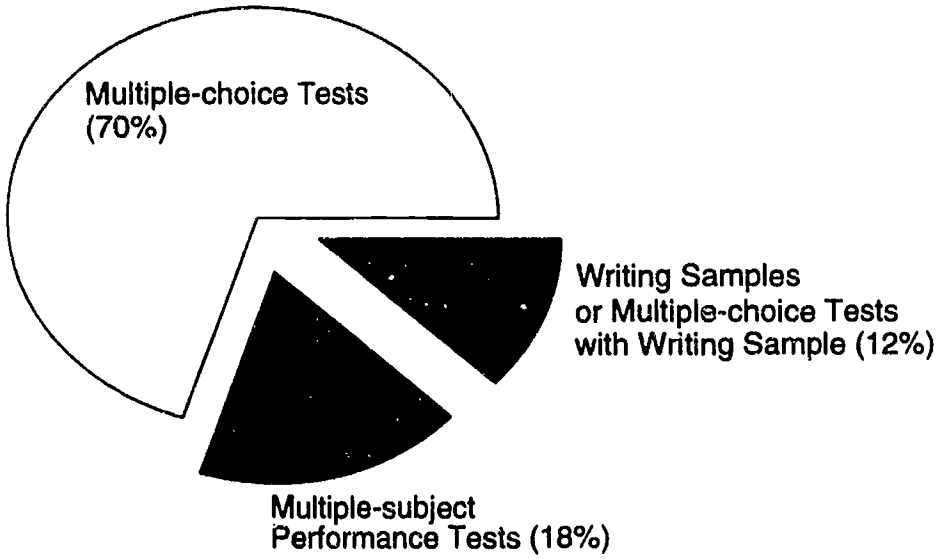
That same GAO study categorized systemwide tests according to the main purpose intended by the test-makers. As shown in Figure 8, 80 percent of all such tests taken in 1990-91 were achievement tests, those that attempt to measure a student's accumulated knowledge or skill. Most of these were commercial tests and many were adapted to match a state's curriculum.

Another 8 percent were designed to measure aptitude or ability, e.g., an intelligence test. Three percent of the systemwide tests were designed to measure school "readiness."

Source for Figures 7 and 8: Data are drawn from U.S. General Accounting Office. *Student Testing: Current Extent and Expenditures, with Cost Estimates for a National Examination*, January 1993.

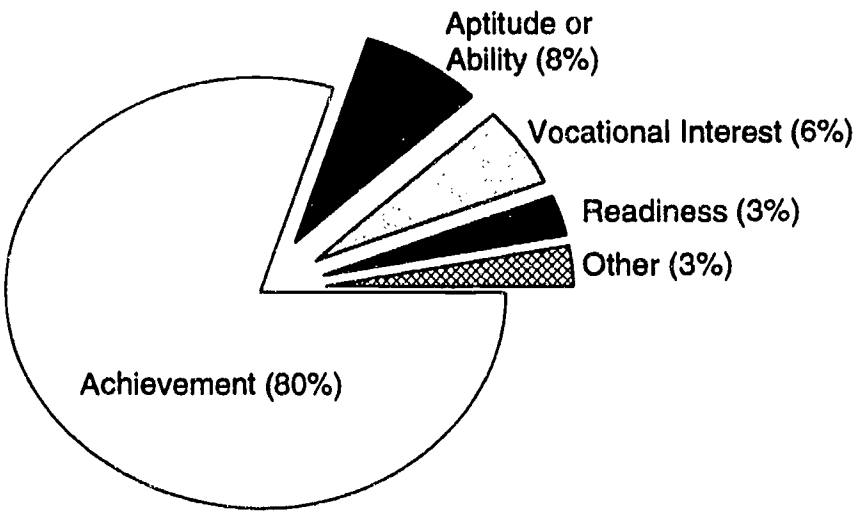
Multiple-choice tests continue to comprise the great majority of systemwide tests in the United States.

Figure 7: Test Design Features for Systemwide Tests in the U.S., 1990-1991



Achievement tests are the most commonly used systemwide tests in the United States.

Figure 8: Types of Systemwide Tests in the U.S., 1990-1991



Teacher Reports on the Frequency of Testing

In the 1992 mathematics assessment, the National Assessment of Educational Progress (NAEP) asked teachers how frequently they assessed their students using four different types of tests:

- Multiple-choice tests
- Problem sets*
- Written responses
- Projects, portfolios, or presentations

The results are summarized in Figure 9 for grades 4 and 8. In general, these data indicate that teachers are using what have been called "alternative" types of assessment in their classrooms. At both grade levels, over half of the students were taught by teachers who used problem sets once or twice a week; four out of ten were given assignments requiring written responses at least monthly; and about one-fifth were given projects, portfolios, or presentations at least monthly. Only 6 and 4 percent of fourth and eighth graders, respectively, were given multiple-choice tests once or twice a week.

While these data indicate that teachers are emphasizing alternative assessment much more than might be expected, and using multiple-choice tests less frequently, there is room for greater use of alternative assessments. Figure 9 also shows that about half of the students at both grade levels had teachers who never or hardly ever used projects, portfolios, or presentations as assessment tools in their classrooms.

*Problem sets are a set of one to five problem situations, word problems from the textbook, or constructed by the teacher (written or verbal).

In Figure 9, note that the frequency categories for multiple-choice tests and problem sets are different than for written responses and projects, portfolios, and presentations.

