DOCUMENT RESUME

ED 377 236 TM 022 497

AUTHOR Buttram, Joan L.
TITLE State-Mandated Student Assessment Programs in the Mid-Atlantic Region.
INSTITUTION Research for Better Schools, Inc., Philadelphia, Pa.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 90
CONTRACT 400-86-0003
NOTE 22p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Tests; *Educational Assessment; Educational Planning; Educational Policy; Educational Practices; Educational Trends; Elementary Secondary Education; Program Descriptions; State Legislation; *State Programs; Test Content; Testing Problems; *Testing Programs; *Test Results; Trend Analysis
IDENTIFIERS Delaware; District of Columbia; *Mandated Tests; Maryland; New Jersey; Pennsylvania; Research for Better Schools Incorporated; *United States (Mid Atlantic States)

ABSTRACT

For the past 8 years, Research for Better Schools, Inc. (RBS) has prepared an analysis of state-wide trends in student achievement in the mid-Atlantic states (i.e., Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania). For this year, instead of quantitatively examining trends in student achievement scores, RBS decided to examine current and planned assessment programs in the five jurisdictions it serves from a programmatic perspective. The report summarizes current or planned efforts in the region, with sections describing activities in each of the five jurisdictions. Review of practices, policies, and plans for the area identifies major questions concerning testing. First concerns the purpose of state-mandated programs, second is the issue of test content, and third is the question of test technology or testing methodology. All of these questions must be addressed to ensure the appropriate use of testing in the region. (Contains 15 references.) (SLD)

TM

STATE-MANDATED STUDENT ASSESSMENT PROGRAMS

IN THE MID-ATLANTIC REGION

JOAN L. BUTTRAM
RESEARCH FOR BETTER SCHOOLS, INC.
444 N. THIRD STREET
PHILADELPHIA, PA 19123

NOVEMBER 2

1990

State-Mandated Student Assessment Programs

in the Mid-Atlantic Region

prepared by:

Joan L. Buttram
Research for Better Schools, Inc.
444 N. Third Street
Philadelphia, PA 19123

November, 1990

This publication is based on work sponsored, wholly, or in part, by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number 400-86-0003. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

Table of Contents

Introduction

For the past eight years, Research for Better Schools (RBS) has prepared an analysis of state-wide trends in student achievement in its region (i.e., Delaware, the District of Columbia, Maryland, New Jersey, and Pennsylvania). These analyses have examined the performance of students in reading, language arts, and mathematics on achievement tests administered by the five state education agencies (SEAs), adding an additional year to the trend analysis each year. Although these analyses have been complicated over time by changes in the standardized test batteries administered, the comparison or normative groups used, and the student samples included in the testing program, they have generally demonstrated that student scores have increased or decreased only slightly over time. In the Mid-Atlantic region, student reading, language arts, and mathematics performance has remained fairly stable and relatively high (in comparison to national norms) over the past decade (Biester, 1990).

This year, RBS decided to examine state-wide trends in student achievement from a somewhat different perspective. Instead of quantitatively analyzing trends in student achievement scores, RBS decided to examine from a programmatic perspective the current and planned assessment programs in the five jurisdictions. In other words, what plans do the five jurisdictions have for assessing student performance over the next five years? Because the laboratory is about to begin a new, five-year contract with the Office of Educational Research and Improvement (OERI) that focuses on improving the outcomes for all students, especially those at risk (OERI, 1990), it seemed especially timely for RBS to document how each jurisdiction planned to measure student outcomes. By documenting how student outcomes would be measured, RBS would have a head start on its work in the region.

This report thus summarizes current and/or planned efforts in the region to assess student performance. These data were gathered in interviews conducted by the author of testing directors and/or other high ranking educational officials in each jurisdiction as well as the review of sudent assessment plans when they existed. The next five sections describe each jurisdiction's student assessment programs, as they exist today and plans for the next five years whenever known. These descriptions will briefly present the development and history of these programs; the samples of students included; the knowledge, skills, and attitudes assessed; and the uses of these data. The last section of this report will identify the major issues facing the student assessment programs in the Mid-Atlantic region as well as suggest areas in which RBS and other R&D organizations can focus their work to help strengthen the valid assessment of student outcomes in the region.

Delaware's Student Assessment Program

Unlike the other three states in the Mid-Atlantic region, Delaware's educational system (approximately 100,000 students enrolled in 164 schools operated by 19 school districts) is comparable in size to a large, urban center. As a result, the Delaware Department of Public Instruction (DPI) is

at least theoretically able to reach out and perhaps work more closely with individual districts and schools in their use of student performance data.

