

**Final Report:
Perceived Effects of the Maryland School
Performance Assessment Program**

CSE Technical Report 409

Daniel Koretz, Karen Mitchell,
Sheila Barron, and Sarah Keith

CRESST/RAND Institute on Education and Training

March 1996

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

Copyright © 1996 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA Catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education, and by grant number 94-02248-000 from The Pew Charitable Trusts.

The findings and opinions expressed in this report do not reflect the policies or positions of the Office of Education Research and Improvement, the U.S. Department of Education, or The Pew Charitable Trusts.

CONTENTS

TABLES.....	v
ACKNOWLEDGMENTS.....	vi
SUMMARY.....	vii
BACKGROUND AND RESEARCH OBJECTIVES.....	1
PROCEDURES	3
Sampling.....	3
Data Collection.....	4
Data Analysis.....	5
Generalizability of Findings.....	7
SUPPORT FOR MSPAP.....	8
Overall Support for the Program.....	9
Educators’ Judgments About MSPAP as a Lever for Reform	10
Educators’ Judgments About MSPAP as a Tool for Monitoring	11
School Performance	
Support for the Testing Domain, Administration, and Scoring	11
Perceived Accuracy of Student Achievement Information.....	13
Perceived Accuracy of School Effectiveness Data.....	14
Educators’ Judgments About MSPAP as an Accountability Instrument	16
TEACHER PROFESSIONAL DEVELOPMENT FOR MSPAP.....	17
Types of Professional Development Activities	18
Activities Related to the Assessment	18
Activities Related to Instruction	20
Professional Development Vehicles.....	21
IMPACT OF MSPAP ON SCHOOL MANAGEMENT AND INSTRUCTION.....	23
School Climate	23
The Changing Role of the Principal.....	24
Expectations for Student Achievement.....	26
Course Offerings	27
Student Grouping and Remediation	28

Effects on Instruction	29
Responses to Forced-Choice Questions About Instruction	29
Teachers' Responses to Open-Ended Questions About Instruction.....	33
PREPARING STUDENTS FOR MSPAP.....	34
Methods for Preparing Students	35
Instructional Approaches.....	35
Motivational Approaches	37
Direct Test Preparation.....	37
Questionable Test Preparation and Administration	41
Perceived Causes of Gains on MSPAP	43
DISCUSSION	44
Implications	46
Lack of Support for Accountability.....	47
The Perceived Burdens Imposed by MSAP	48
Approaches for Providing Student-Level Scores.....	48
Specificity of Curriculum Frameworks	49
Effects on Equity	50
Writing in Instruction and Assessment.....	50
Other Effects on Instruction.....	51
The Potential for Inflated Scores	52
Other Issues of Validity	54
Next Steps.....	55
REFERENCES.....	56

TABLES

Table 1.	Sample Sizes and Response Rates for Maryland Teachers4 by Survey Medium
Table 2.	Percent of Teachers Participating Directly in Testing Activities.....19
Table 3.	Percent of Principals Reporting Rewarding Teachers Based25 on Their Students' MSPAP Scores
Table 4.	Percent of Teachers Reporting That Academic Expectations26 Have "Increased Greatly," by Grade and Type of Student
Table 5.	Percent of Teachers Reporting That the Emphasis on High27 Standards Has Been Harmful or Helpful, by Grade and Type of Student
Table 6.	Percent of Fifth-Grade Teachers Reporting Changes in31 Content Emphasis
Table 7.	Percent of Fifth-Grade Teachers Reporting Changes in32 Content Emphasis Within Language Arts and Mathematics
Table 8.	Percent of Eighth-Grade Mathematics Teachers Reporting33 Changes in Content Emphasis
Table 9.	Reported Hours Devoted to Test-Preparation Activities,39 Fifth-Grade Teachers
Table 10.	Reported Hours Devoted to Test-Preparation Activities,40 Eighth-Grade Mathematics Teachers (Mathematics Only)
Table 11.	Percent of Teachers Reporting Incidence of Questionable42 Test-Preparation and Administration Practices
Table 12.	Percent of Teachers and Principals Reporting That Each43 Factor Contributed "A Great Deal" to MSPAP Gains

ACKNOWLEDGMENTS

We would like to thank many people whose efforts are reflected in this report. Above all, we would like to thank the hundreds of Maryland educators who contributed their time to participate in our surveys, especially the many teachers who took the time to complete both our telephone interview and our lengthy mail survey. We also want to express our gratitude to numerous people in the Maryland State Department of Education and in participating districts whose support and cooperation made this study possible—in particular, Robert Gabrys, Assistant State Superintendent for Research and Development. Several colleagues at RAND assisted us; Melissa Bradley helped us convert our surveys into appropriate forms for CATI interviewing and did the programming of our CATI system, and Brian Stecher offered many helpful suggestions throughout the course of the study. Amina Assaadi at RAND and Katharine Fry at UCLA prepared the document.

SUMMARY

Innovative performance assessment programs are a cornerstone of state education reform efforts nationwide, and the Maryland School Performance Assessment Program (MSPAP) is among the most prominent of these efforts. MSPAP, like many other performance assessment programs, is designed to induce fundamental changes in instruction as well as to measure the educational progress of schools, districts, and the state as a whole. In addition, MSPAP exemplifies the use of statewide assessments in an effort to raise standards. Maryland has established performance standards on MSPAP that are very high relative to the current distribution of achievement, and the Maryland State Department of Education (MSDE) reports every year the percentage of students in each assessed grade in each district who meet or exceed these standards. MSDE also reports a number of other outcomes, such as dropout rates, but performance on MSPAP appears to be the most important element of the state's accountability system.

The goals of MSPAP (and similar reforms) are ambitious, and the program's success is not guaranteed. MSPAP's effects will hinge on the behaviors of educators, and research has shown that the desired changes in practice are not easy to bring about.

The research reported here investigated the effects of MSPAP by surveying teachers and principals in two of the three grades in which MSPAP is administered. The sample included fifth-grade teachers, eighth-grade mathematics teachers, and principals whose schools included either fifth or eighth grade. All groups were surveyed in the spring of the 1994-95 school year, before the 1995 administration of MSPAP, using a computer-assisted telephone interview, and teachers were also administered a lengthy written survey. The surveys explored these educators' support for the MSPAP program; changes they made in school management and instruction in response to it; methods they used to prepare students for MSPAP; and reform-stimulated professional development. Most questions asked about educators' experiences up to the time of the interview, but a small number of questions asked specifically about the preceding (1993-94) school year. Interviews were completed with 112 principals and 224 teachers; mail surveys were collected from 186 teachers. A single sampled district refused to participate in the survey, but the participating districts nonetheless appear

reasonably representative of the state as a whole. (The findings reported here, however, should not be taken to generalize beyond the four populations from which we sampled: elementary- and middle-school principals, fifth-grade teachers, and eighth-grade mathematics teachers.)

Support for MSPAP

When asked a general question about support, roughly three-fourths of the principals and half of the teachers expressed support for MSPAP; 15% of principals and 35% of teachers expressed opposition. The responses to more specific questions were a complex mix of positive and negative views. In general, teachers were less likely than principals to express support for the program.

Support for MSPAP as an instrument of reform (in contrast to its role as an assessment) was reasonably widespread. Almost all principals reported that the program imposed more than a minor burden on their schools, but about two-thirds of the principals reported that the benefits of the program either balanced or outweighed the burdens. A sizable majority of teachers reported that the program has had a moderate amount or great deal of positive impact on instruction. Most principals reported that MSPAP has been useful for encouraging resistant teachers to change their instruction.

Support for the accountability uses of MSPAP was substantial among principals but quite limited among teachers. A majority of principals supported public reporting of MSPAP results and the use of MSPAP to hold schools accountable for meeting state standards, although only about a third of principals supported using MSPAP to identify schools for “reconstitution” (state takeover). Forty-two percent of teachers or fewer expressed support for any specific accountability uses of MSPAP about which the surveys asked.

Teachers’ views of the assessment *per se* were mixed. For example, about 60% strongly agreed that MSPAP assesses a broader range of skills than do multiple-choice tests, and majorities expressed agreement with the view that MSPAP provides accurate views of achievement in specific areas. About 40% of fifth-grade teachers, however, strongly agreed that MSPAP includes developmentally inappropriate tasks, and nearly half of all teachers surveyed strongly agreed that because of the nature of MSPAP, poor writing skills make it hard to judge some students’ achievement in mathematics. Although about half of all teachers and three-fourths of principals said that MSPAP results are at least

somewhat useful for making inferences about school improvement, large majorities also responded that comparisons among schools and estimates of progress may be distorted by irrelevant factors.

Professional Development

Almost all teachers reported some participation in MSPAP-related professional development, and almost all reported that the activities in which they had participated were somewhat or very useful. Also, all but one teacher responded that they felt prepared to help students get ready for MSPAP.

Over two-thirds of teachers participated in at least one activity designed to explain MSPAP or performance assessment in general, and nearly half participated directly in state-, district-, or school-level performance assessment (e.g., by developing or scoring tasks). Over half of the teachers participated in at least one activity providing professional development in MSPAP-related instructional methods and disciplinary content. Eighth-grade mathematics teachers participated in greater numbers than fifth-grade teachers in activities pertaining to general teaching strategies, which suggests that the practices encouraged by MSPAP may deviate more from their previous work.

Teachers had access to many possible vehicles for professional development, and they used those that were convenient and proximate. The vehicles accessed by three-fourths or more of teachers were print materials, collaboration with colleagues, courses offered by the school, and courses offered by the district. Activities sponsored by the state, professional associations, or colleges were typically attended by a small fraction of teachers.