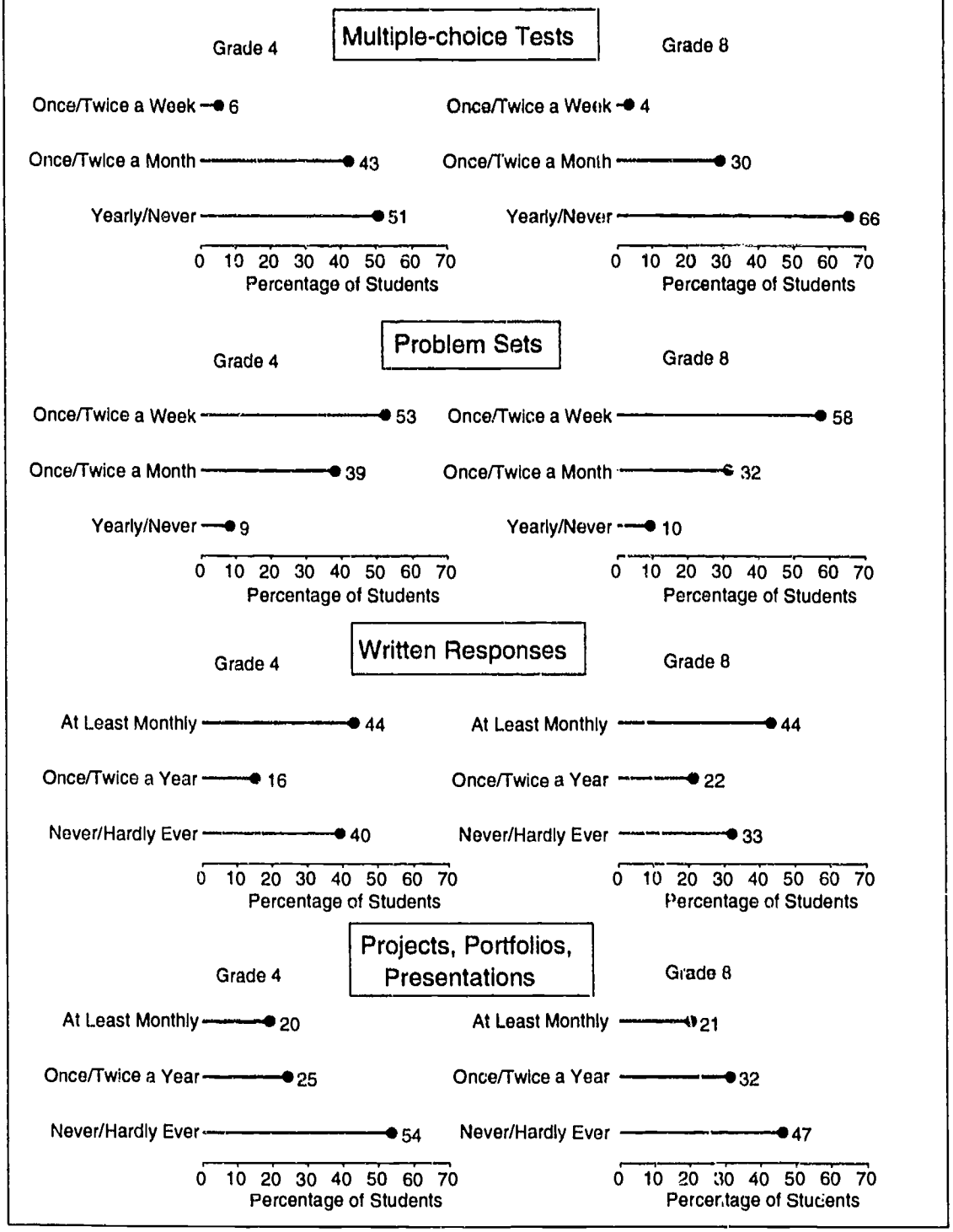
These teacher reports are of *all* the tests they give in the classroom, including tests they make themselves, while most previous pages in this report refer to statewide testing programs.

Source for Figure 9: Data are drawn from National Center for Education Statistics, *Data Compendium for the NAEP 1992 Mathematics Assessment of the Nation and the States*, Educational Testing Service, May 1993, p. 507. While these data are drawn from the Teacher Questionnaire, the data are student-based. That is, the percentage reported is the percentage of students whose teachers report...

See Appendix Table 1 for standard errors.

Many teachers are using alternative assessment within their classrooms once or twice a week; multiple-choice tests are more likely to be used once or twice a month or less.

Figure 9: Teacher Reports on the Frequency of Testing, 1992-93



Student Reports on the Frequency of Testing

The 1992 NAEP math assessment also asked students how often they were tested in math. The results are shown in Figure 10.

In the fourth grade, about 40 percent of the students were tested at least once a week. By the eighth grade this percentage increased to 60 percent.

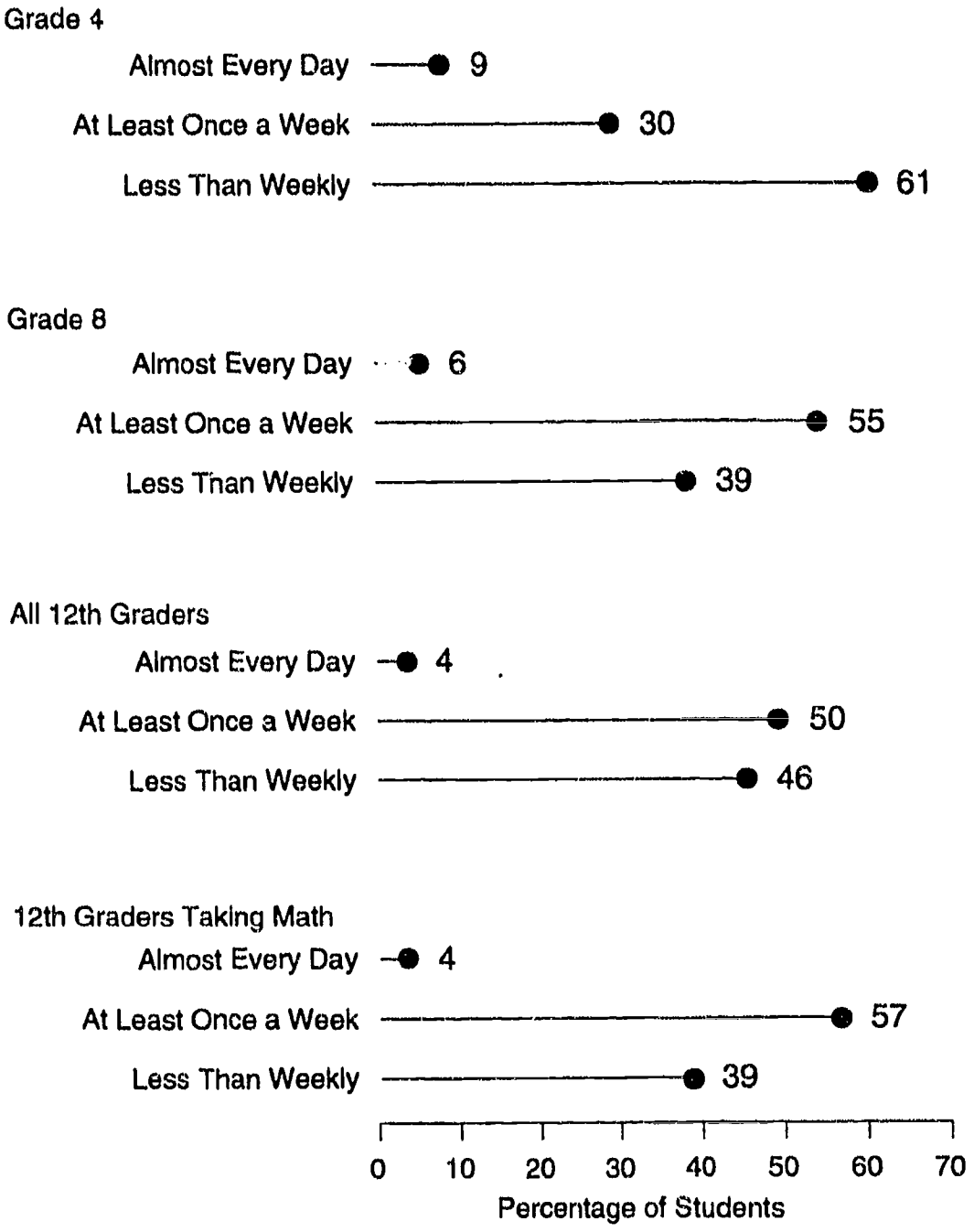
Among all twelfth graders, just over half are tested once a week or more, and among twelfth graders taking math, that figure was just over 60 percent.

