DOCUMENT RESUME

ED 319 763 TM 014 947

AUTHOR Gong, Brian; And Others
TITLE Current State Science Assessments: Is Something Better Than Nothing?
PUB DATE 23 Apr 90
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Assessment; *Educational Change; Educational Improvement; Elementary Secondary Education; Middle Schools; Science Education; *Science Tests; Standardized Tests; State Departments of Education; *State Programs; State Standards; State Surveys; Surveys; *Testing Programs; Test Use
IDENTIFIERS *American Association for Advancement of Science

ABSTRACT

Current science assessments in use in the states were studied, and the extent to which these assessments meet the goals of reform of science education specified by the American Association for the Advancement of Science (AAAS) was investigated. These goals were identified by the AAAS in "Project 2061: Science for All Americans?" Data were derived from surveys conducted by the Education Commission of the States (1987) and the U.S. Office of Technology Assessment (1987). Information was further gathered through telephone calls and a review of the instruments and supporting documents from the states. Commercial tests in use, particularly in middle schools, were examined. It was determined that state science assessments are not a driving force in directing the curriculum or instructional practice toward the goals of science education reform. State assessments varied so widely that their effects on a national science direction were questionable. The content of the assessments was generally not consistent with the AAAS reform goals, and the process aspects of the standardized tests were weak to non-existen An appendix lists the states using standardized tests for science assessment and summarizes the information in a chart. (SLD)

Current State Science Assessments: Is something better than nothing?[1]

Brian Gong, Colleen Lahart, and Rosalea Courtney
Educational Testing Service

23 April, 1990

Background

Many parts interlinked in the educational system are current targets for study and improvement: national and state policies (regarding such issues as funding, standards, and organization of educational institutions), curriculum (guidelines, organization, development) teachers (including training, credentialing, and retention), students (contextualized learning, differential impacts), parents (motivation, out-of-school learning). In this paper we focus on a part that often cuts across and often is attributed as having great leverage on many of the other parts in this complex system: assessment.

We use examination of one subject matter area--science--and of assessment programs at the state level as a lens to examine how assessment and curricular goals do or do not work together to support educational improvement. This paper reports a conceptual analysis of the matches and mismatches between current assessment instruments and desired educational goals; the actual implementation of assessment policies, instruments, and practices is not addressed in this study.

Science Reform and Assessment

Improvement in science education has been linked to the needs for national economic competitiveness, individual functionality in an increasingly technological society, and avoidance of a permanent underclass (e.g., NSB, 1983). Most proposals for science reform have included calls for improved assessment (e.g., AAAS, 1989, p. 166; CCSSO, 1989; OERI, 1988). However, such calls for educational improvement reflect several different models of the role assessments should play in the educational system: as accountability mechanisms to protect the public interest, as evaluation tools to inform administrative decisions about programs and personnel, as systems directly involved in informing classroom instruction and modelling learning outcomes, or perhaps some mixture of these motivations (cf. ETS, 1990; Ewell, 1987; Nickerson, 1989). Accountability mechanisms include high school basic skills exams required for graduation; program evaluation tools support comparison of district, building, teacher, or class performances to other units or over time; instructional monitoring systems are intended to provide diagnostic and prescriptive information at the classroom level.

It is interesting and important to note that regardless of the role they give assessment, most calls for reformed science assessment agree on two

[1]Presented at the American Educational Research Association annual convention, Boston, MA, April 16-20, 1990.

features of desirable assessment: the assessment should exemplify desired performance, and the assessment should report results in ways that inform change towards the desired goals.

The first requirement, that assessment should exemplify desired performance, combines the notions of validity and science reform: the assessments should validly reflect what they are supposed to assess, and what they are supposed to assess should be "good science education" (which is often different from what currently exists, according to the reform-minded reports). The fear is that assessments may reliably measure certain knowledge, skills, or aptitudes, and yet those entities may not be desirable goals or competencies. The second requirement, that the assessment should report results in ways that inform change towards the desired goals, acknowledges that assessments should be intended to inform decision-making and action. As action- or decision-oriented reports, part of assessments' validity stems from their design, implementation, and use for those practical goals, not only from psychometric analysis or content experts' reviews (cf. Messick, 1989).

Purpose

Our overarching purpose is to construct science assessment models and examples that contribute to improved science education. Although our main project, sponsored by ETS, focuses on instructional assessment--or assessment done by teachers and students in classroom settings to improve science learning and teaching--we felt it was important to understand the assessment context in which teachers are working, and also the national trends in science education and assessment. Hence, the broad objective of this study was to survey current state science assessments and determine to what extent the assessments support standards of student performance consistent with the goals of science education reform. Our specific question in this study focuses on the relation between test specifications and the requirements of science reform: How consistent are the currently used state science assessments to the goals of science reforms such as the landmark report of the American Association for the Advancement of Science's Project 2061: Science for All Americans?