Current State-Mandated Testing Program

The Delaware DPI has administered a commercially published, standardized achievement battery to students for over a decade. During this time period, the battery has changed three times, starting with the California Achievement Test (CAT) from 1978-1983, the Comprehensive Test of Basic Skills (CTBS) from 1984-1988, and most recently the Stanford Achievement Test (SAT) since 1989. Selection of a test publisher has typically been made by DPI after considering the psychometric qualities of the test (e.g., knowledge and skills tested, norming sample), the match between the state's curricula and the test's objectives, and price. Until now, the state program has tested all students in elementary and middle grades (1-8) as well as one secondary grade (11) in the spring of each school year; mainstreamed special education students were included in the tested sample. The SAT test battery (used in 1989 and 1990) included eight to ten subtests depending on grade level, including reading, mathematics, language, spelling, listening, study skills, science, social studies, thinking skills, and using information. Student results are reported in average normal curve equivalents (NCEs) scores at the state, district, and school levels. In 1989, the state-wide testing program was budgeted at approximately $100,000.

SAT results are to be used to improve educational programs at all three levels. To this end, DPI provides teacher and parent guides and sponsors teacher/administrator training sessions on administering the test and interpreting the results. However, no ongoing technical assistance is provided by DPI to either teachers or administrators to use test results to improve classroom instruction. The state board of education uses the results for accountability purposes, primarily at the state and district levels. However, there is no system for distributing rewards or penalties based on district/school performance. Unlike some of the other state-mandated testing efforts in the Mid-Atlantic region, Delaware's program does not really fall into the "high stakes" category. High stakes testing occurs when educators and/or students perceive that the results have significant consequences and will be used to make important decisions (Madaus, 1988).

Planned State-wide Testing Program

In September, DPI issued an RFP for a testing contractor to administer a standardized achievement test battery for the next five years. All of the major commercial test publishers have been invited to respond and the selection criteria described above will be used to select a test publisher. The RFP proposes some major changes in the state-mandated testing program. First, no longer will all students in grades 1-8 and 11 be tested. Instead, DPI elected to narrow the sample to include small samples of students in grades 1 and 2 and census testing in grades 3, 6, 8, and 11. In grades 1 and 2, classrooms will be randomly selected across the state; current estimates suggest that approximately 1,000 first and second graders (25-30 percent) will be tested. In grades 1-3, students will be tested on the basic skills; in the other three grades, students will be tested on social

studies and science as well as the basic skills. In addition, all Chapter 1 students in grades 1-12 will be tested on the basic skills; this testing is required for Delaware to continue receiving federal Chapter 1 funds. As before, the results will be used for both school accountability and improvement purposes and no rewards or penalties based on district/ school performance are anticipated.

DPI staff members decided to reduce the amount of standardized testing because they felt that it simply didn't make sense to test every student every year. The revised sampling and testing schedule plans provide sufficient data for DPI to track student and school/district performance over time. The decision to further reduce testing at the early grades (i.e., grades 1 and 2) was made in response to the growing concerns of early childhood educators in Delaware and nationwide about both the validity and potential harm of testing young children. Although DPI has backed away from commercially published standardized achievement tests for young children, staff members privately acknowledge that other assessments will have to be found if DPI is to continue investing in early childhood education.

DPI also is preparing to put into place a writing assessment for students in grade 10. At the current time, DPI expects this assessment to call for students to produce a writing sample that will be scored holistically. Development of this test will be contracted out to an external agency. No other state-wide testing programs are planned for the general student population at this time. Nevertheless, it's important to note that the chief state school officer has recently resigned and so additional state testing programs may be considered after the selection of a new state educational leader.

District of Columbia's Student Assessment Program

The District of Columbia Public Schools (DCPS) is included in this report because it operates as an independent jurisdiction in the Mid-Atlantic region. There are approximately 88,000 students enrolled in the district's 183 schools. Because the district operates as both the SEA and LEA, its testing concerns are more extensive and diverse than the other jurisdictions in this region.

Current Testing Programs

For the past six years, the DCPS has administered the Comprehensive Test of Basic Skills (CTBS) to students in grades 3, 6, 8, 9, 10, and 11. In addition, Chapter 1 students in grades 2, 4, 5, and 7 have been tested on the CTBS. Seven subtests are administered in the spring of each school year, including reading, mathematics, language, science, social studies, and reference skills. Form S, an early version of the CTBS, was administered from 1984 through 1986 and Form U, a more recently published version, has been administered since then. Student progress (grade equivalents based on national norms) is reported by district and school levels. The district spends approximately $37,000 annually scoring the CTBS; the test booklets have been used for several years now and so other costs associated with the

testing program are difficult to calculate.

The CTBS was selected by a panel of elementary and secondary classroom teachers, principals, curriculum and assessment specialists, and parents. Their review was based on the relative match of test objectives with the district's curriculum, test content and reporting options, and price. In order to train teachers to administer the CTBS, chairpersons are named for each school who receive training from the central office; they, in turn, are responsible for turnkey training in their home schools.