Source for Figure 10: Data are drawn from National Center for Education Statistics, *Data Compendium for the NAEP 1992 Mathematics Assessment of the Nation and the States* Educational Testing Service, May 1993, p. 508.

See Appendix Table 2 for standard errors.

About half of 8th and 12th graders say they are tested in math at least weekly.

Figure 10: Students' Reports on How Often They Take Math Tests, 1992



Frequency of Multiple-Choice Testing

The 1992 NAEP math assessment also provides separate information for each of 44 participating jurisdictions. Figure 11 shows the percentage of students whose teachers reported that they used multiple-choice tests to assess student progress at least once a week.

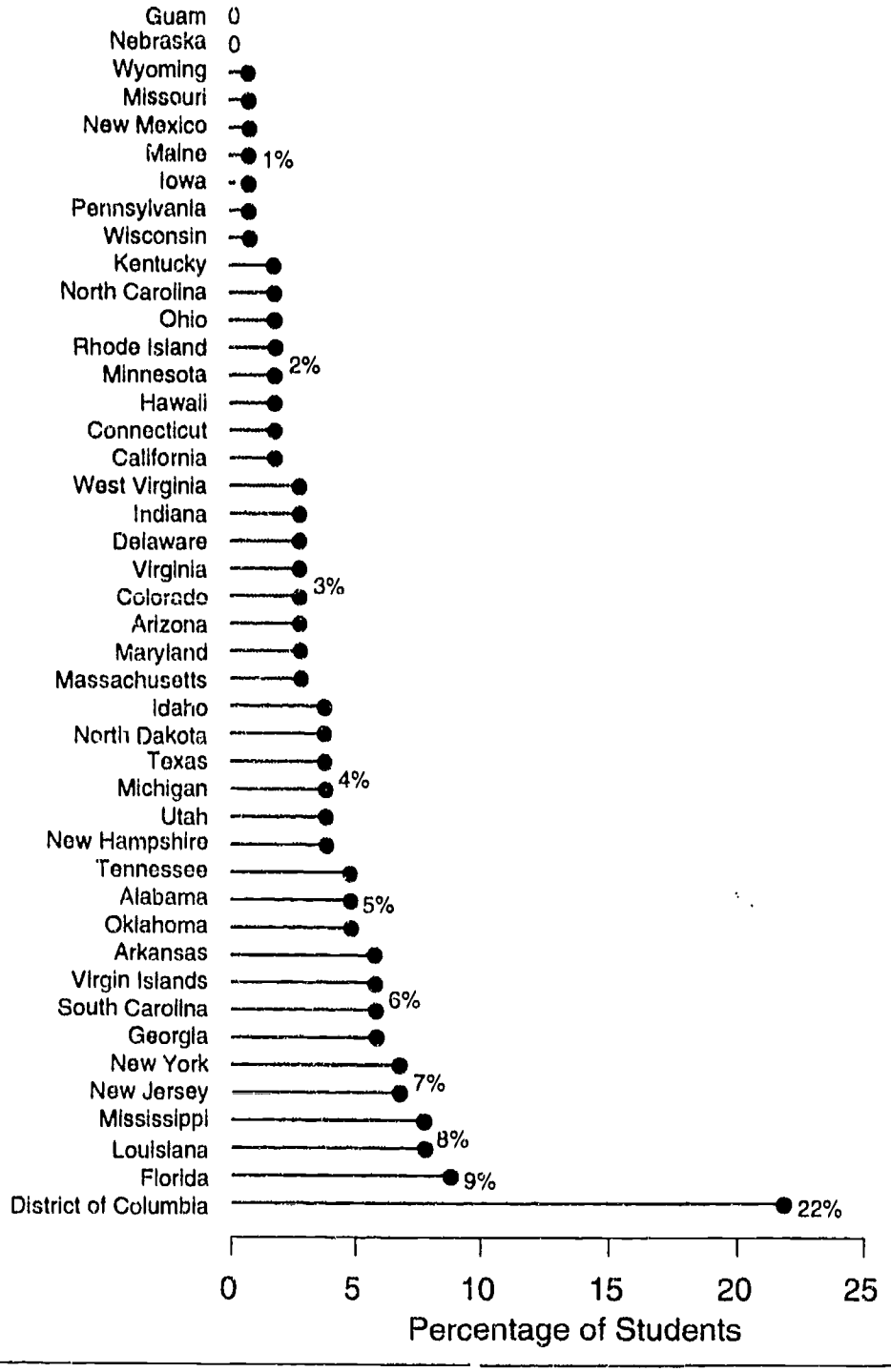
The percentages range from zero in Guam and Nebraska to 22 percent in the District of Columbia. The average for the nation was 4 percent.

Source for Figure 11: Data are drawn from National Center for Education Statistics, *Data Compendium for the NAEP 1992 Mathematics Assessment of the Nation and the States*, Educational Testing Service, May 1993, p. 510.

See Appendix Table 3 for standard errors.

Few teachers reported testing students with multiple-choice tests in math at least once a week.

Figure 11: Percentage of 8th Graders Whose Teachers Report They Use Multiple-choice Tests at Least Once a Week, by Jurisdiction, 1992



The Frequency of Teacher-Developed Tests

In the 1990 NAEP math assessment, teachers in participating jurisdictions were asked to indicate how often they gave teacher-developed tests to their math students. Data for these 40 jurisdictions are shown in Figure 12.