General Specifications for Assessments from Science Reform

Recent reforms in American science education include sponsorship of significant science curriculum development projects by the National Science Foundation at the elementary and middle school grade levels (NSF, 1989), and issuance of a major report, Project 2061, by the AAAS (AAAS, 1989) seeking to define the knowledge, skills, and attitudes all students should acquire in science, mathematics and technology. The NSF-sponsored projects and the AAAS report share an emphasis on development of "scientific literacy" (where scientific literacy is defined as the science and technology to be learned and applied by citizens, as contrasted with science knowledge and skills learned by students intending to major in science in college). One critical criterion is to show relationships between scientific principles and ways of knowing, technological applications, and personal and social issues. A second common emphasis is on depth of experience and mastery of critical knowledge, rather than on breadth of coverage. The AAAS succinctly stated that the goals are to:

> "Identify only a small core of essential knowledge and skills. Do not call on the schools to cover more and more material, but instead recommend a set of learning goals that will allow them to concentrate on teaching less and on doing it better. Focus on scientific significance. Identify only those concepts and skills that are of surpassing scientific importance." (AAAS, 1989, pp. viii-ix).

Three specific points of science education reform, then, against which the state assessments may be analyzed, are to what extent the assessments do:

1. focus on central concepts, skills, and attitudes rather than on uncontextualized, fragmented, or lower-level knowledge skills;
2. focus on or promote development of depth of those central knowledge, skills, and attitudes rather than breadth at the expense of depth;
3. connect science to technical, personal, philosophical, and social applications and phenomena.

Methods and Data Source

We were fortunate to be able to draw upon the comprehensive surveys of science and math assessments conducted by the Education Commission of the States (ECS, 1987) and the U.S. Office of Technology Assessment (OTA, 1987). Both the ECS and OTA reports were based on 1985 data. Much of our work reported here involved updating the ECS data, and examining the assessments from a different angle.

All the states were contacted by phone in April 1990; most had been previously contacted in July 1989. In addition to verbal reports of the state's science assessment, other information about the states' science assessment instruments was gathered, including general descriptions written for public release, sample items, summary results, and actual assessment instruments, if available.[2] Even in less than a year's time there were changes in individual states' science assessment plans and instruments, and policies. Clearly this is a dynamic area, and the results of this survey will be accurate only for a limited time.

Copies of the commercial tests used by the states in their science assessments were examined. Where several forms exist, the form closest to grade level 6-7 was used. The information about the commercial tests' content-process specifications was obtained from the information sections included with each test. Within each assessment, specific items were subjected to content and cognitive analysis. Commercial tests are also modified and revised, and it may be that the specific information about the commercial tests in this report will also soon need to updated.

Although states may test across several grade levels, in this study we focused our analysis (but not the survey) on the middle school science grades 6-7. Pre-high school science, we guessed, would be more likely reflect the science literacy goals than the more discipline-oriented high school grades.

---

[2] State-constructed assessment materials were received from 17 states, including some states that did not have a mandatory state science assessment.

In addition, middle school science has been identified as a key filter, or choice point where students decide long-lasting attitudes about science (e.g., Mullis & Jenkins, 1988) and about future course selection. In addition, other filters are established in middle school, notably academic mathematics preparation (e.g., Beane, 1985). Middle school is an important time for assessments to provide information and encouragement.

Results

The results of our survey and analysis indicate that:

1. state science assessments clearly are not a driving force in directing the curriculum or instructional practices towards the standards of science education reform in the United States.
2. In fact, state science assessments vary so widely that it may be questioned whether they have a major effect at all on a coherent national science direction.
3. In addition, almost all state assessments' content structures do not appear to be consistent with current science education reform goals.
4. The process aspects of standardized tests are weak to non-existent.
5. However, in light of the inconsistency of almost all the state science assessments to the science reform goals examined in this study, it may be a positive thing that state assessments are not powerful determinants of curriculum and instruction.

These conclusions are reported in greater detail below.

Influence of State Science Assessments From a National Perspective

The data show that the majority of states have little direct leverage on science education, either through requirement of a state-wide means for comparison or through accountability measures tied to science performance. Of the 50 states, almost 50% (24 states out of 50) do not have a science assessment at the state level. (See Table 1; full data shown in Appendix A.) Of those states that do have a state science assessment, very few have stronger accountability sanctions than reporting scores publicly. Our conclusion is that state science assessments, from a national perspective, are not a driving force in directing curriculum or instructional practices towards the standards of science education reform. Indeed, for almost half the states there is no state-level assessment to "drive" science towards any standard.