The CTBS results are released publicly in the newspaper and in district-prepared reports and receive widespread attention from the board of education and the public as well. These reports attend most directly to issues of accountability, both at the individual school and at the district level. The district also reports using the results to drive school improvement efforts. Last year, the district superintendent convened a panel of district educators and community representatives who examined the CTBS results in great detail and prepared a plan for improving instruction. Thirty-five schools with low CTBS scores were identified and were to be made the beneficiaries of as many district resources as could be located to help improve their test scores; recent press reports suggest that adequate instructional resources are still lacking in these schools. Nevertheless, in comparison to the other five jurisdictions, the DCPS provides the most extensive assistance to schools in interpreting and using test results for improvement. Supervisors are assigned to each school who are charged with working with classroom teachers to improve instruction (among other responsibilities); much of their interaction focuses on the use of detailed CTBS testing reports at the individual student and test objective level to plan appropriate classroom instructional activities.

In the last few years, the district also put into place a highly publicized, criterion-referenced end-of-course testing program at the secondary level. Examinations were developed for administration in 31 courses across the district; they contain multiple choice and essay items and are to count one-fifth of the students' final grades. However, there are no procedures to ensure their administration, inclusion in the calculation of semester grades, or reporting on student report cards. As a result, the program's initial promise has diminished over time.

In addition to these two, the district administers numerous other programs meant to provide data on the progress of individual students, but not individual schools or the district overall. A locally developed pre-kindergarten observation checklist and the Metropolitan Readiness Test are used to identify young children that are not succeeding in the early grades. The district has just initiated a writing assessment in grades 3 and 7; this assessment relies on a commercially published battery and will be used for instructional planning. Students in grade 8 are given the Ohio Vocational Interest Survey; the results are intended to help students and counselors plan a realistic course of study for all students in their high school program. A life skills test is administered to students in grade 10; if students meet the cut-off score on this test, they are excused from a district, one semester life skills course that focuses on broadly defined

skills needed to survive in today's world (e.g., filling out applications and forms, reading a graph).

Planned Testing Programs

At this point in time, DCPS plans to continue administering the CTBS for another year or so. After that, the testing director expects to switch test batteries so that a more current norm group is used in assessing student performance. Although some consideration is being given to modifying other parts of the district's testing program as described above (e.g., pre-kindergarten, kindergarten, and first grade assessments), no definitive plans exist as yet.

Maryland's Student Assessment Program

In Maryland, there are approximately 690,000 students enrolled in 1,201 schools in 24 school districts, one per county with the addition of Baltimore City. Although larger in population than Delaware, the organizational structure of Maryland's educational system is relatively small, much like that of Delaware, and so the Maryland State Department of Education (MSDE) generally has been able to adopt a fairly proactive, but collaborative relationship with the 24 school districts on a variety of educational issues, including the assessment of student performance.

Current State-Mandated Testing Programs

Since the late 1970s, there have been two state-mandated testing programs in place in Maryland. The first program involves functional tests in reading, mathematics, writing, and citizenship. These four tests were developed as part of a larger state-wide school improvement initiative (SITIP, School Improvement through Instructional Practice) that involved extensive curriculum development in all four areas as well as intensive staff development on four instructional models (e.g., mastery learning, cooperative learning). The functional tests are administered to all students in the ninth grade at different points during the school year and they are expected to pass all four by the end of their high school programs in order to receive a diploma. Results are reported in terms of the percent who have passed each test by school, district, and the state overall.

It's important to note that the test objectives for the functional tests were developed by state and local educators together after reaching agreement on the state's curricula. MSDE issued contracts to external agencies to develop the four functional tests based on these objectives. These tests are criterion-based and thus provide a good measure of how much of the established curricula individual students have mastered. MSDE also has developed diagnostics that provide direction and assistance to the teacher and student when a student fails one of the functional tests.

In addition to the functional testing program, MSDE has administered the California Achievement Test (CAT) to all students in grades 3, 5, and 8 in the fall of each school year for the past six years. The specific

subtests administered are reading, mathematics, and language arts. Average grade equivalents are reported on district (county) and school levels in a state-wide report. The results are fed back to schools to be used for improvement purposes, but as the state testing director noted, "the results often came back too late in the school year to be of much real use to the teachers." The results also are released to the general public in MSDE's annual report.

The results of these two testing programs are intended by MSDE to be used for both school accountability and improvement purposes. The functional tests are considered high stake in that students have to pass all four in order to receive their diplomas; however, most students pass the reading and mathematics tests and so they have had little impact on the instructional programs in the schools. Performance on the writing and citizenship tests has been more problematic for some districts and so the specific content of these two tests has influenced individual district curricula and instructional programs (Corbett and Wilson, 1990). The CAT testing receives attention from the media when the results are first released, but state scores have generally been high and so little controversy is generated. In reality, the results of the CAT do little to drive either the school accountability or improvement agenda.

Planned State Testing Programs

Maryland is probably the most innovative SEA in the Mid-Atlantic region in terms of its student assessment programs. MSDE has modified its normative state-wide testing program (i.e., CAT) so that it will include two parts, the commercially published CTBS and a set of performance assessments currently being designed as part of the Maryland School Performance Program. The state's functional testing program will remain in place in its current form for now, although the state testing director anticipates that eventually students will be able to substitute their scores on the new performance assessment tests in lieu of the functional tests. The remaining parts of this section will describe these two initiatives.