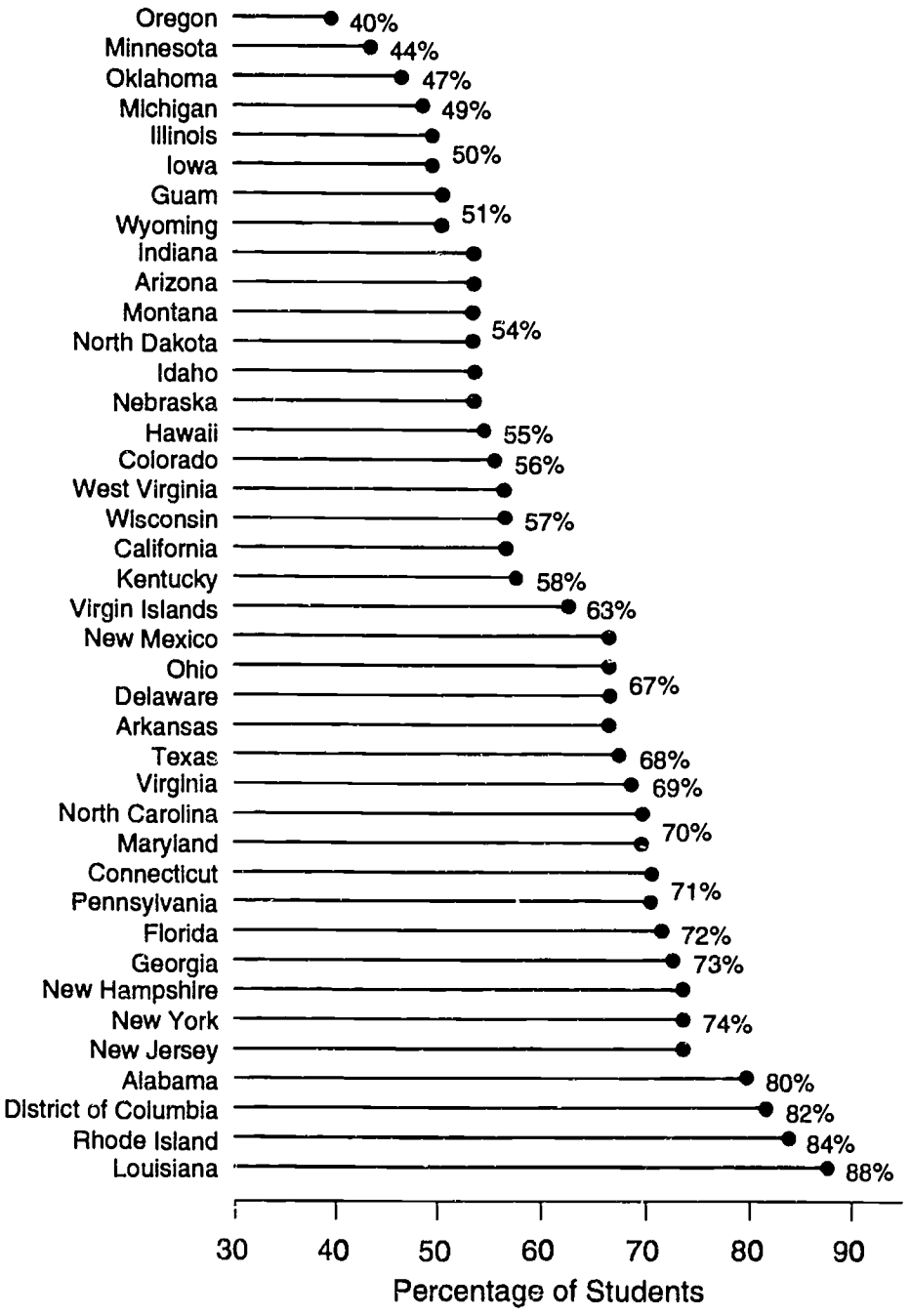
A little under half of the eighth grade students in Oregon, Minnesota, Oklahoma, and Michigan had teachers who gave their own tests at least once a week. At the other end of the scale, more than eight out of 10 students in Alabama, District of Columbia, Rhode Island, and Louisiana had teachers who used their own tests on a weekly, or more frequent, basis. The national average was 61 percent.

Source for Figure 12: National Center for Education Statistics, *The STATE of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States*, Educational Testing Service, June 1991, p. 321.

See Appendix Table 4 for standard errors.

Nationally, about 60 percent of teachers give their own math tests to students at least once a week. The percentage varies from 88 percent in Louisiana to 40 percent in Oregon.

Figure 12: Percentage of 8th Graders Whose Teachers Report Using Teacher-Generated Math Tests Once or More a Week, by Jurisdiction, 1990



Non-traditional State Assessment

As shown in Figure 13, as well as in several of the following sections, the states are busy developing and implementing nontraditional item types. Figure 13 provides an overall sense of this activity level. While the chart gives a dot to a state regardless of the development stage of the assessment (some states have funded an assessment type while others have already implemented the assessment), the direction of state activity is clear.

Extended-response open-ended items, short answer open-ended items, and enhanced multiple-choice* items were the most frequent alternatives cited. Individual and group performance assessments and portfolio or learning records are also being used by a considerable number of states. A few states are using interviews, observations, projects, exhibitions, or demonstrations in their state testing programs.

*The states responding to the survey made their own interpretations of what was meant by "enhanced multiple-choice." Generally, these are efforts to extend a multiple-choice item by asking for such things as an explanation of why a particular option was chosen, or having more than one correct answer.

Source for Figure 13: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

Note: A marker in a cell of Figure 13 means that the state is in some phase of developing the assessment type: funded, not started; begun development; completed development; piloted, being refined; or ready for use. States reporting that they "want to develop" the assessment type are not given a marker in the cell.

Many states are beginning to make considerable use of nontraditional items in their statewide testing programs.