The five states that do have strong state-controlled accountability use mechanisms including tying performance on state assessments to funding sanctions, school accreditation, or student graduation. "Public accountability" reports are thus mixed with reporting results to the school or teacher for curriculum improvement, monitoring student progress, placement, and diagnosis. A mixture of report forms are generated, including reporting individual and group scores. The scores may be in the form of raw scores, criterion-referenced scores, or various transformations of norm-referenced scores. The scores may be sent to schools and teachers to be used in diagnosis, placement, and instructional planning. Group scores are reported at the state, district, and school levels, varying from state to state. Group scores are supposed to be used for state policy, curriculum revision, and development. The important point here is that none of the currently available reports contain information that is useful for assessing performance in relation to the science reform goals (e.g., centrality, depth, or connections), no matter the accountability, program evaluation, or instructional diagnosis purposes, since performance scores are aggregated without regard to the science reform dimensions.

Nature of the State Science Assessment Instruments

On the basis of our analysis of the state science assessment programs' instruments and specifications, we concluded that state science assessments have varied sources, mandates, purposes, and varied curriculum specifications, but they almost all have similar design specifications. Those design specifications include: paper and pencil, multiple-choice format; no more than 60 items; one hour administration time; administered twice a year at most; requires that students work individually; survey Life, Physical, and Earth/Sea/Space sciences and several science process skills; and the reports compress and aggregate results into unidimensional scales and/or single composite scores. In other words, short, highly sampled, multiple-choice tests dominate.[3]

Of the 26 states that reported having science assessments, 10 use only commercial standardized tests, 13 use state-constructed tests, and three states use both commercial and state-constructed tests.

Table 1. Numbers of states with each source of science assessments.

Source of Test	No. of States
No state science assessment	24 (48%)
Commercial test only	10 (20%)
State-made test only	13 (26%)
Commercial and state tests	3 (6%)
Total	50

Table 2 shows the number of states using each of the commercial tests for assessment.

Table 2. Number of states using each of six commercial tests for state science assessment, grades 4-8.

SAT	CTBS	ITBS	CAT	MAT6	TAP	TOTAL

[3]California, New York, and Connecticut are exceptions in that their state science assessments include or specify at least some performance-based or other significantly open assessment component.

No. of states	4	4	4	2	1	1	16*

*Three states use multiple commercial tests

In the commercial standardized tests examined, the Science section varies in length from 25-60 questions, with an average of approximately 40 science items. These items are spread across the traditional disciplines of Life Science, Physical Science, and Earth and Space Science; some tests further subdivide the content categories. Four of the six tests also cross-reference each item to a process skill. The other two tests focus on Content Knowledge in its science sections.

There are two different approaches to testing the Process Knowledge in these tests. Three of the tests use variants of the cognitive skills outlined in Bloom's Taxonomy of Educational Objectives (Bloom et al., 1954): knowledge or recall, comprehension, application and analysis, and synthesis and evaluation. The other organization of process skills used the AAAS scientific task analysis: classify, hypothesize, measure, and infer. The state assessments were very similar in their content-process specifications, especially those assessments that were multiple-choice format and administered annually. The commercial standardized tests are summarized in Table 3.

Table 3. Summary of commercial standardized tests used in state science assessments.

TEST	GRADE LEVELS NO./RANGE	NO. ITEMS	CONTENT FRAMEWORK	PROCESS FRAMEWORK
SAT	4/3.5-9.9	44-60	Physical science Biological sci.	Analysis, infer, predict, classify, experiment, measure, hypothesis
CTBS	7/1.6-12.9	25-40	Botany Zoology; Ecology Physics, Chemistry, Land/Sea/Space	Recall, Explicit information skills, Inferential reasoning, Evaluation
MAT6	6/1.5-12.9	31-50	Physical sci.; Earth and space Life science	Knowledge, Comprehension; Inquiry skills; Critical analysis
TAP	4/9-12	54	Nature of sci. Life science Earth/space Chemistry/physics	Knowledge/information Comprehension Application/analysis Synthesis/evaluation Experimental methods/techniques
CAT	9/1.6-12.9	25-40	Botany; Zoology; Ecology; Physics; Chemistry;	

			Land/Sea/Space
ITBS	5/4-8	40-45	Life science; Earth and space science; Physics; Chemistry; Health and safety; Nature of science (Methods of inquiry; Nature of evidence; Nature of proof; Cause and effect; stability and change)

Stanford Achievement Test (SAT), Comprehensive Test of Basic Skills (CTBS), Metropolitan Achievement Test (MAT6), Tests of Achievement and Proficiency (TAP), California Achievement Tests (CAT), Iowa Test of Basic skills (ITBS).

The CTBS exemplifies a "typical" standardized test. Table 4 shows the Content and Process specifications for the CTBS level H (1983, grades 6.6-8.9).

Table 4. Numbers of items of the Science section of the California Test of Basic Skills (level H, grade level 6.6-8.9) identified with each Content and Process area.