Effects on School Management and Instruction

The surveys identified few changes in school management. Little change was reported in tracking or homogeneous grouping (except for a minor trend away from homogeneous grouping found in eighth-grade mathematics). About a third of principals reported reassigning teachers among grades to improve the relative quality of teaching in the assessed grades.

About three-fourths of principals reported encouraging teachers a great deal to raise expectations, but fewer teachers—only a bit over half—reported that expectations for students had actually increased in their schools. Moreover, while principals were equally likely to report encouraging increased expectations for low-

achieving, average, and high-achieving students, teachers were more likely to report increased expectations for high-achieving students. Over a third of teachers responded that expectations had increased a great deal for high-achieving students, but only 15% reported such an increase for low-achieving students.

Improving instruction is one of the foremost goals of MSPAP, and a large majority of principals and teachers reported that it has been somewhat successful in this regard. Most principals reported that MSPAP has been a useful tool for encouraging instructional changes by resistant teachers. Almost all teachers reported that MSPAP has had a positive effect on instruction in their schools, and about three-fourths reported that instruction had improved by more than a small amount. On the other hand, large majorities of teachers also reported that MSPAP has had at least a small negative effect on instruction in their schools as well.

Most teachers reported changing their own instruction in some manner in response to the MSPAP assessment. The majority of fifth-grade teachers reported shifting time among subject areas since MSPAP began. About two-thirds reported spending more time on writing, and about half spent more time on mathematics. The decreases in time allocations needed to accommodate these increases were scattered among a variety of subjects, with between 10% and 15% reporting decreases in reading, art, music, or physical education. Reported instructional changes within subject areas were largely consonant with the goals of the reform. Majorities of teachers reported increases in writing for a variety of purposes, analysis of text, literary comprehension, mathematical communication, and meaningful problem solving. Between a fourth and a third reported a decrease in emphasis on spelling, punctuation, and grammar; organized play; or free time. Nearly a half reported a decrease in emphasis on number facts and computation. Majorities of eighth-grade mathematics teachers reported increases in data analysis, communicating mathematics, and the use of graphs and tables; about a third reported a decrease in emphasis on computation and algorithms. When asked to describe positive changes, fifth-grade teachers were most likely to comment on the emphases on writing, explaining, and thinking skills; eighth-grade mathematics teachers were most likely to comment on the increase in cooperative work.

Most principals obtained instructional materials aligned with MSPAP and reported that their schools' emphasis on such material had increased. A bit under

half of the teachers reported focusing a great deal on aligning their instruction with the content of MSPAP, and more than a third reported focusing a great deal on the use of MSPAP-like tasks in their instruction. Very few principals reported an offsetting decrease in emphasis on untested material, but many teachers did. About two-thirds of teachers expressed some agreement with the statement that MSPAP has caused teachers to de-emphasize untested material, and 20% to 30% expressed strong agreement.

Test Preparation

Educators reported relying on a wide variety of approaches to preparing students for MSPAP. Broad instructional changes were among the most widespread approaches; for example, the great majority of principals reported encouraging their teachers a great deal to increase their emphasis on higher order thinking skills, and a majority of teachers (61%) reported focusing a great deal on “improving instruction generally.” Somewhat fewer teachers (roughly 40%) reported focusing a great deal on efforts to increase the alignment of their instruction with MSPAP. Few principals reported that untested material had been de-emphasized as a result of MSPAP, but about two-thirds of teachers reported that it had been.

Many educators also relied substantially on “direct test preparation”—that is, the use of practice tests and instruction in test preparation. About three-fourths of principals reported that they encouraged their teachers a great deal to use released MSPAP tasks and other test-preparation materials, and almost all principals reported offering some encouragement to do so. Most teachers (86%) reported that they focused more than a small amount on practice tests, but fewer than half reported focusing a great deal on them. Teachers reported allocating widely varying amounts of time to direct test preparation. For example, the median fifth-grade teacher reported allocating 5 hours to the use of MSPAP tasks (the most frequently used of the five types of practice tests about which we asked), but 25% of teachers reported using them for 2 hours or less, while another 25% reported using them for 9.3 hours or more. Eighth-grade mathematics teachers reported allocating less time than fifth-grade teachers to test preparation, but if other middle-school teachers behaved similarly, the total time spent on test preparation is substantially greater in eighth grade than in fifth. Summing across the five types of practice tests about which we asked, the

median eighth-grade mathematics teacher reported spending roughly 8% of instructional time on these activities, and 25% of teachers reported spending 12% or more. These estimates may overestimate total time spent on these activities, but they do suggest that many teachers are allocating quite substantial amounts of time to them.

As discussed below, the implications of these test-preparation activities are unclear. Some amount of practice with similar tasks is consonant with the goals of the program because the assessment is intended to model improved instruction. Excessive amounts, however, or practice that is too narrowly focused, could lead to inflated scores—that is, score gains that are not accompanied by commensurate increases in students' skills and knowledge.

Educators' Explanations of Score Gains

Educators' explanations of the MSPAP gains in their schools suggest the possibility of score inflation. More than half of teachers (57%) reported that work with practice tests and test-preparation materials contributed a great deal to their schools' gains, and 45% said that familiarity with the assessment had contributed greatly. In contrast, only 15% said that broad improvements in knowledge and skills had contributed a great deal to their MSPAP gains. Perhaps more important, only 20% said that improvements in the knowledge and skills emphasized by MSPAP contributed greatly. Although principals' explanations of gains were markedly more optimistic, even they were twice as likely to point to practice tests or familiarity than to broad improvements in knowledge and skills as having contributed a great deal to increased scores.

Implications

The mix of favorable and unfavorable responses educators provided to these surveys is not surprising. The Maryland reforms are both ambitious—calling for fundamental changes in teaching and large gains in student performance—and young. Time will be required for the program to evolve and for educators to learn how best to respond to it. Nonetheless, the results of these surveys point to issues that could be addressed to improve the program's effectiveness. The findings also highlight the importance of additional research and monitoring of the program.

The results of these surveys raise a concern about insufficiently specific curriculum frameworks, but additional investigation would be required to confirm

whether this is a substantial problem. On the one hand, many reformers eschew excessive specificity of curriculum frameworks in order to focus attention on broader outcomes and to avoid overemphasis on discrete, decontextualized skills. Accordingly, some degree of dissatisfaction among teachers with the specificity of frameworks may be expected and even desirable. However, insufficiently clear frameworks might lead teachers to focus excessively on the assessment *per se* as a surrogate for a curriculum framework, potentially undermining the instructional goals of the program and increasing the risk of inappropriate test preparation.

Teachers' reports of changes in expectations for students point to potentially negative effects of the program on equity. Many education reformers stress the importance of raising expectations for (and the performance of) low-achieving students, but more Maryland teachers report an increase in expectations for high-achieving students than for low-achieving students. Expectations are but one aspect of equity, however, and further investigation is needed to determine whether these perceptions are mirrored in actual practice and what causes them. Modifications of the program—for example, to give schools more public credit for improving the performance of students who nonetheless fall below Maryland's lowest performance standard—may be warranted to induce a more favorable effect on equity.

The results reported here suggest that the program has met one of its goals in increasing the amount of writing students do in school. At the same time, teachers' responses suggest the possibility that this change may have negative ramifications as well, in terms of both instructional impact and test validity. Many teachers maintain that the emphasis on writing is excessive and that instruction has suffered because of the amount of time required for writing. In addition, virtually all teachers reported that MSPAP's emphasis on writing makes it difficult to judge the mathematical competence of some students. Both of these concerns could be addressed by further research, but they also require judgments by policy makers—for example, judgments about the relative value of instruction forgone to accommodate greater writing, about the relative importance of communication compared to other aspects of mathematical competence, and about trade-offs between instructional effects and test validity.

The potential use of MSPAP to provide scores for individual students is currently the focus of debate in Maryland. These surveys confirmed that almost all educators would find the assessment more useful if such scores were provided.

Few principals would be willing to add testing time, however, to obtain higher quality student scores, while most teachers and middle-school principals (but only half of the elementary principals) would accept a change in the assessment's mix of task formats to this end. These findings suggest that one option would be to experiment with changes in format mix while monitoring educators' responses to determine whether instructional incentives are degraded.

A variety of the results of these surveys raise the critically important issue of potentially inflated gains on MSPAP. These findings include the substantial amount of reported test preparation, the reports of a majority of educators that some schools have been able to raise scores without improving education, and the fact that far more educators (especially teachers) attributed the score gains in their own schools to familiarity and test preparation than to improvements in knowledge and skills. The potential for inflated scores when standardized multiple-choice tests are used for accountability is now widely accepted, and there are reasons to expect that similar problems can arise in performance assessment programs as well. At the same time, the findings of these surveys, while cause for concern, are not sufficient to confirm the existence of serious problems. For example, given the novelty of MSPAP's formats, some increase in scores from familiarity could represent an increase in the validity of scores even if it did not reflect a commensurate increase in skills and knowledge. Moreover, teachers' perceptions of the causes of gains may not be fully accurate. Nonetheless, given the importance of the issue, the striking responses of educators on this topic suggest the need for further investigation of potential score inflation and its correlates.

Finally, responses to a variety of questions in these surveys point to the need for additional investigations of the validity of the MSPAP assessment. Validation requires a variety of different types of evidence, particularly in the case of assessments, such as MSPAP, that employ innovative formats, have multiple purposes, and are intended to alter educational practice. The task of accumulating evidence pertaining to validity should be ongoing.

In all, these surveys suggest that Maryland's reforms are meeting with important initial successes but are also encountering substantial difficulties, at least in the perception of educators. The state now has the opportunity to use data from these and other studies to guide midcourse corrections to improve the program's impact and minimize its negative effects.