CTBS. The CTBS will be administered to a random sample of 750 students in each district, 250 in grades 3, 5, and 8 each. This means that there could be as few as three students per grade per school taking the CTBS. This is in sharp contrast to previous years where all third, fifth, and eighth grade students were tested. This substantially reduces the testing time that most students will face, though it may create problems logistically for some schools in locating and segregating students for testing during the school day. Nevertheless, it does represent a significant reduction in school time devoted to standardized testing programs. As in Delaware, all Chapter 1 students will be administered the CTBS in order for Maryland to continue receiving these federal funds.

As before, the reading, mathematics, and language arts subtests of the battery will be administered. Average grade equivalents will be calculated and reported by district as well as for the state overall. These scores are meant for comparison purposes only. Very simply, these will provide a base on which to compare the performance of districts and/or the state to the

nation as a whole. There is no intention that student performance on the CTBS be used to drive school improvement efforts in Maryland.

Maryland School Performance Program. Performance assessments of Maryland students will be conducted as part of the state's initiative to develop a comprehensive system of public accountability at the individual school, district, and state level. As part of this effort, MSDE will begin collecting information on the performance of students, schools, districts, and the state overall on a broad range of variables, including the assessed knowledge of students, their participation in school (i.e., attendance and drop-out data), student attainment or promotion rates, and their post-secondary plans and decisions. In addition, MSDE will publish other supporting information on the numbers of students enrolled in school as well as entrants and withdrawals; the wealth and expenditure per pupil; the instructional, professional, and instructional aide staff support for students; the length of the school day and year for pupils; and the number of students receiving special programs (i.e., special education, bilingual, Chapter 1, and free/reduced meal programs). This program is intended as an outcome-based educational approach which identifies crucial indicators of student and school performance, collects and publishes data on each area, compares the results against a set of state-wide standards, and develops and implements school improvement plans based on the needs identified by the data.

The state's plans for student performance assessments are most germane to our discussion here. Although the assessments are being developed over time, the system will be described in finished form (Governor's Commission on School Performance, 1989). Performance assessments will be developed in five areas -- reading, mathematics, writing/language usage, science, and social studies. They will be administered at grades 3, 5, 8, and 11 and are expected to tap the residual concepts, skills, competencies, and processes that students will have accumulated in the preceding grades. In other words, students in grade 3 will be tested on what they've gained from the primary grades, students in grade 5 will be tested on grade 4 and 5 material, and so on. Students will be asked to complete a series of performance tasks that are related to the state's identified learning outcomes; are authentic, real-world activities; include pre-assessment and instructional activities that set the stage for the context and themes of the task; require students to use higher order thinking skills and connect the concepts, context, and processes within the discipline; and require multiple responses within a context.

Educators at both the local and state levels have been involved in the development of the curriculum content and specifications of the performance assessments, constructing and reviewing tasks, and will be involved in reviewing the results as well as reporting formats. At the current time, MSDE expects to report results in terms of school performance in the five content proficiency areas and provide instructional guides that will recommend instructional activities and materials to address content area weaknesses. It should be noted that MSDE is currently focusing its attention on the development of the tasks and so has completed only rudimentary plans for reporting results or using results to plan appropriate

instruction. It is assumed that other levels of reporting (e.g., student, district) as well as the interpretation and use of results will be explored in more depth once the design work for the various performance tasks is completed. Nevertheless, the results of these tests are expected to drive major instructional reform in the state and thus are expected to have fairly high stakes.

New Jersey's Student Assessment Program

Over 1.1 million students are enrolled in 2,304 schools operated by 592 school districts. The organizational structure of the New Jersey educational system is very complex. In addition to the large number of individual school districts, there are four intermediate service agencies (ISAs) and county education offices. A new state commissioner has recently been appointed to lead the New Jersey Department of Education (NJDE) and so it is difficult to predict what stance he will take in terms of his relationship with the individual districts or his position on student assessment; however, the department has a history of using state-wide assessment programs to hold school districts accountable and push particular school improvement agendas. Unlike the three previously described jurisdictions (Delaware, District of Columbia, and Maryland), NJDE applies sanctions when school districts do not meet expected state performance levels and so state-mandated testing programs are clearly high stake here.

Current State-Mandated Testing Programs

The NJDE state-mandated testing programs have changed considerably over the past decade, and as will be clear in the following section, are expected to go through additional changes in the 1990s. The state's involvement in testing programs dates back to the 1975 legislation surrounding the Thorough and Efficient (T&E) decision. At that time, NJDE instituted a state-wide assessment program to make sure that all students were receiving a thorough and efficient education by tracking student performance at particular grade levels; the testing program at that time was developed by an external contractor.