Content	No. of items
Botany	8
Zoology	10
Ecology	5
Physics	5
Chemistry	6
Land/Sea/Space	6
Total	40

Process	No. of items
Recall	4
Explicit Info. skills	21
Inferential reasoning	5
Evaluation	10

The examination of assessment specifications and content analysis of sample items shows that both commercial tests and state-constructed assessments fail to meet the science education reform goals exemplified by the AAAS.

The content structures are inconsistent with current science education reform goals. There is a clear mismatch between the breadth of content coverage specified by the test and what most middle school science classes actually cover, let alone what is advocated by Project 2061 and other reform documents. In addition, the items appear to represent a sample of topics within each discipline. Such treatment certainly is not conducive to encouraging learning an area in depth. Sparse sampling neither addresses what is or should be taught, nor provides incentives for addressing topics in depth. On the CTBS test shown in Table 4 for example, there are five items that deal with ecology. However, the California state curriculum guidelines identify four major concepts for ecology; middle school textbooks commonly divide treatment of ecology into three main chapters and over 20 major headings (Gong et al., 1990). Most of the assessment items examined, however, deal with facts, not principles, and appear designed to tap lower-level thinking skills. For example, none of the CTBS items appeared directly to require knowledge of central concepts such as California curriculum guidelines and Project 2061 advocate. None of the tests had clusters of questions to probe understanding of principles in multiple contexts, applications, or levels of difficulty. Thus, there was no way to assess whether a student answered a question by reasoning from high-level principles, by memorizing a specific fact, or by guessing. (This larger issues of whether the assessments validly tap process skills is addressed in greater detail by Gong, 1990.)

Even on paper, the Process aspects of state science assessments are weak to non-existent. It is interesting to note that the Process categorizations for the CTBS example in Table 4 did not reflect any developmental sequence across forms. That is, there are not more items that supposedly tap "higher order thinking skills" (e.g., Evaluation) in the test forms for the higher grade levels than there are in the tests for younger students. In fact, the CTBS form for grades 6-8 shown in Table 4 has twice as many "Evaluate" items as the form for grades 11-12. In any case, almost every test will be of little use to a teacher or science educator because the reports generally report a highly aggregated score for "content." And even though they have Process specifications, none of the commercial standardized tests report skill or process performance. Thus, administrators, teachers, students, and parents do not even get a report that attempts to reflect performance on higher order thinking.

None of the items examined appear to tap knowledge of technology or current social issues. All the items examined deal with content as it might be presented in a science textbook, or in common personal experience (e.g., growing a plant). On the point of dealing with science in ways and forms appropriate for developing a scientifically literate citizenry, then, these assessments also fall short of the desired mark.

<u>Summary</u>

In summary, state science assessments have the potential for being a driving force in barely half of the states currently. For the states that do have a state science reform, almost none appear to have a current state science assessment program that will help direct curriculum or instructional practices towards the standards of science education reform. The majority of state science assessments' content structures appear inconsistent with current science education reform goals, and the process aspects are weak to non-existent.

In light of the inconsistency of almost all state science assessments to the science reform goals examined in this study, it may be a positive thing that state science assessments are currently not powerful determinants of curriculum and instruction nationally. We may wonder whether the "something" of science assessments in place in half the states is better than the "nothing" at all favored by the other half of the states. In any case, we may be glad that the weakness of most state science assessments provides opportunities to reform the assessments as well.

<u>Discussion and Recommendations</u>

When Is "Something" Better Than "Nothing"?

The low number of states with state assessments in science is surprising and disturbing. It reflects the low value placed on science education. Not counting the seven states that have no state assessments of any kind, it shows that 60% (26 of 43) of the states that have state assessments--usually in math and reading have put science assessment on a back burner. Thus, having a state assessment in science may be a good sign, one that says the state legislature or other have agreed that science is important. In such a climate, whether or not the assessments are valid, they may have good side-effects. As one district science curriculum assistant noted, "I would welcome a state test in science, because it would buy science time in the crowded day. As it is now, we are constantly pushed aside in favor of math and language arts--subjects that are being tested." In the upper elementary grades, science is often allocated only 20 minutes a day for instruction; some districts require as little as 12 minutes a day, or one hour a week. Buying time is a large issue. Thus, assessments may be poor evaluation instruments, but still play beneficial roles in the educational system.