**FINAL REPORT: PERCEIVED EFFECTS OF THE MARYLAND SCHOOL
PERFORMANCE ASSESSMENT PROGRAM**

**Daniel Koretz, Karen Mitchell, Sheila Barron, and Sarah Keith
CRESST/RAND**

BACKGROUND AND RESEARCH OBJECTIVES

A cornerstone of the current education reform movement, manifested in initiatives in many states and districts across the nation, is innovative assessment systems intended to encourage good classroom practice, help faculty focus on standards for student performance, and provide data pertaining to school improvement. Proponents of assessment-based reform claim that challenging, authentic assessments will better address higher order thinking skills, communication facility, and conceptual understanding (e.g., Resnick & Resnick, 1992; Stiggins, 1991). They also believe that systems that use these assessments for accountability will encourage high levels of student performance while avoiding the negative effects associated with accountability systems based on multiple-choice tests, such as narrowing of the curriculum and fragmentation of content (Maryland School Performance Program Office, 1994; Wiggins, 1989; see Koretz, Linn, Dunbar, & Shepard, 1991; Smith, 1991; and Smith & Rottenberg, 1991, on the impact of conventional, test-based accountability systems).

One of the most prominent of these assessment-based reforms is the Maryland School Performance Program (MSPP), sparked by a 1989 report that set ambitious goals for student achievement and school performance. MSPP is a school accountability program, the cornerstone of which is the Maryland School Performance Assessment Program, or MSPAP. MSPAP consists of assessments in reading, mathematics, science, social studies, writing, and language administered in Grades 3, 5, and 8. It is performance-based, focuses on communication, higher order thinking skills, and problem solving, and entails both multidisciplinary tasks and group work. MSPP reports data on a number of other measures as well, such as promotion rates, dropout rates, and attendance rates, but MSPAP scores are the most important element of the system. Schools were originally charged with meeting specified performance standards by 1996; the

target date has since been changed to 2000. They are put on alert if their performance is poor, and in extreme cases, they can be “reconstituted” (i.e., taken over by the state). MSPAP relies heavily on public awareness of the results to influence performance; the state publishes annual reports of MSPP results at the district level, and districts are required to publish comparable data at the school level.

The success of assessment-based reform, however, depends on the responses of educators. Research on assessment-based reform suggests that school change is difficult to achieve and that changes in teachers’ practices—and the resulting changes in student work—are not easily brought about (Druker & Shavelson, 1995). For example, early research on the reading component of the state reform in Maryland (Afflerbach, Guthrie, Schafer, & Almasi, 1994; Guthrie, Schafer, Afflerbach, & Almasi, 1994) suggested that change was impeded by a lack of alignment between teachers’ beliefs and practices and those implicit in the program, a lack of alignment between existing and mandated instruction and performance assessment, a lack of resources to help implement change, and insufficient communication from the jurisdiction about program mandates. These and other studies indicate that assessment-based reform is influenced by a number of complex school and classroom variables, including incentive systems, local beliefs and norms, financial resources, and available materials and support services (see also McLaughlin, 1990).

The research reported here investigated the impact of MSPP and MSPAP by surveying elementary- and middle-school principals, fifth-grade teachers, and eight-grade mathematics teachers. The surveys explored these educators’ support for the program, changes they made to school organization and management in response to the program, changes in classroom instruction, methods used to prepare students for MSPAP, and reform-stimulated teacher professional development. The results of this research should be of interest to participants and stakeholders in Maryland schools and to educators and policy makers in other states contemplating school reform based on innovative performance assessment programs.

PROCEDURES

This report summarizes the results from mail and telephone surveys that were administered during the 1994-95 school year to assess Maryland principals' and teachers' opinions of and responses to MSPAP. A total of 112 principals and 226 teachers (both split approximately evenly between the fifth and eighth grades) were interviewed.

Sampling

A multistage design was employed for sampling. In the first stage, districts were chosen using both purposeful and random sampling. Because difficulties were encountered in obtaining data from MSDE, it was impractical to sample schools from the state's 24 districts. Instead, four districts judged important due to their size and population characteristics were selected, along with a random sample of five of the remaining districts. One selected district elected not to participate in the study. Within each participating district, a random sample of elementary and middle schools was chosen. Schools with fewer than ten students in the target grade (fifth or eighth) assessed using MSPAP were excluded. In the large districts, stratification on percent free lunch (as a proxy for school-level SES) was used to ensure that representative numbers of low-SES schools were selected. Ninety-one elementary schools and 88 middle schools were sampled.

In the spring of 1995, the principals of the selected schools were contacted and asked to participate in the study and to provide us with the names of teachers for our teacher survey. Because the survey focused in part on changes in practice over time, we only attempted to interview principals who had served as an administrator in a Maryland school for at least four years (since 1991-92). Approximately 20% of the principals in the sample were too new to administration to satisfy this eligibility screen. Interviews were conducted with 112 (78%) of the eligible principals.

All sampled principals, regardless of their own eligibility, were asked to provide names of teachers for our survey of teachers, and 80% of the total sample of principals (including ineligible principals) provided this information. Principals selected for the elementary-school sample were asked to provide the names of teachers who were then teaching fifth-grade students. In schools with three or fewer fifth-grade teachers, all teachers were contacted and asked to participate in the study. In schools with more than three fifth-grade teachers, only a random

sample of three teachers was contacted and asked to participate in the study. A total of 192 fifth-grade teachers were selected for inclusion in the study.

Principals selected for the middle-school sample were asked to provide names of teachers who were then teaching mathematics to eighth-grade students. All eighth-grade mathematics teachers in sampled schools were contacted and asked to participate in the study. A total of 148 eighth-grade mathematics teachers were selected for inclusion in the study.

Because of our focus on changes in practice, we also attempted to interview only teachers who had taught the relevant grade (and subject area in the case of eighth grade) in at least one of the previous three years (1991-92, 1992-93, or 1993-94). Twenty-two percent of the fifth-grade teachers contacted and 16% of eighth-grade mathematics teachers did not satisfy this eligibility screen. Of the eligible teachers, 113 (76%) and 111 (75%) were interviewed from the fifth and eighth grades, respectively.¹

Data Collection

Data were collected using both a computer-assisted telephone interview (CATI) and written surveys. Both the surveys and interviews were developed at RAND by project staff and piloted with Maryland educators. CATI was used to collect data from principals; both CATI and the written surveys were used to collect data from teachers. We attempted to get both interview and written survey data from all of the eligible teachers; however, for some teachers we have data from only one of the two sources. Table 1 presents the participation rates by group and survey medium.

Table 1
Sample Sizes and Response Rates for Maryland
Teachers by Survey Medium

	Fifth grade	Eighth grade mathematics
CATI	113 (76%)	111 (75%)
Mail survey	89 (59%)	97 (64%)
Both	82 (55%)	90 (61%)

¹ To be conservative in estimating participation rates, we assumed that teachers we failed to contact were eligible and refused to participate. If some of these teachers were in fact ineligible, our true participation rates would be higher.

Principal interviews were designed to collect information about school demographics, general support for the reform effort, principal's own responses to the reform (including her or his role as an instructional leader), and effects of the reform effort on the school, its teachers, and its students. Questions were also asked about how the information provided by MSPAP was used in the school and the burdens imposed by the program.

The teacher interviews asked some questions that mirrored questions asked of principals (e.g., questions about general support) as well as questions about test preparation practices, classroom practices, and understanding of the MSPAP program. The written surveys asked teachers questions concerning their opinions of the usefulness and reasonableness of MSPAP results for specific purposes, school climate, and teacher professional development. Also, additional questions were asked on the written surveys about preparing students for MSPAP.

Most questions in all of the instruments asked about educators' opinions at the time of the survey or about their experiences with MSPAP up to that time. However, a small number of questions focused specifically on the previous school year (1993-94). For example, questions about test-administration practices had to focus on the 1993-94 school year, because MSPAP had not been administered in 1995 at the time our surveys.

The vast majority of questions asked were presented in a closed format (e.g., 3-, 4-, and 5-point Likert scales) or required the respondent to answer with a single number. An effort was made to balance the questions, in that some implied a positive view of the program, while others implied a negative view or raised possible criticisms. (Many were neutral.) This balance was particularly important in the case of the relatively few questions that asked respondents to express the strength of their agreement or disagreement with assertions, such as "all students can learn to a high level" and "the emphasis on high standards for all students is putting undue pressure on schools and students."

Data Analysis

Because the sampling procedure was complex and exploratory analyses suggested that differential sampling rates across districts might affect the results, weighting was used to offset differential sampling. The weight assigned to each case was the product of the inverses of the probabilities that the principal or teacher would be selected at each stage of the sampling process and the

probability that sampled individuals would participate (complete a survey).² Weights were assigned by assuming that the eligibility rates in the sample were equal to the eligibility rates in the population. The sum of weights was constrained to be equal in the two grades rather than to be proportional to the numbers of principals and teachers in each grade statewide.

Weighted frequencies for the Likert questions were calculated overall for the two grades combined and separately by grade level. Questions requiring a numerical response were summarized using weighted means, standard deviations, medians and other selected quantiles.

The CATI interviews included a small number of open-ended questions asking principals and teachers to provide more involved answers. Respondents' answers to these questions were audio-taped whenever the respondent gave his or her permission for taping.³ These responses were coded from the tapes after the survey was completed. This permitted the use of a substantial number of responses in devising a coding scheme.

Because this study was designed to provide a description of educators' perceptions of the implementation and impact of the MSPAP program, we were interested in collecting a wide variety of information from principals and teachers. Accordingly, we decided to devote resources to increasing the breadth of data collected at the expense of sample size and statistical power. Because of the large amount of information collected from each principal and teacher and the relatively small size of the samples, we have taken a descriptive rather than statistical approach in this report.