Figure 13: The Use of Nontraditional Tests Items in State Assessment Programs, 1992-1993

	Enhanced Multiple-Choice	Short Answer Open-ended	Extended-Response Open-ended	Interview	Observation	Individual Performance Assessment	Group Performance Assessment	Portolio or Learning Record	Project, Exhibition, Demo	Other
TOTALS	14	18	22	3	4	14	7	9	5	4
Alaska			●							
Alabama			●							
Arkansas			●					●		
Arizona		●	●							
California	●	●	●	●		●	●	●	●	
Colorado			●			●		●	●	
Connecticut	●	●	●			●	●			●
Delaware						●			●	
Florida			●			●				
Georgia		●			●					
Hawaii		●		●	●					
Idaho						●				●
Illinois	●									
Indiana		●						●		
Kansas	●		●			●				
Kentucky		●	●				●	●		
Maine	●	●	●							
Maryland	●	●					●			
Massachusetts			●							
Michigan		●								
Minnesota		●				●	●		●	
Missouri	●					●				
Nevada	●									
New Jersey		●	●							●
New Mexico								●		
New York	●	●	●	●	●	●	●	●	●	
North Carolina	●	●	●							
Ohio	●		●							
Oregon			●							
Pennsylvania	●	●	●				●			
Rhode Island						●		●		●
South Carolina	●					●				
Tennessee						●				
Texas			●							
Utah		●	●		●					
Vermont			●					●		
West Virginia		●				●				
Wisconsin	●	●	●							

The Status of Alternative State Testing Programs

Figure 14 provides a closer look at the status of state testing programs that use alternative item types. Each type of assessment is shown for each development category — ready to use; piloted, being refined; begun or completed development; and funded, not started.

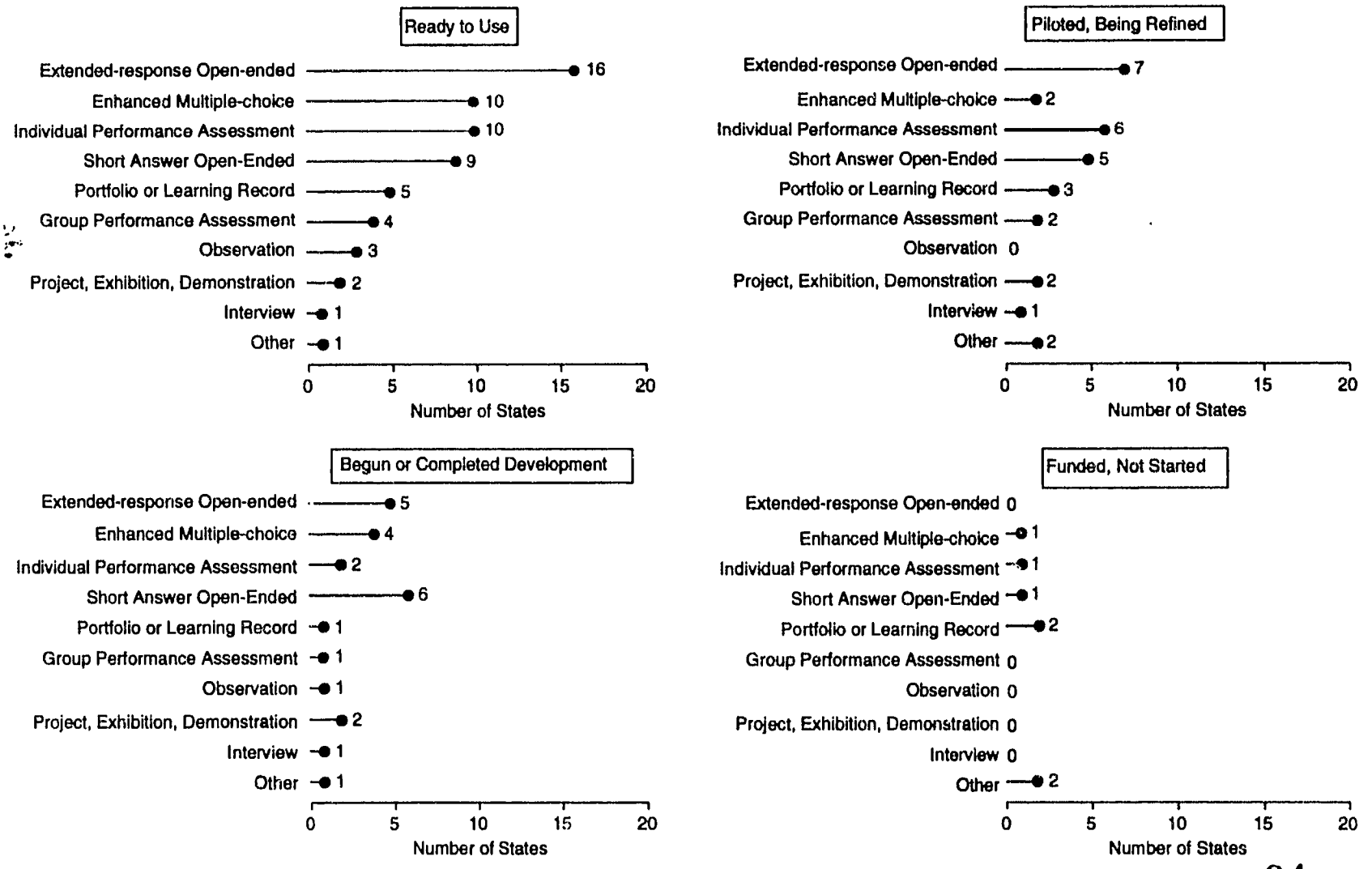
A considerable number of states are ready to implement (or are already using) extended-response open-ended items, enhanced multiple-choice items, individual performance assessments, and short answer open-ended items. A number of states are also moving ahead in developing and piloting a variety of alternative item types.

Source for Figure 14: Data are drawn from the "State Student Assessment Program Database, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

Note: In Figure 14, a state could be counted more than once if, for example, it had an enhanced multiple-choice test that was ready to use and another that was being piloted. Thus, the counts differ from Figure 13.

A considerable number of states are using, or are ready to use, alternative item types in their state testing programs.