Of course, it would be better if assessment programs' direct effects were also beneficial. Carefully conceived and implemented state-mandated assessments can play a significant role in promoting good science education (Armstrong et al., 1988). To construct such assessments we must have more detailed models that go beyond empirical and conceptual links between performance and objectives as contained in state and local assessment; we must identify how much, why, and how tests are related to curriculum, instruction, and cognition. The links between assessment and science education have not yet been established in detail. In particular, testings' influence on learning is not well-understood, although the research literature (and common knowledge claims) are full of impassioned yet often contradictory claims, such as tests drive the curriculum, tests reflect the textbooks; the format of tests channelize students' attention and lead to piecemeal learning, tests promote coaching; tests contribute to bias, tests help eliminate local disparities. To sort out the complicated picture of the connections between assessment and learning, teaching, and curriculum will require additional studies that address the political, social, and psychological decisions and organizations related to what happens in classrooms, students' heads, state education offices, and textbooks' publishing houses. Even more important than documenting what assessments currently are, we must construct models--conceptual and actual prototypes--for making assessments positive forces in educational reform. The models should be research-based, decision/action-oriented, have strong ties to curriculum and instruction, and be more responsive to teacher/student needs.

Good assessments should be based on well-articulated educational goals. This requires that the ecological relations of assessment and other parts of the educational system should be clearly articulated and coordinated. The assessment, curriculum guidelines, teacher education, local staff development, and funding infrastructures should be consistent and integrated. An evolutionary plan, championed by a committed "evangelist" with a power base, may be most appropriate for most states: periodic, systematic improvement of a science assessment program can go beyond the current assessments in a series of steps. (See reports of California's or Connecticut's experiences, e.g., Shavelson et al., 1990; Baron, 1990.)

Research and Development Work

Tests must be "tuned," or constructed for particular purposes. Those who mandate, design, and use tests should be clear on the purpose of the test: is it to hold student performances accountable, for program evaluation, program improvement, student diagnosis, or something else such as enforcing usage of a syllabus.

We need additional information about how test information is used so we can design tests and administration procedures that provide the necessary information. For example, three distinct uses that have distinct information design requirements are: comparison of individuals or groups to others for the purpose of assessing outcome or performance; longitudinal comparisons for assessing progress or change within a group or individual; and assessment of performance to inform instruction or curriculum design. Another way of addressing this concern is to note that the criteria and sources of criteria for interpreting the assessment must be clear: are they norm-referenced (e.g., ranking, percentile), criterion-referenced, self-referenced (e.g., progress over time), contextualized (e.g., compared against other districts or students with similar backgrounds); and the standards vary from being minimal competency to a core requirement to a moderately high target for all.

The majority of current science assessments must be examined and revised before using them to show standards of desired student performance. The reporting structures especially are weak. They are over-simplified for policy-makers, and do not have the right information for those at local levels who need to make decisions about curriculum and instruction. Too often the state assessments treat education like a game of "Blind Man's Bluff," with score reports calling out "Warmer!" or "Colder!" once a year--hardly the best information for informing the educational decisions that teachers and superintendents have to make sometimes daily.

It is hoped that studies such as this one will help inform policy makers, test developers, educators, and parents so they can guide the development of more appropriate assessments that will support a population that is truly more scientifically literate.

References

AAAS. (1989). Science for all Americans (AAAS Project 2061 overview report). Washington, DC: American Association for the Advancement of Science.

Armstrong, J., Davis, A., Odden, A., & Gallagher, J. (1989). Designing state curriculum frameworks and assessment programs to improve instruction. Presentation at the 1989 AERA Annual Meeting, San Francisco, CA.

Baron, J. B. (1990). What is the Connecticut Multistate Performance Assessment Collaborative Team (COMSPACT) in math and science and what are its goals? Presentation at AERA, Boston, MA.

Beane, D. B. (1985. Mathematics and science: Critical filters for the future of minority students. Washington, D.C.: The Mid-Atlantic Center for Race Equity, The American University.

Bloom, B. S. (Ed.). (1954). Taxonomy of educational objectives. Book 1: Cognitive domain. NY: Longman.

CCSSO. (1989). State-by-state indicators of science and mathematics education. Preliminary report. Washington, D.C.: Council of Chief State School Officers.

ECS. (1987). State initiatives to improve science and mathematics education. Denver, CO: Education Commission of the States (ECS).

ETS. (1990). The uses of standardized tests in American education. Proceedings of the 1989 ETS Invitational Conference. Princeton, NJ: Educational Testing Service.

Ewell, P. T. (1987). Assessment, accountability, and improvement: Managing the contradiction. Washington, D.C.: American Association for Higher Education Assessment Forum.

Gong, B. (1990). An analysis of science test items' cognitive demands: When is Bloom's comprehension equivalent to the scientific method? Presentation at the annual AERA national convention, April 16-20, 1990, Boston, MA.

Gong, B., Courtney, R., and Richardson, G. (1990). A survey of current science curriculum guidelines and materials. (unpublished manuscript).

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement, 3rd ed. NY: American Council on Education.

Mullis, I. V. S. & Jenkins, L. B. The Science report card: Elements of risk and recovery. Trends and achievement based on the 1986 National Assessment. (NAEP report 17-S-01). Princeton, NJ: Educational Testing Service.