In 1982, NJDE decided that a more appropriate route to assess the delivery of a thorough and efficient education would be a minimum basic skills, criterion-referenced test (Minimum Basics Skills Testing Program, or MBS). In response, educators at the local and state levels worked together to develop test specifications in the areas of reading and mathematics and tests were developed for grades 3, 6, 9, and 11 to match those specifications. The test were administered and results were reported back at the individual school, district, and state level to track student performance in the target grades and provide accountability data.

About this same time, a new governor was elected who, in turn, appointed a new educational commissioner. Together, they decided that significant numbers of New Jersey students were not adequately prepared to meet the increasing demands of the world of work. In response to pressure from the business community and others, New Jersey moved to a new, more stringent testing program that included two basic parts. First, the third

and sixth grade testing of the earlier program was eliminated and districts were allowed to substitute testing on commercially published standardized achievement batteries in reading and math; cut-off scores were established for these tests and districts were required to submit their results to the state. NJDE provided funds for additional instructional support to those districts that did not meet the state standards. In addition, the state's monitoring program became more intensive as these standards were not met.

Very soon after that, the state developed a High School Proficiency Test (HSPT) that all high school students are required to pass if they are to receive a high school diploma. The test has three subparts -- reading, mathematics, and writing, all roughly at the 9th grade level. All three subtests include a multiple choice section, the writing subtest also asks students to produce a sample in response to a test-provided stimulus. This test was seen as raising the standard that New Jersey students were expected to meet. Similar to the state's experience with earlier testing programs, more and more districts (and students) met the state standard over time.

However, the business community was not yet satisfied with the performance of New Jersey high school graduates and so continued to pressure the governor and the education commissioner to raise the standard a second time. Since NJDE's response represents its future work, we will move on to the next part.

Planned State Testing Programs

Starting with the graduation class of 1994, all high school students will have to pass the 11th grade New Jersey High School Proficiency Test. This test will have three parts -- reading, mathematics, and writing. In many ways, the 11th grade HSPT mirrors the current 9th grade HSPT, except that the test will cover knowledge and skills expected of 11th grade students.

As might be expected, the development of the 11th grade HSPT followed the same procedures as were used for the 9th grade HSPT (NJDE, 1990). Reading, mathematics, and writing committees of local and state educators, parents, students, and representatives of the business community met to identify the skills that high school students will need to function politically, economically, and socially in an increasingly complex technological society. Each committee's specific charge was to identify skills to be assessed on the 11th grade test, using the 9th grade test as the starting point. The knowledge and skills identified by the committees have incorporated and expanded upon the ninth grade skills to emphasize thinking, problem solving, reasoning, and decisionmaking appropriate for 11th graders. These lists were circulated to all public school districts in the state as well as to other interested groups for review and comment. They generally reinforced the work of the individual committees, although some modifications were made based on the review process. Following this, reading, mathematics, and writing committees comprised of school educators developed sample test items and specifications that were forwarded to an external contractor for development of the test items. This process is almost completed and the first of three years of "due process testing"

(i.e., required legally to give students sufficient notice of change in graduation requirements) is expected to occur in December, 1990. As with the 9th grade HSPT, the 11th grade HSPT will establish cut-offs that students must score above in each area; students will have to pass all three areas in order to obtain their high school diploma. NJDE has allocated approximately 1.1 million dollars for the development of the 11th grade test.

The NJDE also is developing an 8th grade, early warning test. Although it will test reading, mathematics, and writing skills appropriate to the 8th grade, it is meant to provide advance notice to students, their parents, and schools that students are in danger of not meeting the state standard. Development of the 8th grade test has followed the same procedures described above for the 11th grade test. The 8th grade test will be administered for the first time in March, 1991. Because it is an early warning test, and not a graduation test, the state is not required to go through "due testing" procedures. Approximately one million dollars have been allocated for development of the 8th grade test.

These two tests are seen by state officials as driving the state's school improvement agenda. They provide an explicit standard that school districts are expected to meet. Districts that fall short are provided test data that pinpoints their weaknesses. These tests are clearly high stakes, for both the students and the districts in which they're enrolled. As a result, most New Jersey districts view the state-mandated program as pushing the school accountability and not the school improvement agenda.

Pennsylvania's Student Assessment Program

Pennsylvania is the largest state in the Mid-Atlantic region. Approximately 1.7 million students are enrolled in 500 districts that operate 3,248 schools. Similar to New Jersey, there is a network of 29 ISAs that provide assistance and channel funds to individual school districts. Like New Jersey, Pennsylvania has a new state leader, however his agenda for state-wide student assessment programs is more public than his counterpart's in New Jersey and will be discussed below in terms of PDE's plans for state testing programs.

Current State-Mandated Testing Programs

Until the 1988-89 school year, Pennsylvania had two complimentary testing programs operating state-wide that approached student performance from two very different perspectives. The first program, Educational Quality Assessment (EQA), reported on student performance at the school level. The second program, Testing for Essential Learning and Literacy Skills (TELLS), provided performance data at the individual student level. Both are described in more detail below.