Although we chose not to conduct formal statistical tests, we were guided in our presentation of results by the degree of confidence we believe can be placed in specific results. For example, we were relatively confident in the estimates of the percent of principals or teachers reporting a specific view or practice. Confidence intervals for simple random samples can provide a rough guide to the level of confidence one can have in estimates from these surveys. For example, the 95% confidence intervals for proportions of .60 (or .40) would be $\pm .098$ for a simple random sample of 100 and $\pm .069$ for a simple random sample of 200. The

² Very large weights were trimmed to avoid inflating error variance and giving excessive weight to any single case.

³ In the small number of cases in which the respondent was unwilling to be taped, the interviewer typed in the respondent's comments.

corresponding error bands for a proportion of .20 (or .80) would be smaller: $\pm .080$ for a sample of 100, and $\pm .057$ for a sample of 200. These examples are typical of the results we reported.⁴ Comparisons across groups (e.g., the percent of elementary principals as compared to the percent of middle-school principals) generally have larger margins of error, and we therefore present such differences only when they are large or for other reasons suggestive.

There are, of course, other threats to the validity of the conclusions we reach in this study. Not all educators that were sampled chose to participate, and this may introduce some degree of bias into the reported results. In addition, as in all surveys, respondents may show a tendency to shade their responses in the direction they feel is socially desirable, and some results therefore may be less than accurate reflections of educators' true beliefs. This "social-desirability bias" could include a tendency to fail to report or underreport activities that are considered of questionable propriety.

Generalizability of Findings

The resulting weighted teacher samples are representative of fifth-grade teachers and eighth-grade mathematics teachers with sufficient experience statewide, with the exception of the one district in which we were not permitted to interview. The resulting weighted principal samples are representative of elementary- and middle-school principals, with the same restrictions. For simplicity, we sometimes refer to these groups together as "educators," but this does not imply that findings would have been similar for other groups of Maryland educators.

The performance of our sampled counties on MSPAP, after weighting for our differential sampling rates, was quite similar to that of the state as a whole despite the loss of one sampled county. Our weighted sample very closely matched the state (within 2 percentage points) in terms of the percentage of students reaching the satisfactory level on MSPAP in fifth-grade mathematics and eighth-grade reading. Our sample scored somewhat lower in fifth-grade reading (25% satisfactory, versus 30% in the state as a whole) but somewhat higher in eighth-grade mathematics (44% satisfactory, versus 40% in the state as

⁴ Because of the complex sample design, confidence intervals are not straightforward to estimate. The clustering in our sample would increase the margin of error relative to that in a simple random sample; a finite sample correction would reduce it modestly.

a whole). Our sample appeared to have a somewhat higher incidence of poverty, based on the rather weak proxy of receipt of free or reduced-price school lunches, despite our having stratified on that variable in our three largest districts. MSDE reported a statewide rate of 29%; we found 29% in our weighted middle-school sample but 37% in our weighted elementary-school sample.

Because the sum of weights was equal in the two grades, both grades counted equally when we combined results across grades. Had we instead weighted each grade to reflect its numbers in the population of teachers, fifth-grade teachers would have counted far more in our results, even though they teach essentially the same number of students as do the eighth-grade mathematics teachers. However, in most instances, results were combined only when they were similar in the two grades, so weighting proportionally would not have greatly changed the results reported here.

SUPPORT FOR MSPAP

We asked educators about their support for MSPAP, both overall and with respect to each of the three functions MSPAP serves: inducing reform, monitoring school performance, and providing the basis for accountability. Additionally, we asked principals and teachers about their use of information from MSPAP in drawing conclusions about student achievement and educational effectiveness.

When asked global questions about support for MSPAP, about three-fourths of Maryland principals expressed support for the program. Fewer teachers than principals expressed positive views—a pattern that appeared frequently when we asked principals and teachers similar questions about support for the program. Just over half of the teachers said they somewhat or strongly support the program. A sizable majority of teachers, however, said they believe the program has had a moderate or great deal of positive effect on instruction.

Educators' responses to questions about more specific aspects of MSPAP revealed a complex mix of opinions; both negative and positive views were reported by substantial majorities of respondents. These mixed opinions appeared in response to questions about the quality of MSPAP as an assessment, a number of issues of administration and scoring, and factors that potentially distort MSPAP comparisons among schools and over time.

Support for the accountability functions of the assessment was inconsistent but, in many instances, lower than support for MSPAP's other functions. A sizable majority of principals supported the public reporting of MSPAP scores, and over half said they support MSPAP's use in holding schools accountable to state performance standards. Considerably fewer said they support the role of the assessment in identifying schools for reconstitution. Fewer teachers than principals supported any of the accountability uses about which we asked.

Overall Support for the Program

While the majority of educators in our samples said they support the state's performance-based assessment program, more principals than teachers expressed this view. About three-quarters of the principals but just over half of the teachers responded that they support MSPAP. The three-fourths of principals who supported MSPAP were about evenly split between those who characterized themselves as strong supporters and those who were "somewhat supportive." Seventy-two percent of principals said their attitude toward MSPAP had become more positive over the prior few years. Fewer teachers than principals were supportive of the program. Fifty-six percent of the teachers said they support MSPAP, while 35% said they oppose it.

When asked about the burden MSPAP imposes on their school, over 90% of principals said that MSPAP imposed more than minor burden (reporting a "moderate" or "great" burden). By far the most frequently stated reason principals found MSPAP burdensome was the time demands it imposed; 96% of principals gave time demands as a burden. Management and record-keeping (71%) and the need for staff retraining (68%) also were cited as reasons that many respondents found the program burdensome. Staff stress or low morale was given as a reason by 63% of principals. Other program burdens include unclear achievement targets (49%), difficulty motivating staff (41%), and the need for rapid instructional change (38%).

On the other hand, 46% of the principals who perceived the program as burdensome reported that the program's benefits as a tool for improving the quality of instruction in their school are greater than the burdens it imposes, and an additional 22% said the benefits and burdens balance. Moreover, 81% of principals said the program has become easier to accommodate in their school in the several years the program has been in place.

Many principals and teachers agreed that the emphasis on high standards for all students is putting “undue pressure” on schools and students. More teachers (66%) than principals reported this view. About half of the elementary principals (47%) and a third of the middle-school principals (31%) somewhat or strongly agreed that the emphasis on high standards for all students is putting undue pressure on schools and students.

Greater numbers of teachers reported undue pressure when we moved from general statements about the impact of high standards on schools to questions about the pressure they feel as teachers to improve student performance on MSPAP. Almost all agreed (43% strongly and 45% somewhat agreed) that teachers are under undue pressure to improve student performance on MSPAP.

Educators’ Judgments About MSPAP as a Lever for Reform

The sections below discuss educators’ judgments about three functions served by MSPAP: serving as an agent of reform, as a tool for monitoring student achievement and school effectiveness, and as an accountability instrument.

MSPAP and MSPP are based on a number of fundamental beliefs about student performance and educational opportunity. Like school reformers elsewhere, MSPAP’s architects assert that all students can learn to high levels and that the job of Maryland educators is to provide the means by which that happens. We asked teachers whether they agreed with some of the premises of the program.

Teachers expressed mixed opinions about some of the fundamental tenets of the program. Sixty percent of teachers said they somewhat or strongly agree with the notion that all students can learn to high levels. Even more (85%) agreed that this is the right message to give to Maryland students, regardless of whether it is possible for all students to learn to high levels. The majority of teachers, however, still expect variation in student performance; only 21% of teachers agreed with the statement “All students can learn to the *same* high level.” Similarly, almost all (88%) somewhat or strongly agreed with the statement that for some students the lowest MSPAP proficiency level is a high level of performance.

Few principals (8%) agreed that the goal of meeting MSPAP standards by 1996 (the original program goal) was a realistic goal for their schools. It should be noted that the schedule for meeting standards has since been revised: The current

expectation is that schools will meet performance standards on MSPAP, attendance rates, promotion rates, and other outcomes by the year 2000.

Most principals and teachers were positive about MSPAP's value as an agent of reform. Eighty-four percent of principals said that in response to MSPAP they have encouraged their teachers a great deal to focus on improving instruction generally. Eighty-one percent of principals said MSPAP has been a useful tool for encouraging positive instructional change among teachers who are very resistant to making changes to their instruction; 44% said that it has been very useful in this respect. About half of the teachers agreed that MSPAP has caused some teachers who are resistant to change to improve their instruction. Most teachers reported that MSPAP has had positive effects on instruction in their schools. Eighty-three percent of fifth-grade teachers and 63% of eighth-grade mathematics teachers said MSPAP has had a moderate or great deal of positive effect on instruction in their school; almost all of the remainder said MSPAP has had a small amount of positive effect. (For more discussion of perceived positive and negative effects of MSPAP on instruction, see the section below, Impact of MSPAP on School Management and Instruction.)

Educators' Judgments About MSPAP as a Tool for Monitoring School Performance

The surveys probed educators' views of three aspects of the adequacy of MSPAP as a measurement tool: the adequacy of the test itself (e.g., the domain it assesses and the way it is administered and scored); the accuracy of the information it yields about student performance; and its accuracy as a measure of school effectiveness.

Support for the Testing Domain, Administration, and Scoring

Despite generally positive views of the role of MSPAP in instructional improvement, teachers' opinions about the adequacy of the MSPAP assessment were clearly mixed: Large majorities offered both positive and negative opinions. On the positive side, teachers were nearly unanimous (93%) in their agreement that MSPAP tests a wider range of skills than do multiple-choice tests. Sixty-three percent of fifth-grade teachers but only 45% of eighth-grade mathematics teachers agreed that MSPAP more closely resembles what they teach than do standardized tests. Fifty-six percent of teachers somewhat or strongly agreed that

tasks reflect essential concepts and content, and 65% said assessment tasks are based on realistic situations.