National Science Board (NSB) Commission on Precollege Education in

Mathematics, Science and Technology. (1983). Educating Americans for the 21st Century. CPCE-NSF-03. Washington, DC: National Science Foundation.

Nickerson, R. S. (1990). New directions in educational assessment, Educational Researcher, 18(9), 3-7.

NSF. (1989). Guide to Programs: Fiscal year 1989. Washington, DC: National Science Foundation.

OERI. (1988). Creating responsible and responsive accountability systems: Report of the OERI State Accountability Study Group. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, Program for the Improvement of Practice.

OTA. (1987). State educational testing practices: Background paper. Washington, DC: U.S. Office of Technology Assessment.

Shavelson, R., Pine, J., Goldman, S., Baxter G., & Hine, M. S. (1990). Alternative technologies for assessing science achievement. Paper presented at AERA, Boston, MA.

STATES USING COMMERCIAL STANDARDIZED TESTS FOR STATE SCIENCE ASSESSMENTS

	Test Name	Admin.	Grades
Alabama	SAT	Annual	(4,8,)
Arizona**	ITBS SAT	Annual Annual	(1-8) (9-12)
Arkansas	MAT6	--	(4,6,7,8,10)
Georgia**	ITBS	--	(2,4,7,9)
Idaho	ITBS TAP	Annual	(6,8,11)
Louisiana*	CAT		
New Hampshire	CAT	Annual	(4,8,10)
New Mexico	CTBS	Annual	(3,5,8,10)
S. Carolina*	CTBS	Annual	(4,5,7,9,11)
S. Dakota	SAT	Annual	(4,8,11)
Tennessee	SAT CTBS4	Annual	(2,5,7,9,12)
Virginia*	ITBS	Annual	(4,8,11)
West Virginia	CTBS	Annual	(3,6,9,11)

13 TOTAL

* 3 states also use state-constructed tests.
** State-constructed test planned.

STATES USING STATE-CONSTRUCTED TESTS FOR STATE SCIENCE ASSESSMENTS

	Test Name	Admin.	Grades
California	CAP	Annual	(8, in '90 6,12)
Colorado	no name	1987 (trial only)	(3,6,9,11)
Florida	no name	88/89 (trial only)	(3,4,5)
Indiana	1 step no name	88/89	(3,6,8,9,11)
Louisiana*	LEAP	Annual	(11)
Maine	MEA (matrix items in science)	Annual	(4,8,11)
Massachusetts	MEAP	88/90	(4,8,12)
Michigan**	MEAP	1986&88 1989 1992 New Draft	(4,8,10) (5,8,11)
Minnesota**	MEAP	1987 (voluntary) 1993 (mandatory)	(4,8,11) (6,9,11)
Missouri	MMAT	4 yr cycle	(3,6,8,10)
North Carolina	NCAT	Annual	(K-3,4-6,7-8)
Oklahoma	no name	Annual	(3,7,10)
New York	Science Evaluation Manipulative Skills; Objective section	1989	(K-4)
Pennsylvania	EQA TELLS	Annual	(4,6,7,9,11)
S. Carolina*	Basic Skills	1988-89	(3,6,8)
Virginia*	SEPAM		

16 TOTAL

Delaware and Minnesota also have Item Banks available which include Science.
* 3 states also use commercial tests.
** New Science Assessment in Development

NO STATE ASSESSMENT IN SCIENCE

Connecticut	planned for 1993 (state-constructed, performance assessment)
Delaware	(has item Bank in science available)
Hawaii	
Illinois	planned for 1992 (state-constructed, grades 3,6,8,11)
Kansas	
Kentucky	
Maryland	planned for early 1990's (state-constructed)
Mississippi	
Nevada	
New Jersey	
Oregon	planned for 1993 (state-constructed, to be given every 5 years for program evaluation)
Rhode Island	
Texas	planned for 1994-95
Utah	
Washington	
Wisconsin	
Wyoming	

17 TOTAL

NO STATE ASSESSMENT

Alaska
Iowa
Montana
Nebraska
North Dakota
Ohio
Vermont

7 TOTAL

State	Level of Centraliz.	Assm't Publ.	When Admin.	Grade Levels	Source of Specs	Individ. Reported	Scores For	Group Reported To	Scores For	
ALABAMA	1	2 SAT	A	4, 8		Yes	compar. between	S.D.B.	indicator of National Standing	
		1 BCT		3,6,9	NOT SCIENCE TESTS					
ALASKA	0	NO STATE ASSESSMENT								
ARIZONA	2	2 ITBS SAT	A	1-8 9-12		School	diagnosis	B.D.	diagnosis	In development: Arizona Student Assessment Plan will include science '93-94
ARKANSAS		MAT6		8, 10		Teacher	diagnosis	D.S	4, 5	
CALIFORNIA	1	1	A	8 (in '90 6, 12)				B S	4 policy	
COLORADO	?	?	1987	3, 6, 9, 11		B	Achiev.	D.S.	5	
CONNECTICUT		NOT IN SCIENCE			In development: Common Core of Learning Assessment used in HS will contain Science component.					
DELAWARE	NOT IN SCIENCE 1	DEAP		11		B	diagnosis placement	D.S.	policy	