EQA. EQA was developed in 1968 to provide school-based assessments on the state's twelve goals for quality education (i.e., collect data on how well districts were meeting the state's twelve goals). These goals covered

a wide variety of areas, including reading, writing, mathematics, analytical thinking, social studies, arts and humanities, science and technology, environment, self-concept, health practices, and health knowledge. Districts were allowed to decide when to administer the EQA, as long as it was administered once every five years. Prior to 1985, students in grades 5, 8, and 11 completed the test battery in the spring; from 1986-1988, students in grades 4, 6, 7, 9, and 11 took the test at the same time of the school year. Results were reported by school in raw scores and percentiles based on state norms and were meant to be used by schools as part of their long-range planning and school improvement efforts. The test was not administered in 1989.

TELLS. TELLS was developed in 1984 in direct response to the growing concern nationwide on the need for educational reform. At that time, the state had not established formal standards for student performance and in some circles, there was pressure to follow the direction of Maryland and New Jersey and many other states who had adopted minimum proficiency or other types of tests required for graduating and receiving a high school diploma. Rather than following this route, Pennsylvania elected to set standards for performance (via the TELLS) and then provide assistance (via funds) to students not meeting those standards.

TELLS tested all students in grades 3, 5, and 8 in reading and mathematics in March of each school year. Items were chosen from a nationally standardized item pool each year so that student results could be reported in terms of raw scores, percent correct, percentage above/below state-established cut-off scores as well as percentiles estimated from national norms. The results were reported to districts at the school and district levels, and released to the public at district and state levels.

Although there was pressure to release scores at the school level, PDE resisted until the 1988-89 school year when the scores were released and used to rank individual schools across the state.. This prompted widespread criticism from educators at all levels and partially contributed to the resignation of the secretary of education. Many educators at the local level felt that this act represented a breach of faith in how TELLS data were to be used. Although the test continues to be administered, there has been deliberate de-emphasis by the new secretary in the reporting of results. At one point this fall, the state board of education included on its agenda a motion to drop the TELLS test altogether, however this motion was dropped before being considered; it is expected to be resubmitted following the election in November. In addition, no funds were allocated to provide additional support to districts in the current state's budget and so its future is somewhat suspect.

Planned State Testing Programs

At this time, PDE is probably the least sure of the five SEAs in the Mid-Atlantic region on its plans for assessing student performance. PDE officials privately acknowledge that of the two current state assessment programs, EQA has the most potential to push a school improvement agenda. The state board of education has continued to affirm the twelve goals on

which the EQA test items are based and so measuring school performance in relation to these goals (or standards) has some appeal. Nevertheless, the test has been administered inconsistently and has few strong advocates. Although calls for school accountability have typically favored TELLS over EQA, the failure of the state legislature to fund the follow-up assistance part of the TELLS program and the recent stir over the use of TELLS results has left this testing program with few advocates. In reality, there are few calls for either program.

If PDE wished to pursue more innovative alternative student assessment programs, it is at a particularly advantageous turning point. Since re-election of the current governor is almost certain, the current secretary of education will most likely remain in place. He has been an outspoken critic of standardized testing programs and so the climate is ripe for more innovative testing efforts in Pennsylvania if support for a state-wide effort can be found. And that is a big if in Pennsylvania. Instead, it's more likely that the state will end up with one of its existing programs. Innovative student assessment efforts at the local level will most likely be supported in one way or another (e.g., verbal support, state waivers on some reporting requirements), but it's unlikely that PDE will lead a state-wide effort.

Testing Questions Facing the Mid-Atlantic Region

There are three major testing questions facing the jurisdictions in the Mid-Atlantic region. First, and perhaps the most fundamental question of all, concerns the purpose of state-mandated testing programs. More simply, do they exist principally as ways to hold schools and/or teachers accountable for student learning or to drive school improvement agendas by providing input on a school's instructional strengths and weaknesses? Second is the issue of test content, or on what knowledge and skills should students be examined? For the purposes of this paper, the third question will be referred to as test technology, or what methods should be used to examine student progress? The answers to these three questions vary from jurisdiction to jurisdiction in this region; in some cases, the answers are changing and have the potential to radically reform the assessment of student performance over the next five years.

Purpose of Testing

The two purposes of state-wide testing that seem to cause the most difficulty for all concerned are school accountability and school improvement. As Emerson Elliott (1990) acknowledged at the most recent Educational Testing Service (ETS) conference, these two purposes are more often than not at loggerheads and so it may be impossible for one test to serve both purposes, at least as they are configured now. Because of the high stakes involved in most state programs, attempts to use their results to drive school improvement initiatives quickly become overshadowed by demands from the state legislature and the public to rank and label schools as effective or ineffective. Once this happens, educational practitioners are not likely to see test results from these programs as providing much

useful information for improving their instructional programs.