However, many teachers reported negative opinions concerning MSPAP's content domain, administration, and scoring. Eighty-six percent of fifth-grade teachers and 61% of eighth-grade mathematics teachers believe the assessment materials and testing requirements are developmentally inappropriate for students in their grade. Over three-quarters of teachers (77%) agreed that the curriculum content for the assessment is not defined well enough for them to adequately prepare their students, and 61% strongly or somewhat disagreed that the content of the assessment is consistent across years. About three-quarters (74%) said they believe scoring standards are not consistent across years and disciplines. Fifty-eight percent somewhat or strongly agreed that good writing skills make it appear that some students know more mathematics than they do; conversely, virtually all teachers agreed that poor writing skills make it hard to judge some students' mathematics achievement. Sixty-four percent of fifth-grade teachers and half of eighth-grade mathematics teachers agreed that testing times for some MSPAP tasks are too short for students to show what they have learned. More (72%) said the use of group tasks places a heavy burden on students who are more able or inclined to do the work.

A somewhat clearer pattern emerges if one considers only those teachers who reported strong opinions (i.e., who responded that they "strongly agree" with statements in the survey). The most frequently stated strong agreement was with the view that MSPAP tests a wider range of skills than multiple-choice tests; this was expressed by 61% of the teachers. Almost half of the teachers (45%) strongly agreed that poor writing makes it hard to judge some students' mathematics achievement using MSPAP. Other concerns about the assessment were voiced by fewer teachers overall, and by fewer eighth-grade mathematics teachers than fifth-grade teachers. For example, 41% of fifth-grade teachers but only 18% of eighth-grade mathematics teachers strongly agreed that MSPAP assessment materials and testing requirements are developmentally inappropriate. About a third of fifth-grade teachers (32%) and about a fourth of eighth-grade mathematics teachers (23%) strongly agreed that testing times for some MSPAP tasks are too short. Half of the fifth-grade teachers but only a third of the eighth-grade mathematics teachers strongly agreed that group tasks place a heavy burden on students able or inclined to do the work.

Perceived Accuracy of Student Achievement Information

Notwithstanding these concerns about factors that could distort the information MSPAP provides about student performance, the majority of teachers expressed positive opinions about the accuracy of student achievement information provided by MSPAP. Our surveys asked fifth-grade teachers about the accuracy of the student performance information on learning outcomes in reading, mathematics, social studies, science, writing, and language usage; eighth-grade mathematics teachers were questioned about individual process outcomes (problem solving, communication, reasoning, and so on) and content outcomes (estimation, geometry, measurement, etc.).

Across the Maryland learning outcomes, between 51% and 72% of teachers said that the information MSPAP provides about student achievement is somewhat or very accurate. For example, 65% of fifth-grade teachers said that MSPAP information accurately portrays students' achievement in writing and language usage, and just over half said the same about reading, mathematics, social studies, and science.

The judgments of eighth-grade mathematics teachers were similar. Sixty percent of eighth-grade mathematics teachers said that MSPAP provides somewhat or very accurate information about students' achievement on the mathematics process outcomes; over 50% responded similarly about the mathematics content outcomes. Somewhat more teachers said MSPAP provides accurate information about problem solving (72%) and communication (69%) than about the reasoning (58%) and connections outcomes (54%). Among mathematics content, teachers said MSPAP provides somewhat or very accurate information about students' achievements in arithmetic operations (69%), number relations (72%), statistics (67%), probability (67%), estimation (57%), geometry (61%), measurement (65%), and patterns/relationships (60%).

Elementary- and middle-school principals' judgments about the usefulness of MSPAP data on student performance were similarly positive. Over three-quarters of principals (77%) said MSPAP provides them and other school principals with somewhat or very useful information about student performance; an additional 15% said the information is slightly useful.

The surveys did not ask comparable questions about the accuracy of student achievement information provided by standardized multiple-choice tests. We did,

however, ask about the reasonableness of MSPAP and standardized test data for assessing school performance; these responses are described next.

Perceived Accuracy of School Effectiveness Data

MSPAP and other Maryland School Performance Program data are used to draw conclusions about the effectiveness of educational programs. We asked both principals and teachers numerous questions specifically about this use, because a test can provide good information about student performance even if it does not provide a valid basis for conclusions about educational effectiveness. For example, comparisons among schools could reflect differences in students' backgrounds rather than differences in educational effectiveness. The survey asked educators for their opinions of the reasonableness of MSPAP data for drawing inferences about school effectiveness and asked them to compare it in this respect to traditional, standardized, multiple-choice achievement tests. Both teachers and principals were also asked to respond to questions about a number of specific factors that might distort comparisons among schools or measures of change over time. Although the survey did not explicitly point this out to respondents, the factors about which we asked have little to do with the format of MSPAP; most would apply similarly to traditional multiple-choice tests if they were used in the same way.

Principals' statements about the value of MSPAP information in judging school performance echoed their opinions about MSPAP's utility as a lever for reform. Recall that 76% of Maryland principals voiced support for the MSPAP program and 84% said they have encouraged their teachers a great deal to focus on improving instruction generally in response to MSPAP. Most principals (70%) said that MSPAP data are somewhat or very reasonable for making inferences about school improvement. Sixty-nine percent of elementary principals but only 46% of the middle-school principals agreed that MSPAP provides a better view of school effectiveness than do standardized tests.

Principals' positive views of the usefulness of MSPAP data for measuring educational effectiveness, however, were somewhat inconsistent with their comments about specific limitations of the data for this purpose. Sixty-seven percent agreed that some schools have found ways to raise MSPAP scores without improving education. Moreover, 93% of principals said that score differences between schools often reflect students' characteristics more than

school effectiveness. About as many principals (88%) somewhat or strongly agreed that score changes from year to year often reflect differences in the characteristics of students taking the test rather than changes in school effectiveness. Ninety-two percent agreed that schools with highly transient populations are at an unfair disadvantage.

Teachers' global judgments of the accuracy of MSPAP data for monitoring school performance were less positive than those of their principals. Half of the teachers disagreed with the statement that MSPAP is sensitive to most important improvements in education. Fifty-six percent of teachers—compared to 70% of principals—said the test provides a somewhat or very reasonable basis for drawing conclusions about the effectiveness of schools' educational programs. In contrast, most eighth-grade mathematics teachers (86%) responded that the Maryland Functional Test (a fixed-response minimum competency test) is a reasonable basis for judging the effectiveness of schools' programs. Just a third of the teachers agreed that MSPAP provides a better view of school effectiveness than do standardized tests in general.

Like their principals, a large majority of teachers responded that other factors could distort differences across schools and over time in monitoring school performance. Most (73%) responded that some schools have found ways to raise MSPAP scores without really improving education. Almost all somewhat or strongly agreed that student characteristics (94%) and high mobility (96%) could distort school comparisons. Ninety-one percent of teachers said differences in student motivation could distort comparisons across schools or years. In addition, 67% agreed that differences in test administration could distort comparisons. Thirty percent of the teachers agreed that school comparisons are distorted by the retention of students in non-MSPAP grades to improve school scores.

Some of the teachers' concerns about the use of MSPAP data in judging school effectiveness are lessened by their combination with other information about student performance. The *Maryland School Performance Program: District and School Reports* display data on MSPAP and other student performance indicators (attendance, promotion, graduation, etc.). Ninety-five percent of the teachers somewhat or strongly agreed that combining MSPAP with other student performance and supporting information provides a better indication of school quality than MSPAP scores alone. When asked about individual data elements, half of the teachers said attendance rates are somewhat or very reasonable

indicators of educational effectiveness. The same percentage of fifth-grade teachers responded similarly about promotion rates.

The vast majority of principals (84%) claimed they and other administrators would find the information provided by individual student scores on MSPAP to be somewhat or very useful (5% said it would be slightly useful). Further, 94% of teachers said that MSPAP would be somewhat or very much more useful if it provided scores for individual students. One way to obtain more reliable scores for individual students would be to allocate more time to the assessment. Most principals, however, were willing to trade little or no extra testing time to obtain MSPAP scores for individual students. Eighty-three percent of elementary principals and 57% of middle-school principals said they were unwilling to schedule any additional testing time for this purpose. Some principals (18%) said they would be willing to schedule an additional hour of testing; very few (9%) said they'd be willing to add up to four hours in exchange for individual scores. More middle-school than elementary principals were willing to extend testing time.

Another way to obtain more reliable MSPAP scores for individual students would be to change the mix of item types—for example, by adding some multiple-choice items to the current battery. Most of the middle-school principals (85%) and teachers at both levels (81%) but only 49% of the elementary principals said they would support changing the mix of formats in order to obtain scores for individuals.

Educators' Judgments About MSPAP as an Accountability Instrument

As noted, the Maryland School Performance Program calls for publication of MSPAP scores and attendance, promotion, and other school performance data, and the state maintains (but has rarely exercised) the prerogative to “reconstitute” schools that fall short of accountability targets—that is, to seize control of them.

While principals' support for public reporting of results was almost as widespread as positive views of the data's usefulness to them, slightly fewer said they supported MSPAP's use in holding schools accountable to state standards. Sixty-five percent of principals said they somewhat or strongly support the publication of MSPAP and other school data. Fifty-seven percent said they

support holding schools accountable for meeting performance standards on MSPAP and other MSPP data. Not surprisingly, the strongest accountability lever—identifying schools for reconstitution—had the fewest supporters: only 31% of the principals.

A smaller percentage of teachers than principals supported any of these accountability uses. Forty-two percent of teachers said they somewhat or strongly support public reporting of data, and the same percentage oppose it. Fewer supported the other accountability uses. Only 31% of teachers supported holding schools accountable to state standards, and only 14% supported the use of MSPAP and other school data in identifying schools for reconstitution. In both instances, the remaining teachers mostly held negative rather than neutral opinions: 61% opposed holding schools accountable to state standards, and 76% opposed using these data to identify schools for reconstitution. Similarly, 74% of teachers agreed that judging schools against the state’s performance standards unfairly rewards and punishes many teachers. About three-fourths of teachers said they oppose the use of MSPAP and other program data in identifying schools for state control.