State	Level of Centraliz.	Assm't Publ.	When Admin.	Grade Levels	Source of Specs	Individ. Reported	Scores For	Group Reported To	Scores For
FLORIDA	1	1 no name	1989	3,4,5	(pilot)				
GEORGIA		ITBS		2, 4, 7, 9	Teacher committee	Teacher	Program Decisions	D.S.	5 Use is determined by the local level. Changing to state test in early 1990's
HAWAII	N O	S C I E N C E		T E S T					
IDAHO	1	2 ITBS TAP	A	6, 8, 11	1	B	curr & instr. improvement	D.S.	curr & instr. improvement
ILLINOIS	N O	S C I E N C E		11 T E S T	(PLANNED FOR 1992)		No		
INDIANA	1	1 ISTEP	1989	3, 6, 8, 11	2	1	remed.	D S	4 4 or 5
IOWA	N O	S T A T E	A S S E S S M E N T						
KANSAS	1	1	A	2, 4, 6, 8, 10	N O T I N	S C I E N C E			
KENTUCKY	1	1	1987	K-12	N O T I N	S C I E N C E			
LOUISIANA	1	1 LEAP 2 CAT	A	11	2		Graduation requirement	S.	policy

State	Level of Centraliz.	Assm't Publ.	When Admin.	Grade Levels	Source of Specs	Individ. Reported	Scores For	Group Reported To	Scores For
MAINE	1	1 MEA	A	4, 8, 11	1	No		S.D.B	3. Results published 8. Curriculum planning
MARYLAND	1	1	NO SCIENCE TEST					(State test being developed)	
MASSACUSETTS	1	1	88/90	4, 8, 12	1	No		S.D.B.	curr. improve., policy
MICHIGAN	1	1 MEAP	1986,88 1989	(4,8,10) (5,8,11)	1	Yes	diagnosis remed.	S.D.B.	In development: A new draft of MEAP to be used in '92 parent reporting, curr. impr., policy (4.5?)
MINNESOTA	1	1 MEAP	1993 1987	4, 8, 11	1	No		S D B	Building a new mandatory science assessment. curr. improvement
MISSISSIPPI			NO SCIENCE TEST						
MISSOURI	1	MMAT 1	4-year cycle	3, 6, 8, 10	2	Yes	?	S.D.B.	policy, inst. improv. materials rec'd.
MONTANA	2	NO STATE ASSESSMENT							
NEBRASKA	2	NO STATE ASSESSMENT							
NEVADA	1	1	NOT IN SCIENCE						
N HAMPSHIRE	2 w/1	CAT 2	A	4, 8, 10	3 w/1	Yes	S ?	D, B	policy, comparisons instruc. improvement

State	Level of Centraliz.	Assm't Publ.	When Admin.	Grade Levels	Source of Specs	Individ. Reported	Scores For	Group Reported To	Scores For	
NEW JERSEY		N O T I N S C I E N C E								
NEW MEXICO	1	CTBS	A	3, 5, 8,10		Yes	Student Progress diagnosis	B.D.S.	monitor prog., accred., policy	
NEW YORK	1	1 sci. eval. manip. skills; p & p test	A 89	9-12 K-4	2	No		B.S.D.	policy, school improvement	
NORTH CAROLINA	NC Science Achievement Test	1	K-3 4-6 7,8	A	3, 6, 8	2	Yes	Report to instruct. planning	S.D.B.	Teaching prescribed curriculum
NORTH DAKOTA		N O SCIENCE ASSESSMENT	2?	A	All but grade 3	?	Yes	?	B.D.S.	policy
OHIO	N O	S T A T E	A S S E S S M E N T							
OKLAHOMA			A	3, 7, 10	2	Yes	diagnosis	S.D.B.	4, 5 policy	
OREGON	N O T	I N	S C I E N C E		(may be cut altogether)		In development: Science Assessment for use every 5 years for program evaluation. Target date: Spring '93.			
*PENNSYLVANIA	1	EQA *voluntary 1 1 TELLS *mandatory	1	A	4, 6, 7, 9, 11	1 and 2	Yes	?	S.D.	8. program planning program evaluation