Figure 14: Number of States Using Alternative Item Types, 1992-1993, by Status



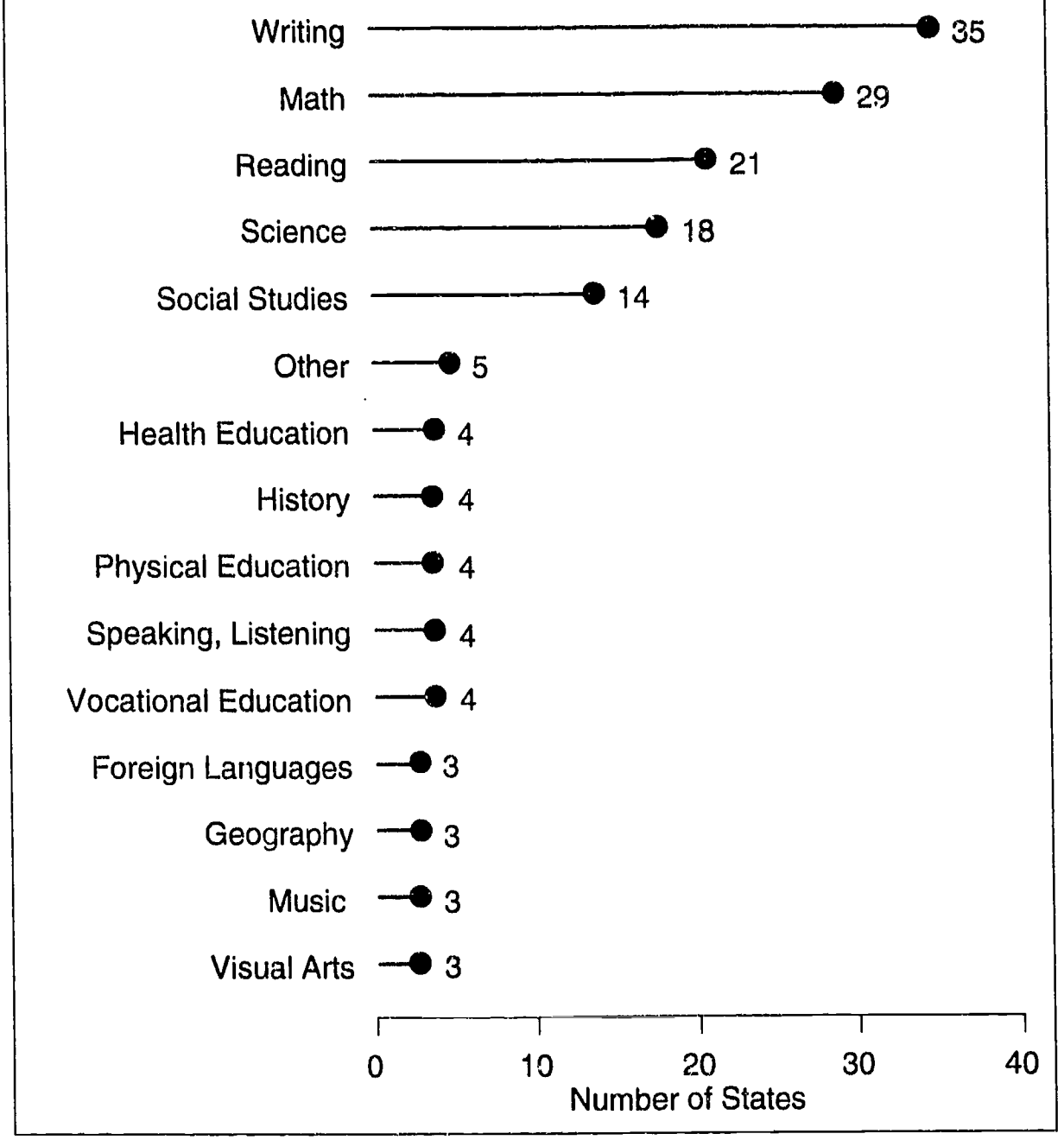
Subjects Assessed in Non-traditional State Testing Programs

As might be expected, state alternative assessments are given in the same subjects that are the province of traditional multiple-choice tests. Figure 15 shows the variety of subjects in which nontraditional item types are used. The subjects of writing, mathematics, reading, science, and social studies top the list.

Source for Figure 15: Data are drawn from the "State Student Assessment Program Data base, 1992-1993." Council of Chief State School Officers and North Central Regional Educational Laboratory.

States are using nontraditional items to assess student progress in the traditional subjects of writing, math, reading and science.

Figure 15: Number of States Developing or Using Nontraditional Test Items, by Subject, 1992-1993



Teacher Use of Different Types of Tests, by Race/ Ethnicity

The 1992 NAEP mathematics assessment provides data that can be used to determine whether different groups of students are exposed to different assessment types. One can see, for example, whether students from minority groups, students in lower ability groups, or students in classrooms with lower levels of resources available to their teachers are being given different types of tests.

Table 1 shows the extent of multiple-choice tests, problem sets, written responses, and projects or portfolios given to students grouped by race/ethnicity. The pattern is about the same among the racial/ethnic groups, after taking sampling error into account, and there are few differences that are significant.

Data in Table 1 are drawn from *NAEP 1990, 1992 - National Math Assessment-Data Almanac - Grade 8: Teacher Questionnaire, Weighted Percentages and Composite Proficiency Means.*

Table 1: Teachers' Reports on How Frequently They Assess Students with Different Assessment Types, Grade 8, by Race/Ethnicity, 1992

	White	Black	Hispanic	Asian	American Indian
Multiple-Choice					
Once/Twice a Week	4% (1.1)	6% (3.0)	6% (1.5)	8% (4.1)	1% (0.6)
Once/Twice a Month	28 (3.1)	40 (3.4)	32 (3.0)	28 (5.7)	22 (5.6)
Once/Twice a Year	29 (2.6)	24 (3.9)	27 (3.3)	28 (5.0)	31 (4.6)
Never	40 (2.5)	30 (3.4)	36 (2.6)	36 (5.2)	47 (6.7)
Problem Sets					
Once/Twice a Week	61 (2.7)	50 (3.2)	51 (3.4)	63 (5.5)	67 (9.6)
Once/Twice a Month	29 (2.8)	43 (3.7)	38 (3.6)	33 (6.0)	26 (8.1)
Once/Twice a Year	6 (1.6)	4 (1.3)	7 (1.6)	3 (1.2)	5 (2.4)
Never	5 (1.2)	3 (1.0)	4 (1.2)	2 (1.8)	3 (2.2)
Written Responses					
Once/Twice a Week	8 (1.6)	12 (2.9)	14 (2.2)	8 (2.2)	3 (2.0)
Once/Twice a Month	37 (3.1)	33 (4.3)	31 (2.7)	30 (5.6)	20 (6.2)
Once/Twice a Year	22 (2.5)	26 (3.8)	22 (3.3)	25 (5.3)	28 (6.6)
Never	34 (3.2)	29 (3.0)	34 (4.1)	38 (7.1)	49 (9.2)
Projects or Portfolios					
Once/Twice a Week	2 (0.7)	5 (1.9)	3 (0.8)	3 (1.0)	1 (0.5)
Once/Twice a Month	18 (2.1)	19 (3.2)	20 (2.3)	25 (4.5)	16 (6.1)
Once/Twice a Year	32 (2.0)	32 (2.7)	32 (2.8)	21 (3.0)	31 (7.4)
Never	48 (3.0)	44 (3.8)	45 (3.3)	51 (4.7)	52 (8.4)

Percentages are followed by standard errors (in parentheses).