TEACHER PROFESSIONAL DEVELOPMENT FOR MSPAP

The MSPAP assessment represents a dramatic shift in the state’s expectations for students and teachers. Teacher professional development in preparation for MSPAP therefore plays an essential role in the reform. Our survey asked fifth-grade teachers and eighth-grade mathematics teachers about the professional development activities in which they took part, their opinions of the usefulness of the activities, and the vehicles through which they obtained professional development.

In preparing for MSPAP, over 90% of responding fifth-grade teachers and eighth-grade mathematics teachers attended at least one workshop, and 98% made use of one or more printed resources. To aid their teachers in acquiring the necessary knowledge and skills for preparing students for MSPAP, 95% of principals responded that they obtained training for their staff, and the same number reported that they provided their teachers with release time for training.

Teachers appear to think that their participation in these activities has consequences for their students’ performance on MSPAP. At least 85% of

participating teachers rated each of the activities about which we inquired as somewhat or very useful. Additionally, 80% of teachers commented that differences in access to staff development opportunities affect school comparisons. Another, less direct indication of teachers' positive views of professional development is the fact that all but one teacher responded that he or she felt prepared to some degree to help students get ready for MSPAP, and over three-quarters of teachers reported feeling mostly or fully prepared.

Types of Professional Development Activities

We categorized professional development activities into two broad classifications, each with two categories:

Activities related to the assessment:

- understanding MSPAP and performance assessment
- working directly on MSPAP or other performance assessments—e.g., scoring

Activities related to instruction:

- teaching disciplinary material
- using general teaching strategies

We asked about numerous specific activities within each of these four categories. In the case of each category, more than half of the responding teachers attended at least one of the activities about which the survey asked.

Activities Related to the Assessment

Measurement-driven reform requires teachers to learn about the assessment by which the teachers and their students are judged. We divided professional development related to the assessment into two categories: informational activities that explained MSPAP and performance assessment to the teachers, and activities that involved teachers directly in MSPAP or other performance assessment programs. The former included courses and workshops introducing MSPAP, professional-association conference sessions about the assessment, and informal discussion and collaboration with peers about MSPAP and performance assessment. We asked Maryland teachers about three ways in which they worked on MSPAP or other (locally developed) performance assessments: developing tasks for the assessment, piloting tasks, or scoring assessments.

Over two-thirds of teachers attended at least one informational activity about MSPAP or performance assessment generally. With the exception of some state and regional conferences that drew relatively few teachers, other individual activities in this category generally drew between one-third and one-half of respondents, and usually a larger percentage of eighth-grade mathematics teachers than fifth-grade teachers. Over half attended an activity introducing them to MSPAP. Among the professional development activities concerning MSPAP and performance assessment, the most frequently cited (reported by 85% of the teachers) was informal discussion and collaboration with other teachers—a type of professional development often overlooked.

Nearly half of fifth-grade teachers and 70% of eighth-grade mathematics teachers participated directly in at least one aspect of performance assessment by developing, piloting, or scoring tasks for MSPAP or a local performance assessment (Table 2). Local activities usually drew about twice the percentage of teachers that participated in comparable MSPAP activities. Only two of the specific activities about which we asked drew more than a fourth of the teachers, and those only for eighth-grade mathematics teachers. Over 50% of eighth-grade mathematics teachers reported developing tasks for local performance assessments, and 35% participated in local-level piloting. Relatively few fifth-grade teachers participated in either of these activities. While the state relies on teachers at the same grade level as that being assessed to work on MSPAP, these

Table 2
Percent of Teachers Participating Directly in Testing Activities

Activity	Grade 5	Grade 8
State level		
Developing MSPAP tasks	10	24
Piloting MSPAP tasks	6	15
Scoring MSPAP	10	12
Local level (school or district)		
Developing performance assessment tasks	12	53
Piloting performance assessment tasks	11	35
Scoring performance assessments	18	25

results may suggest that some local districts are relying more heavily on middle-school subject area specialists to develop the tasks for their local performance assessments.

Among informational activities about MSPAP and performance assessments attended by at least 20% of teachers, fifth-grade teachers cited two more frequently than others as being very useful: training for scoring performance assessments (60%) and training pertaining to the development of performance assessments (55%). In contrast, none of these activities was rated as very useful by more than 30% of the eighth-grade mathematics teachers who attended them. Close to half of the teachers at both grade levels who said they work informally with other teachers rated it as very useful. The more general introductory sessions were rated as very useful by fewer than one-fourth of their participants.

Overall, direct participation in assessment was more frequently rated very useful than were informational sessions about MSPAP and performance assessment. Over half of the participants found each of the local activities of this type to be very useful. At the state level, between one-third and one-half of the participants rated the activities very useful.

Activities Related to Instruction

MSPAP encourages numerous changes in instruction: in the content teachers present, in the ways in which they teach it, and in the instructional strategies they use across subject areas. We separated instructional issues into two categories: (a) disciplinary content and discipline-specific instruction and (b) general teaching strategies. Courses and workshops about teaching a subject area (math, reading/language arts, science, or social studies) were categorized as disciplinary content and discipline-specific instruction. Activities that taught teachers about thematic instruction, cooperative learning, higher order thinking skills, or writing across the curriculum were categorized as general teaching strategies.

Over 50% of teachers participated in at least one activity about disciplinary content and instruction. With the exception of a state-sponsored conference with low attendance, between 45% and 55% of teachers in both grades attended each activity about which we asked. A similar proportion of fifth-grade teachers attended at least one activity related to general teaching strategies. Eighth-grade mathematics teachers, in contrast, were more likely to attend activities

pertaining to general teaching strategies: Over 80% attended at least one. Over three-quarters of the eighth-grade mathematics teachers took a course or workshop on cooperative learning.

About half of the participating teachers ranked the discipline-oriented activities as very useful, with one exception. Less than a quarter rated their activities on the teaching of reading/language arts as very useful. The general strategies topic rated very useful by the largest percentage of teachers—almost 70%—was writing about math, which was included only on the eighth-grade survey. (Fifth-grade teachers were asked a comparable question about writing across the curriculum.) All the other activities about general teaching strategies were rated as very useful by between one-third and one-half of the teachers.

In interpreting these participation rates, one could conclude that teachers of both fifth and eighth grade are working to change many areas of their classroom instruction, since in both instructional categories, more than 50% of the teachers reported participating in at least one activity. It appears that eighth-grade mathematics teachers are concentrating especially heavily in general teaching strategies, perhaps suggesting that strategies called for by MSPAP differ markedly from the previous norms of middle-school mathematics instruction.

Professional Development Vehicles

Teachers have numerous vehicles from which to obtain professional development in preparation for MSPAP: courses and workshops, state and regional conferences, print and video resources, and informal collaboration with colleagues. We asked about courses and workshops sponsored by schools, districts, colleges or universities; conferences sponsored by national, state or regional groups; and print and video resources produced by national, state, regional, or local groups.

Print materials were the vehicle accessed by the largest percentage of teachers, with 98% using at least one of the many printed items about which the surveys asked. Teacher collaboration was also used by most teachers: 85% of teachers worked with their colleagues in preparation for MSPAP. About 70% of teachers took at least one course sponsored by their schools, and essentially the same number took at least one sponsored by their districts. Less than a third participated in any of the listed conferences and other activities sponsored by

national, state, or regional organizations or universities. Similarly, less than a third used one or more of the listed video resources.

Schools and districts are the most common providers of workshops and courses in all areas that teachers take in preparation for MSPAP. Sixty-four percent of fifth-grade teachers and 74% of eighth-grade mathematics teachers participated in at least one activity provided by their school. The percentages participating in district-sponsored activities were similar, but the grades were reversed: 75% of fifth-grade teachers and 65% of eighth-grade mathematics teachers took at least one activity from the district. Both grades took on average a similar number of courses at their schools, just over two, but at the district level eighth-grade mathematics teachers took an average of almost three courses, one more than the average number taken by fifth-grade teachers.

About 12% of teachers participated in activities sponsored by colleges and universities, and about 20% participated in at least one course or workshop sponsored by other organizations. Teachers at both levels utilized these sponsors in roughly similar numbers.

Teachers were much less likely to participate in state-sponsored professional development than in local activities. Only 10% of fifth-grade teachers and 35% of eighth-grade mathematics teachers attended at least one of a number of workshops, conferences, and academies available at the state level. Most of these activities were attended by 5% or fewer of the teachers. The two exceptions were for eighth-grade mathematics teachers, of whom a quarter attended sessions on MSPAP sponsored by the National or Maryland Council of Teachers of Mathematics. Thirteen percent attended the Governor's Academy, sponsored by MSDE to focus on math, science, and technology.

In comparing the vehicles teachers used, it appears that access and convenience play strong roles in teachers' choices of professional development activities. Print materials are easy to distribute and keep as references, one's colleagues are nearby for collaborative work, and courses and workshops offered by the school and district are somewhat easily accessible. In contrast, activities sponsored at the university or state level are probably less convenient than local ones, and video resources may be difficult to distribute and use.

IMPACT OF MSPAP ON SCHOOL MANAGEMENT AND INSTRUCTION

The goal of MSPAP is to change fundamentally what is occurring in schools both in and out of the classroom. We questioned principals and teachers about their perceptions of the changes that have occurred in their schools as a result of the program. According to teachers, MSPAP has affected the climate of Maryland schools in both positive and negative ways. On the positive side, MSPAP appears to have increased the interaction between the principal and the teachers in the school. On the negative side, teachers report a decrease in morale and job satisfaction.