State	Level of Centraliz.	Assm't Publ.	When Admin.	Grade Levels	Source of Specs	Individ. Reported	Scores For	Group Reported To	Scores For
RHODE ISLAND			N O T I N S C I E N C E						
S. CAROLINA		1	CTBS (89) SAT (90) 1 and 2 A Basic Skills	(3,6,8)	4,5,7, 9,11	2	Yes	Program dev.	B.D.S. 4,5
S. DAKOTA	1	SAT 2	A	4, 8, 11		Yes	eval.?	B.D.S.	policy, curr. eval.
TENNESSEE	1	2 SAT CTBS4	A	2, 5, 7, 9, 12; 2-8,10	1 and 2	Yes	diagnosis	S.D.B.	policy
TEXAS	1	1	A	N O T I N S C I E N C E			(Planned for 1994-95 for grades 3,5,7,9,11)		
UTAH	N O T I N S C I E N C E								
VERMONT	N O S T A T E A S S E S S M E N T								
VIRGINIA	1	2 ITBS SEPAM?	Annual	4, 8, 11	2	Yes	diagnosis prog. eval.	B.D.	3. public report
WASHINGTON	1	MAT 2	SCIENCE ASSESSMENT SECTION OF MAT OPTIONAL						
W. VIRGINIA	1	2 CTBS	A	3, 6, 9, 11	1/2	Yes	diagnosis	B.D.	policy, curr. eval.

Appendix A

State	Level of Centraliz.	Assm't Publ.	When Admin.	Grade Levels	Source of Specs	Individ. Reported	Scores For	Group Reported To	Scores For
WISCONSIN	1	CTBS 2	N O T	I N	S C I E N C E				
WYOMING	1	NAEP 2	N O T	I N	S C I E N C E				

CODE KEY

<u>Level of Centralization</u>
0 = no state requirements
1 = state test
2 = local tests

<u>Assessment Publisher</u>
1 = state
2 = commercial (name)

<u>When Administered</u>
A = annually

<u>Results Reported to</u>
S = state
D = district
B = school building

Source of Assessment Specs.
1 = state assess. committee
2 = state curr. guidelines
3 = local
4 = want but don't have

<u>Use of Assessment Results</u>
1 = report to state
2 = report to local
3 = public report
4 = state sanction, reward
5 = local sanction, reward
6 = graduation requirement
7 = diagnosis, placement
8 = program/curriculum improvement
9 = policy

SUMMARY OF STANDARDIZED TESTS
OF SCIENCE ACHIEVEMENT
USED IN STATE ASSESSMENTS

NB: Several tests have more recent editions than those cited here, but the modifications to the content/process categorization specifications appear relatively minor. In many of the tests, the science section is an addendum to a core battery.

1982-83 Comprehensive Test of Basic Skills (CTBS)

* 25-40 questions
* Seven levels - Grades 1.6 to 12.9
* Content / Process

Content	Process
Botany	Recall
Zoology	Explicit Info Skills
Ecology	Inferential Reasoning
Physics	Evaluation
Chemistry	
Land/Sea/Space	

1986 Iowa Test of Basic Skills (ITBS)

* 35-45 questions
* Five levels - Grades 3-9; Grades K-3 levels do not offer science
* Content / Process

Life Science
Earth & Space Science
Physics
Chemistry
Health & Safety
Nature of Science
  - Methods of Inquiry
  - Nature of Evidence
  - Nature of Proof
  - Cause and Effect
  - Stability and Change

1986 Metropolitan Achievement Test (MAT6)

* 31-50 questions
* Six levels - Grades 1.5 to 12.9
* Content / Process

Content	Process
Physical Science	Knowledge
Earth and Space Science	Comprehension
Life Science	Inquiry Skills
	Critical Analysis

1987 Tests of Achievement & Proficiency (TAP)

* 54 questions
* Four levels - Grades 9-12
* Content / Process

Content	Process
Nature of Science	Knowledge/Info
Life Science	Comprehension
Earth/Space	Application/Analysis
Chemistry/Physics	Synthesis/Evaluation
	Experimental methods/ Techniques

1986 California Achievement Tests (CAT)

* 25-40 questions
* Nine levels - Grades 1.6 to 12.9
* Content

Botany
Zoology
Ecology
Physics
Chemistry
Land/Sea/Space

STANFORD ACHIEVEMENT TEST

1982 Stanford Achievement Test (SAT)

* 44-60 questions
* Four levels - Grades 3.5 to 9.9
* Content / Process

Physical Science
Biological Science
Inquiry Skills (Process)

Analysis	Infer
Predict	Classify
Experiment	Measure
Hypothesis	

LEVEL	GRADE
Primary 3	3.5-4.9
Intermediate 1	4.5-5.9
Intermediate 2	5.5-7.9
Advanced	7.0-9.9

18 Biological Science
- 7 Living objects
- 11 Environmental Interactions

15 Inquiry Skills
- 3 Infer
- 3 Measure
- 6 Analyze
- 2 Hypothesis
- 1 Classify