Teacher Use of Different Types of Tests, by Ability Group

Table 2 shows how frequently teachers use different assessment types for each student ability group — low, middle, high, and mixed. Again, the pattern is very similar among the ability groups.

Data in Table 2 are drawn from *NAEP 1990, 1992 National Math Assessment - Data Almanac - Grade 8: Teacher Questionnaire, Weighted Percentages and Composite Proficiency Means.*

Table 2: Teachers' Reports on How Frequently They Assess Students with Different Assessment Types, Grade 8, by Ability Grouping, 1992

	Low	Middle	High	Mixed
Multiple-Choice				
Once/Twice a Week	2% (0.8)	5% (1.6)	3% (1.3)	5% (1.9)
Once/Twice a Month	30 (3.6)	35 (3.7)	22 (3.7)	29 (4.4)
Once/Twice a Year	28 (3.9)	25 (3.9)	34 (4.5)	27 (4.7)
Never	40 (3.9)	35 (3.1)	42 (3.1)	9 (5.6)
Problem Sets				
Once/Twice a Week	54 (6.4)	58 (3.6)	67 (3.5)	50 (5.8)
Once/Twice a Month	29 (3.8)	33 (3.9)	26 (3.3)	41 (5.3)
Once/Twice a Year	9 (3.5)	6 (1.7)	3 (1.4)	3 (1.6)
Never	9 (4.3)	3 (1.0)	4 (0.9)	6 (2.4)
Written Responses				
Once/Twice a Week	7 (1.9)	7 (1.5)	13 (3.5)	8 (2.1)
Once/Twice a Month	28 (4.6)	45 (3.2)	31 (3.6)	29 (4.0)
Once/Twice a Year	26 (4.5)	21 (2.6)	21 (2.5)	24 (4.7)
Never	39 (5.3)	27 (2.9)	35 (4.7)	39 (4.6)
Projects or Portfolios				
Once/Twice a Week	1 (0.4)	3 (1.6)	2 (0.9)	3 (1.2)
Once/Twice a Month	18 (3.0)	20 (3.0)	19 (2.8)	17 (3.8)
Once/Twice a Year	31 (4.9)	33 (3.6)	32 (4.7)	30 (3.9)
Never	0 (5.9)	45 (3.7)	47 (4.7)	51 (4.6)

Percentages are followed by standard errors (in parentheses).

Teacher Use of Different Assessment Types, by Resource Availability

Table 3 shows the types of assessments students are exposed to when grouped by their teachers' estimation of the sufficiency of classroom resources. The pattern does not vary much with differences in the availability of classroom resources.

Data in Table 3 are from the *1992 National Assessment of Educational Progress Mathematics Assessment* and were calculated by special computer runs conducted at Educational Testing Service.

Table 3: Teachers' Reports on How Frequently They Assess Students with Different Assessment Types, Grade 8, by Resource Availability, 1992

	Get All Needed Resources	Get Most of Needed Resources	Get Some or None of Needed Resources
Multiple-Choice			
Once/Twice a Week	10% (4.5)	2% (0.9)	5% (1.8)
Once/Twice a Month	30 (6.3)	35 (4.2)	23 (3.2)
Once/Twice a Year	21 (4.6)	27 (3.1)	32 (4.9)
Never	39 (6.3)	36 (3.4)	40 (4.0)
Problem Sets			
Once/Twice a Week	56 (7.1)	58 (2.9)	59 (4.6)
Once/Twice a Month	32 (7.6)	31 (2.8)	33 (4.0)
Once/Twice a Year	8 (5.2)	6 (1.6)	4 (1.3)
Never	5 (2.9)	5 (1.2)	4 (1.5)
Written Responses			
Once/Twice a Week	14 (3.7)	6 (1.5)	11 (2.5)
Once/Twice a Month	35 (5.6)	35 (3.6)	36 (3.8)
Once/Twice a Year	14 (4.2)	24 (3.0)	24 (2.9)
Never	37 (6.6)	35 (3.9)	30 (4.9)
Projects or Portfolios			
Once/Twice a Week	2 (2.1)	2 (0.6)	3 (1.3)
Once/Twice a Month	23 (4.8)	19 (2.2)	17 (3.8)
Once/Twice a Year	26 (5.9)	30 (3.6)	35 (4.0)
Never	48 (7.5)	50 (3.9)	44 (4.6)

Percentages are followed by standard errors (in parentheses).

In Conclusion

There is a lot of testing in American schools. A visitor who had not seen our schools for 20 years would likely see the volume of testing as the most noticeable change in American education. It appears that the future holds more testing, unless the alternative assessments under development start taking the place of traditional standardized tests.

Testing is very much alive in our schools, but a national investigation of testing has been under way now for a few years. The jury is still out on the condition of the testing system. The statistics in this report tell a lot about the quantity and type of testing, but nothing about its impact and its value. One feature of testing that needs much closer examination is the "test all students" approach for accountability in 41 state systems, with just one state sampling students (and three states with mixed systems). The National Assessment of Educational Progress has been operating on a sampling basis for over 20 years. Its sampling approach has been used at the state level and could be used at the district level, as well. This approach can serve accountability needs without intruding so much on instructional time; testing every student in a class can then be reserved for serving instructional purposes.

The progress in developing alternative item types in statewide programs — constructed responses, performance assessments, and portfolios — appears to be substantial. And when we look at the *total* of testing... including teacher constructed testing... that goes on in the classroom, alternative approaches appear to be doing well. There is, however, wide variation among the states in the frequency of multiple-choice testing. The individual teacher creates the tests most closely aligned with instruction, and this also varies considerably among the states, with 60 percent of 8th grade students getting such tests in mathematics at least once a week. Perhaps we need more attention given to helping teachers construct good assessments.

It remains to be seen whether these performance-type assessments will begin to substantially displace the traditional multiple-choice, machine-scored tests that have predominated. We do note that seven in 10 tests in state testing systems are still multiple-choice. Establishing a baseline, as we attempt to do in this report, will permit educators and policy makers to track the profile of testing in this period of ferment and change.