Educators reported changes in instruction, and these changes largely correspond to the emphases of the reform program (e.g., more group work and writing, and less emphasis on computation and language mechanics). MSPAP does not appear to have changed the structure of tracking and course offerings in most elementary and middle schools. (One exception to this may be a small trend away from homogeneous grouping at the middle-school level.)

School Climate

We asked teachers to respond to a series of comments concerning issues of school climate using a Likert response scale (strongly disagree, somewhat disagree, somewhat agree, strongly agree). In addition, teachers were asked to indicate whether they believe MSPAP has affected the status of the school with respect to each statement.

There is good news and bad news to report concerning both school climate and the impact of MSPAP in this area. However, in the case of most aspects of school climate addressed by the survey, only a minority of teachers indicated that MSPAP had caused a change.

Most teachers (82%) reported that the principal in their school had provided constructive feedback to the teachers on their performance. Moreover, in the opinion of almost a third of the teachers, MSPAP facilitated this activity. (Virtually none—2%—said MSPAP led to a decrease.) Eighty percent of teachers report that they are encouraged to experiment in their teaching. Here again, roughly a third (29%) responded that MSPAP had led to an increase in the degree to which they are encouraged to experiment, while 11% said MSPAP had led to a decrease.

Most teachers (74%) responded that too much of their time is diverted from instruction to deal with classroom management issues. However, only 17% responded that MSPAP has led to an increase in the amount of time that is diverted. (Virtually none—2%—responded that MSPAP has led to a decrease in the amount of time diverted from instruction.)

Few teachers reported that morale is high, and a majority reported that MSPAP has harmed it. Nearly three-fourths of teachers (71%) disagreed with the following statement about their school: “Teacher morale is high.” In addition, 57% of teachers responded that MSPAP has led to a decrease in teacher morale in their school, while only a few (4%) reported that MSPAP has produced an increase. This was the only instance in which a majority of teachers responded that MSPAP had led to a change in a specific aspect of school climate. Responses pertaining to job satisfaction were more positive: 74% of teachers reported being at least somewhat satisfied with their jobs. However, roughly a third of teachers responded that MSPAP has led to a decrease in job satisfaction (as opposed to 9% reporting MSPAP has led to an increase).

Teacher opinions were mixed on the current status of student morale: 42% agreed that it is high, while 58% disagreed. Although most teachers did not feel MSPAP has affected student morale, about one-fourth reported that MSPAP had lowered it. (Virtually none—3%—said that MSPAP had improved student morale.) Similarly, most teachers reported that MSPAP has not affected students’ sense of accomplishment, but 20% responded that MSPAP has led to a decrease in students’ sense of accomplishment, and only a few (6%) said that MSPAP has led to an increase.

The Changing Role of the Principal

Most principals reported that MSPAP has substantially changed their role as the instructional leader of their school. A large majority of principals (68%) reported that the percent of time they devote to instructional issues has increased as a result of MSPAP, and most principals (84%) reported that in response to MSPAP, they have given their teachers a great deal of encouragement to “improve instruction generally.”

MSPAP has also influenced the staffing decisions made by roughly a third of principals. Across the two grades, 30% reported moving teachers among grades (either into a tested grade, out of a tested grade, or both) because of MSPAP. A

third of elementary principals report moving good teachers into Grade 5, and about a fourth (27%) reported transferring weaker teachers out of Grade 5. The corresponding numbers for middle-school principals were much smaller: 15% and 13%, respectively. It seems plausible that this difference between grades reflects staffing constraints faced by principals: An elementary school is likely to have a number of teachers qualified to teach fifth grade, whereas a middle school is likely to have relatively few teachers qualified to teach eighth-grade mathematics.

Principals report using MSPAP as a lever to change teachers’ practices. Over 80% of principals responded that MSPAP has been at least somewhat useful as a tool to encourage resistant teachers to make positive changes in their instruction, and nearly half (44%) said it was very useful for this purpose.

A substantial number of principals reported using public recognition or resources to reward teachers for their students’ good performance on MSPAP (Table 3). Overall, roughly half of the principals reported publicly recognizing teachers within the school for their students’ MSPAP performance, and roughly one-fourth reported recognizing teachers outside of the school. A substantial number reported giving teachers additional resources for use within the school as a reward for MSPAP performance. Only a few principals reported using other rewards about which we asked, including extra pay, a lighter teaching load, and relief from administrative and disciplinary duties.

Table 3
Percent of Principals Reporting Rewarding Teachers Based on Their Students’ MSPAP Scores

Type of reward or recognition	Fifth grade	Eighth grade mathematics
Public recognition within the school	60	37
Additional resources for use within the school	30	45
Public recognition outside the school	21	28
Priority on requests for materials	15	15
Relief from administrative or disciplinary duties	1	14
Greater choice of students to teach	2	5
A lighter teaching load	1	1
Extra pay	1	0

Expectations for Student Achievement

Principals reported taking the “high standards for all” message seriously. We asked principals to indicate the degree to which they encouraged their teachers to focus on setting higher expectations for various groups of students in response to MSPAP using a Likert response scale with the options “not at all,” “somewhat,” and “a great deal.” Most (about 65% of elementary-school principals and 85% of middle-school principals) reported encouraging their teachers a great deal to focus on setting higher expectations for students. Furthermore, there was virtually no difference in the percent of principals reporting “a great deal” for low-achieving, average, and high-achieving students.

Teachers’ responses painted a somewhat less optimistic picture: Fewer reported that expectations really had increased, and they reported a less equitable change in expectations. Sixty-three percent of fifth-grade teachers and 48% of eighth-grade mathematics teachers responded that expectations had increased since MSPAP began. When asked separately about different groups of students, however, far more teachers (particularly eighth-grade mathematics teachers) reported that expectations had increased greatly for high-achieving students than for other groups (Table 4).

Most teachers reported that the emphasis on high standards for all students has been helpful for students. However, more teachers thought that the emphasis on high standards has been helpful for average students and high-achieving students than for special-education or low-achieving students (Table 5). In addition, more of the eighth-grade mathematics teachers reported that the emphasis on high standards has been somewhat or very harmful for special-education and low-achieving students than thought they were helpful.

Table 4
Percent of Teachers Reporting That Academic Expectations Have “Increased Greatly,” by Grade and Type of Student

	Fifth grade	Eighth-grade mathematics	Combined
Special education	17	18	18
Low achieving	20	9	15
Average achieving	27	9	19
High achieving	42	28	36

Table 5

Percent of Teachers Reporting That the Emphasis on High Standards Has Been Harmful or Helpful, by Grade and Type of Student

	Fifth grade		Eighth-grade mathematics		Combined	
	Harmful	Helpful	Harmful	Helpful	Harmful	Helpful
Special education	3	48	43	39	21	44
Low achieving	15	66	46	41	28	56
Average achieving	1	95	26	69	12	84
High achieving	1	95	13	70	6	84

Course Offerings

We asked principals a number of questions about course offerings, and an additional caveat about these specific responses should be noted. Survey questions are not fully adequate to characterize changes in student coursework; changes that appear dramatic might be only superficial in reality, and vice versa. For instance, a school might decide to eliminate a course called remedial mathematics and place students who previously would have been assigned to it in a general mathematics class. This could be a major shift for the school that raises the level of content delivered to the lowest achieving students, or it could be no more than a name change, with few implications for the students involved. Thus, the survey results reported here are only a first step in getting a picture of the changes in course offerings stemming from MSPAP and the broader school reform effort.

Only 16% of middle-school principals reported that MSPAP has affected course offerings in their schools. Algebra and pre-algebra were the courses most often mentioned as additions, and basic or general math was the course most often eliminated.⁵ Also, almost all middle-school principals (85%) reported that the number of students assigned to remedial courses had not changed since MSPAP began.

A larger but still modest percentage of middle-school principals (30%) reported that MSPAP has affected the courses offered in the high schools in their

⁵ The former suggests that some principals may have interpreted the question to include adding sections of courses already offered.

area. However, it is likely that many middle-school principals have only limited information about course offerings in the high school in their area. (In fact, when prompted for specific course additions or deletions that have occurred in area high schools, most middle-school principals responded with “I don’t know.”) Of those who gave specifics, computer-assisted math lab was the course most often mentioned as an addition. Reading/language arts, higher mathematics (e.g., calculus), functional mathematics, and remedial mathematics were also mentioned by more than one principal. Only four principals could specify courses that had been dropped at the high school level.⁶

Student Grouping and Remediation

Grouping students who are similar in terms of previous achievement is a common practice that often begins at the middle-school level. Many current education reformers, with their emphasis on high standards for all, take a negative view of this practice.

The responses of elementary- and middle-school principals differed markedly with respect to homogeneous grouping: Middle-school principals were much more likely to report grouping but were also more likely to report that grouping had decreased in response to MSPAP. While most (85%) middle-school principals reported grouping students homogeneously, only about a fourth (28%) of elementary principals reported doing so.⁷ Few elementary principals reported changes in grouping practices in response to MSPAP, but some middle-school principals indicated that there has been a trend away from homogeneous grouping in response to the program. Thirty-three percent of principals who reported still using homogenous grouping said that they use it less in response to MSPAP. In addition, 28% of the principals who reported that they do not currently group students homogeneously stated that they did use homogeneous grouping before MSPAP began.

Teachers, on the other hand, indicated that their use of homogeneous grouping has declined. Almost half (43%) of teachers reported that they group students homogeneously based on ability within their classes less that they did

⁶ Two principals indicated that remedial math was eliminated, and a third said Introduction to Algebra was eliminated. The fourth principal noted a decline in vocational education courses, home economics, and electives in the arts.