In the Mid-Atlantic region, all five of the jurisdictions are still grappling with this dilemma. Top educational leaders and test directors argue that both purposes are important, and some even privately give the edge to school improvement over school accountability. Nevertheless, they all acknowledge the political realities that can and do push the school accountability agenda to the forefront and subvert the school improvement process. Assistance will be needed to successfully join these two purposes in building the understanding and support of governing officials and the general public for state-mandated programs that can truly drive a school improvement agenda.

Test Content

Since the publication of A Nation At Risk (National Commission on Excellence in Education, 1983), there has been increasing recognition that too much attention has been paid to the development of basic skills at the expense of higher order skills. This is especially the case for disadvantaged children who have become victims of the federal and state programs (e.g., Chapter 1) ironically designed to redress their learning deficits. State-mandated testing programs and other commercially published achievement batteries were developed to make sure that these programs were effective and students mastered their abcs. However, these testing programs have had other, unintentional effects that eventually served to narrow the curriculum to discrete skill bands. By concentrating on well-defined reading, writing mechanics, and mathematics skills, students, in essence, were deprived of the opportunity to learn more broadly defined knowledge and skills that would help them to become independent, thoughtful learners prepared to face the demands of the ever changing world.

In the past few years, both the educational R&D community and practitioners alike have realized that the U.S. curriculum has become too narrow in scope and students must be taught how to learn if they are to succeed. Students must learn how to apply their knowledge and skills to solve problems in real world settings; learning must be contextualized (Resnick, 1990). And if this is the case, then the content focus of tests must be altered to provide students with opportunities to identify and solve problems, transform information, explain events and relationships, apply principles, and design and execute their solutions (Baker, 1990). To use the current test lingo, authentic assessments must be designed that require students to demonstrate their higher order thinking skills.

The five jurisdictions in the Mid-Atlantic region are in very different stages of responding to this need. Although all five acknowledge the need to expand the content focus of current testing programs, only Maryland, and New Jersey to a lesser extent, have begun to deal with this issue in any significant way. Both have revised their test content objectives to pay attention to higher order skills. In Maryland, the new test specifications explicitly call for the inclusion of a particular higher order thinking skills framework (ASCD, 1985) to be used in the development of tasks. In New Jersey, higher order skills are to be embedded in the test items, but no

specific link was made between testing content and any framework for defining thinking skills. Both states also have relied on the expanding curriculum guidelines developed by professional associations and other groups (e.g., National Council of Teachers of Mathematics and Project 2061) to ensure that their test batteries reflect recent thinking about particular content areas. However, both state testing efforts are in their infancy and so it is too soon to predict their success in either developing appropriate content measures or the impact of these measures on classroom instruction or student learning.

The other three jurisdictions (i.e., Delaware, the District of Columbia, and Pennsylvania) recognize the need to assess student development of higher order skills but so far have been unable to expand their current student assessment programs to include higher order thinking skills beyond what is now tested in commercially published standardized achievement batteries. In two of the three SEAs (i.e., Delaware and the District of Columbia), there has been a strong, almost exclusive reliance on commercially published test products to assess student performance; this reliance precludes increased emphasis of higher order skills until these companies include more higher order skills in their batteries. In the third SEA (i.e., Pennsylvania), a fairly extensive history exists in the development of state-mandated tests (e.g., EQA, TELLS). Unfortunately, much of PDE's staff expertise has been lost over the past few years and so it's questionable whether PDE could again muster the resources needed to develop such a test. High ranking PDE officials acknowledge the need to include higher order skills in student assessment programs, but suspect that this will most likely happen as a result of individual school district efforts rather than state-wide initiatives.

Test Technology

Most state-mandated programs currently rely on multiple choice items (except in the writing content area which almost always includes a writing sample as part of the assessment in the Mid-Atlantic region), and there is almost universal dissatisfaction with these items to assess student knowledge and skills. The litany of complaints about multiple choice items is extensive, including their over-emphasis on simple recognition and recall rather than higher order skills and their lack of authenticity, or the low correlation between performance on these items and application of knowledge and skills in more real-world settings. Clearly, most critics calling for the inclusion of higher order thinking skills in state-mandated testing programs are also calling for alternative, authentic assessment techniques. They argue that assessment strategies must be consistent with the content tested and that it is questionable, at best, whether one can validly test student's attainment of higher order skills with multiple choice test items, and even if one could, whether one should.

As was noted over and over again at the recent OERI conference, The Promise and Peril of Alternative Assessments (October, 1990), the psychometric technology on alternative assessments is growing day by day. Richard Shavelson (1990) at this conference identified six different strategies that are currently being developed to assess student progress in science that

incorporate a higher order thinking skill perspective. Other presenters (e.g., Eva Baker, Joan Boykoff Baron, Dennie Palmer Wolf) shared additional examples of developmental work in a broad array of disciplines. At the same time, many of the same conference presenters and discussants (e.g., Baker, Baron, Robert Linn, and Shavelson) noted that many methodological problems still plague these alternatives. For example, alternative test constructors have identified procedures to establish interrater reliability for scoring these measures, but the problem of intertask reliability has not yet been addressed satisfactorily. Unlike the issue of test content discussed above, there are still significant questions to be answered before student performance can be validly and reliably assessed using alternative measures.