Notes and References

The State Student Assessment Program Database, 1992-1993 was produced by the Council of Chief State School Officers and the North Central Regional Educational Laboratory using results from a statewide survey of state testing programs. For information about database content, contact:

Edward Roeber, Director,
Student Assessment Programs
CCSSO
202-386-7045

Linda Bond
Director of Assessment
Regional Policy Information Center
NCREL
317-328-9704

Student Testing: Current Extent and Expenditures, with Cost Estimates for a National Examination was prepared for Congress by the U.S. General Accounting Office. First copies of GAO reports are free. Additional copies are $2 each. Order No. GAO/PEMD-93-8 from:

GAO
PO Box 6015
Gaithersburg, MD 20877
or call 202-275-6241

The NAEP 1992 Mathematics Report Card for the Nation and the States was prepared by Educational Testing Service under contract with the National Center for Education Statistics. For ordering information, write to:

Education Information Branch
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Ave., NW
Washington, DC 20208-5641

The Data Compendium for the NAEP 1992 Mathematics Assessment of the Nation and the States was prepared by Educational Testing Service under contract with the National Center for Education Statistics. For information, write to the address listed above.

The STATE of Mathematics Achievement: NAEP's 1990 Assessment of the Nation and the Trial Assessment of the States was prepared by Educational Testing Service under contract with the National Center for Education Statistics. For information, write to the address listed above.

Appendix Table 1: Teachers' Reports on the Frequency of Testing (Percentage of Students with Standard Error Reported in Parentheses)

	Grade 4	Grade 8
Multiple-Choice Tests		
Once or Twice a Week	6 (1.0)	4 (1.0)
Once or Twice a Month	43 (2.8)	30 (2.5)
Yearly or Never	51 (2.7)	66 (2.8)
Problem Sets		
Once or Twice a Week	53 (2.8)	58 (2.3)
Once or Twice a Month	39 (2.3)	32 (2.4)
Yearly or Never	9 (1.4)	10 (1.7)
Written Responses		
At Least Monthly	44 (2.6)	44 (2.7)
Once or Twice a Year	16 (1.5)	22 (2.0)
Never or Hardly Ever	40 (2.0)	33 (2.7)
Projects, Portfolios, or Presentations		
At Least Monthly	20 (1.7)	21 (2.0)
Once or Twice a Year	25 (1.8)	32 (2.5)
Never or Hardly Ever	54 (2.4)	47 (2.6)

Appendix Table 2: Students' Reports on How Often They Take Mathematics Tests, Grades 4, 8, and 12, NAEP, 1992 (Percentage of Students with Standard Error Reported in Parentheses)

	Grade 4	Grade 8	Grade 12 All Students	Grade 12 Taking Math
Almost Every Day	9 (0.6)	6 (0.3)	4 (0.3)	4 (0.4)
At Least Once a Week	30 (1.2)	55 (1.2)	50 (1.2)	57 (1.4)
Less Than Weekly	61 (1.5)	39 (1.3)	46 (1.2)	39 (1.5)

Appendix Table 3: Percentage of Students Whose Teachers Report They Use Multiple-Choice Tests at Least Once a Week, 1992 (Percentage of Students with Standard Error Reported in Parentheses)

Alabama	5 (1.7)
Arizona	3 (1.2)
California	2 (0.7)
Colorado	3 (1.0)
Connecticut	2 (0.7)
Delaware	3 (0.2)
District of Columbia	22 (0.8)
Florida	9 (1.9)
Georgia	6 (1.3)
Hawaii	2 (0.2)
Idaho	4 (1.3)
Indiana	3 (1.1)
Iowa	1 (1.1)
Kentucky	2 (0.4)
Louisiana	8 (2.4)
Maine	1 (1.1)
Maryland	3 (1.1)
Massachusetts	3 (1.2)
Michigan	4 (1.9)
Minnesota	2 (1.0)
Missouri	1 (0.2)
Nebraska	0 (0.2)
New Hampshire	4 (2.3)
New Jersey	7 (2.0)
New Mexico	1 (0.9)
New York	7 (2.2)
North Carolina	2 (0.8)
North Dakota	4 (2.0)
Ohio	2 (1.0)
Oklahoma	5 (1.8)
Pennsylvania	1 (0.9)
Rhode Island	2 (0.2)
South Carolina	6 (1.8)
Tennessee	5 (1.7)
Texas	4 (1.5)
Utah	4 (1.5)
Virginia	3 (1.0)
West Virginia	3 (1.1)
Wisconsin	1 (0.6)
Wyoming	1 (0.2)

Appendix Table 4: Teachers' Reports on How Often Students Take Teacher-Generated Mathematics Tests (Percentage of Students with Standard Error Reported in Parentheses)

	At Least Several Times a week	About Once a Week
Alabama	9 (2.8)	70 (3.1)
Arizona	7 (2.2)	47 (2.7)
Arkansas	6 (2.3)	61 (3.1)
California	5 (1.2)	52 (3.6)
Colorado	6 (1.5)	50 (3.8)
Connecticut	5 (1.7)	66 (2.9)
Delaware	2 (0.3)	65 (1.1)
District of Columbia	27 (1.0)	55 (1.1)
Florida	11 (1.8)	61 (3.1)
Georgia	7 (1.7)	66 (3.4)
Hawaii	4 (0.4)	51 (0.9)
Idaho	2 (0.4)	51 (2.2)
Illinois	4 (1.7)	46 (4.2)
Indiana	5 (1.5)	49 (4.0)
Iowa	6 (2.9)	44 (4.8)
Kentucky	6 (1.8)	52 (4.4)
Louisiana	11(2.4)	77 (3.6)
Maryland	4 (1.4)	65 (3.4)
Michigan	5 (1.6)	44 (3.7)
Minnesota	3 (1.1)	41 (3.5)
Montana	3 (0.8)	50 (2.9)
Nebraska	5 (1.2)	49 (3.5)
New Hampshire	5 (0.8)	69 (1.3)
New Jersey	11 (2.5)	64 (3.7)
New Mexico	4 (0.6)	64 (1.5)
New York	6 (2.1)	68 (3.3)
North Carolina	10 (1.6)	60 (3.2)
North Dakota	10 (0.6)	46 (3.0)
Ohio	4 (1.3)	63 (3.8)
Oklahoma	2 (0.8)	45 (3.6)
Oregon	3 (1.1)	37 (3.3)
Pennsylvania	5 (1.4)	66 (3.3)
Rhode Island	12 (1.0)	72 (1.6)
Texas	3 (1.1)	65 (3.5)
Virginia	12 (2.2)	57 (3.4)
West Virginia	8 (2.1)	49 (4.6)
Wisconsin	5 (1.8)	52 (3.7)
Wyoming	4 (0.7)	46 (1.2)
Guam	10 (0.6)	41 (0.7)
Virgin Islands	5 (0.4)	58 (0.9)

04202-06057 • S123M6 • 204882