In the Mid-Atlantic region, alternative assessment strategies currently in use at the state level are limited to writing samples. The District of Columbia, Maryland, and New Jersey all have state programs that include such assessments and Delaware expects to issue an RFP shortly for test services in this area; only Pennsylvania currently does not use any form of alternative assessment in its state programs. In terms of other types of alternative assessments, Maryland is the only SEA currently developing performance assessments, and not surprisingly, it is grappling with many of the methodological issues commonly discussed. The other four SEAs have not yet begun to tackle these difficult issues and so if alternative assessments are to be the future for assessing student performance, there is much to be done in this region.

Implications for Mid-Atlantic Educational R&D Community

To summarize, the educational leadership and testing directors of the Mid-Atlantic region generally support the need for state-mandated testing programs to assess student development in broad knowledge and skill areas, including higher order skills; alternative assessments to validly measure student attainment of those skills; and the use of results from those measures to improve instructional programs in the state. At the same time, close attention must be paid to calls for school accountability from state governors and legislatures, the business community, and the voting public. And too often, the demands for school accountability overshadow and even drown out the calls for school improvement.

At the recent OERI conference on alternative assessments, Resnick (1990) called for the marriage of these two purposes if American education is going to adequately prepare all students for the challenges they will face after leaving school. As noted earlier in this paper, these two purposes are too often seen as being in conflict, primarily because the standards for school accountability (e.g., emphasis on basic skills as defined in most state-mandated or commercially published achievement batteries) do not match and may actually conflict with the standards envisioned by educators for schools (e.g., emphasis on higher order skills). If Resnick is right, then the educational R&D community and practitioners can no longer accept current measures of student performance as sufficient. They must join together to build the understanding, commitment, resources, and technology so that school accountability standards and assessment methods more accurately reflect current thinking on the essential knowledge

and skills students must achieve.

In terms of understanding and commitment, the educational community must continue to emphasize the importance of teaching students how to learn. This message must be carried to governing groups and to the general public, and pressure must be applied so that students' development of basic skills is no longer seen as an adequate curriculum for school districts to follow. Development of alternative assessments is an expensive undertaking and so the need for resources will be great; resources must be found not only to support the psychometric development of these measures but also to provide teachers with the training and time they will need to learn how to develop and use them as part of their instructional program. In terms of technology, the educational R&D community must continue to work hard to develop alternative assessment strategies and to transfer these strategies, once developed, to both commercial test publishers and educational practitioners. If the educational R&D community is to be successful in helping to carry out this major reform, the demands will be great. However, the potential payoffs for success are significant and the penalties for failure severe.

References

Association for Supervision and Curriculum Development. (1985). Developing minds: A resource book for teaching thinking. Alexandria, VA: Author.

Baker, E. L. (1990, October). What "probably works" in alternative assessment. Conference conducted by the Office of Educational Research and Improvement, The Promise and Peril of Alternative Assessment, Washington, DC.

Baron, J. B. (1990, October). Use of alternative assessment in state assessments: The Connecticut experience. Conference conducted by the Office of Educational Research and Improvement, The Promise and Peril of Alternative Assessment, Washington, DC.

Biester, T. W. (1990). Trends in school improvement: Regional test results. Philadelphia: Research for Better Schools.

Corbett, H. D. & Wilson, B. L. (1990). Testing, reform, and rebellion. Norwood, NJ: Ablix.

Elliott, E. (1990, October). Opening comments. Conference conducted by the Educational Testing Service, The Assessment of National Educational Goals, New York City.

Governors' Commission on School Performance. (1989). The report of the Governor's Commission on School Performance. Annapolis, MD: Author.

Linn, R. L. (1990, October). Comments. Conference conducted by the Office of Educational Research and Improvement, The Promise and Peril of Alternative Assessment, Washington, DC.

Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), Critical issues in curriculum: 87th yearbook of the NSSE, Part 1. Chicago: University of Chicago Press.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, DC: U.S. Department of Education.

New Jersey Department of Education. (1990). The reports of the reading, mathematics, and writing committees. Trenton, NJ: Author.

Office of Educational Research and Improvement. (1990). Regional educational laboratory request for proposal. Washington, DC: U.S. Department of Education.

Resnick, L. B. (1990, October). Assessment and educational standards. Conference conducted by the Office of Educational Research and Improvement, The Promise and Peril of Alternative Assessment, Washington, DC.

Shavelson, R. J. (1990, October). What alternative assessment looks like in science. Conference conducted by the Office of Educational Research and Improvement, The Promise and Peril of Alternative Assessment, Washington, DC.

Wolf, D. P. (1990, October). Assessment as an episode of learning. Conference conducted by the Office of Educational Research and Improvement, The Promise and Peril of Alternative Assessment, Washington, DC.