⁷ Because this question was posed to principals and not teachers, it is possible that grouping within classrooms was overlooked.

before MSPAP. Also, 70% of fifth-grade teachers and 51% of eighth-grade mathematics teachers reported having students work in groups of mixed ability more often than before MSPAP was first administered.

However, there also appears to be a slight trend toward providing students with remediation outside of normal class time. Roughly a quarter of the teachers reported an increase in the number of students participating in before- or after-school remedial programs. Also, 12% of fifth-grade teachers reported an increase in participation in remedial summer school programs.

Effects on Instruction

Has MSPAP influenced the way that teachers teach? And, if so, what are the most prevalent changes teachers have made? These were two of the key research questions that motivated this study. We asked teachers a number of questions pertaining to changes in practice: both closed-ended questions that would indicate the percent of educators perceiving specific instructional changes, and open-ended questions aimed at getting educators' opinions of the most salient changes.

A majority of teachers (61%) reported that they have focused a great deal on "improving instruction generally" in their efforts to improve scores on MSPAP. In addition, when asked if holding schools accountable for meeting MSPAP standards causes teachers to increase their efforts to improve learning, 71% agreed.

Responses to Forced-Choice Questions About Instruction

Teachers reported that MSPAP has produced both positive and negative instructional effects. Overall, slightly more teachers reported that there were positive effects than negative effects. Virtually all fifth-grade teachers (98%, including even those opposed to MSPAP) felt MSPAP had had at least a small positive effect on instruction, and most eighth-grade mathematics teachers (92%) concurred. In both groups, 22% of teachers felt MSPAP has had "a great deal" of positive effect on instruction in their school. However, teachers also indicate that MSPAP has had negative effects on instruction: 85% of fifth-grade teachers and 71% of eighth-grade mathematics teachers reported at least a small negative effect. Approximately 12% of teachers felt that MSPAP had had a great deal of negative effect on instruction in their school.

Fifty-five percent of fifth-grade teachers and 33% of eighth-grade mathematics teachers reported focusing "a great deal" on improving the match

between the content of their instruction and the content of MSPAP. The difference across grades is most likely related to a greater percentage of fifth-grade teachers perceiving a lack of alignment. (More fifth-grade teachers than eighth-grade mathematics teachers noted concern for the lack of alignment between their curriculum and MSPAP in response to an open-ended question about the negative effects of MSPAP on instruction.)

Seventy-four percent of teachers indicate that there is content they emphasize more because of MSPAP, and 52% indicate that there is content they emphasize less because of MSPAP. There appears to be a difference among the grades in this respect: Fifth-grade teachers are quite balanced in terms of the percents indicating that there have been increases (70) and decreases (62), whereas eighth-grade mathematics teachers were more likely to report increases (76) than decreases (45).

Principals appear to be facilitating a change in the curriculum taught in their schools. Almost all principals (95%) reported obtaining new instructional materials aligned with MSPAP. About half of the principals reported encouraging their teachers a great deal to focus instruction on “skills or content likely to be on MSPAP.” Most (85%) reported that their schools’ emphasis on material likely to be emphasized by MSPAP had increased, but only 34% said it had increased greatly.

Surprisingly, principals did not report lessened emphasis on untested material, but many teachers did. Very few principals stated that there had been a decrease in their schools in the emphasis given to the pre-MSPAP curriculum (4%), to untested subject areas (6%), or to material unlikely to be tested even though it is in a tested subject area (10%). On the other hand, many teachers did report that some areas had been de-emphasized because of MSPAP. Sixty-nine percent of fifth-grade teachers agreed (21% strongly) that “MSPAP has caused some teachers to de-emphasize or neglect untested subject areas.” Eighth-grade mathematics teachers were asked a narrower question: whether MSPAP had caused some math teachers to de-emphasize or neglect untested mathematics topics. The responses, however, were similar: 69% agreed (30% strongly) that such changes in emphases had occurred.

Because the content taught differs in substantial ways across the two groups of teachers we surveyed, we asked different questions in the two grades

concerning specific changes in emphasis. Overall, we found a greater agreement concerning what specifically had been increased than we did about what had been decreased. It is not surprising, given that MSPAP clearly emphasizes certain skills and content areas, that teachers reported increasing the instructional time they devoted to those areas. However, what schools and teachers take time away from in order to accommodate MSPAP is not clearly dictated by the assessment. Therefore, teachers reported time being taken away from a wide variety of content areas as well as various classroom activities.

Fifth-grade teachers. Fifth-grade teachers were asked how they have shifted available instructional time between subject areas since MSPAP was first administered. The subject area to which by far the most teachers indicated allocating more instructional time was writing (Table 6). The subject areas for which the most teachers indicated a decrease since MSPAP began were reading, art, music, and physical education. Seventy-two percent of the teachers who reported shifting instructional time indicated that the changes were due largely to MSPAP.

Fifth-grade teachers were also asked how they have changed emphasis *within* math and language arts (Table 7). Within language arts, fifth-grade

Table 6

Percent of Fifth-Grade Teachers Reporting Changes in Content Emphasis

Content area	Decreased	Increased
Writing	0	66
Mathematics	0	47
Science	2	41
Social studies	8	30
Reading	14	32
Art	14	11
Physical education	10	0
Music	11	0

Table 7

Percent of Fifth-Grade Teachers Reporting Changes in Content Emphasis
Within Language Arts and Mathematics

Content	Decreased	Increased
Language Arts		
Writing for a variety of purposes	1	89
Analysis and evaluation of text	2	67
Literary comprehension	9	51
Spelling, punctuation, and grammar	28	21
Mathematics		
Mathematical communication	1	81
Problem solving using meaningful tasks	5	80
Application	10	79
Number facts and computation	44	18

teachers reported an increase in emphasis on writing for a variety of purposes, analysis and evaluation of text, and literary comprehension. A decrease in emphasis was reported on spelling, punctuation, and grammar. The majority of fifth-grade teachers indicated that they have increased the emphasis on mathematics communication, meaningful problem solving, and mathematics applications. The only aspect of mathematics in which a sizable proportion of fifth-grade teachers indicated a decrease in emphasis was number facts and computation.

Because much of the time spent in elementary grades is on education that is not traditional desk work, we also asked about a variety of other activities. The two activities that a sizable proportion of the teachers reported decreasing in order to accommodate MSPAP were organized play (30%) and student free time (25%).

Eighth-grade mathematics teachers. Teachers indicated that they have increased the emphasis on a number of aspects of mathematics that we asked about, especially data analysis, communicating mathematical ideas, and problem solving (Table 8). The only aspect of mathematics for which a sizable proportion of eighth-grade mathematics teachers indicated a decrease in emphasis was computation and algorithms.

Table 8

Percent of Eighth-Grade Mathematics Teachers Reporting Changes in Content Emphasis

Content	Decreased	Increased
Data analysis	0	76
Communicating mathematical ideas and solutions	0	76
Problem solving and reasoning	0	70
Use of graphs and tables	0	60
Conceptual knowledge	0	55
Space, dimensionality, measurement	5	49
Ratios, proportions, and percents	3	30
Computation and algorithms	35	1

Teachers' Responses to Open-Ended Questions About Instruction

As noted, teachers were asked open-ended questions about the most positive and negative effects of MSPAP on instruction in their schools. The positive comments made by the most eighth-grade mathematics teachers concerned cooperative work: MSPAP has led teachers to emphasize cooperative work more and to have their students do more of it. Some also commented that MSPAP has led teachers to focus more instruction on writing, problem solving, thinking skills, real-life applications, and hands-on activities. Teachers also commented that MSPAP has led them to make greater use of an interdisciplinary approach and to emphasize process more while de-emphasizing rote teaching and learning.

Fifth-grade teachers were positive about many of the same things as the eighth-grade mathematics teachers. However, fifth-grade teachers were more likely to comment on an increase in emphasis on having students write and explain and an increased focus on teaching thinking skills.

Teachers also noted a number of negative effects that MSPAP has had on instruction. The negative comments noted by the most teachers in both grades concerned the time taken away from instruction in order to prepare for or administer MSPAP (more eighth-grade mathematics teachers made this comment), and, as noted above, the lack of integration of MSPAP with the teachers' curriculum. Also, some teachers in both fifth and eighth grades were negative because they believe MSPAP forces teachers to teach to the test.

PREPARING STUDENTS FOR MSPAP

The success of Maryland's education reform hinges substantially on the approaches teachers follow to prepare students for the MSPAP assessments. Some methods of test preparation improve instruction, while other methods may degrade it. Similarly, the validity of MSPAP scores and gains will depend on test preparation: Some methods may produce real, generalizable gains in student performance, while others will inflate scores and create an illusion of progress.

We asked principals and teachers a variety of questions pertaining to test preparation and administration of the assessment. Test preparation shades into instruction, and one goal of the current education reform movement is to diminish further the distinction between them. Our questions therefore spanned a range from generalized instructional changes (e.g., giving more homework, raising expectations, placing more emphasis on higher order thinking skills, etc.) to methods that are tightly tied to the test itself (e.g., practicing old MSPAP items). The methods about which we asked also ranged from clearly legitimate (e.g., familiarizing students with the format of the assessment) to clearly illegitimate (e.g., providing hints on correct answers while administering MSPAP).

Both principals and teachers reported widespread reliance on a wide variety of approaches to preparing for MSPAP, including setting higher expectations, placing more emphasis on higher order thinking skills, attempting to improve students' motivation to do well on the assessment, using practice tests, and giving instruction on test-taking skills. However, few teachers reported placing heavy emphasis on more and harder schoolwork or more homework. Few teachers reported instances in their schools of most of the clearly inappropriate test-preparation or test-administration practices about which we asked, although there were several important exceptions. More than a fourth reported that test items had been rephrased during testing, and nearly half reported that potentially helpful materials had been left visible in rooms in which MSPAP was administered.

Educators' perceptions of the effectiveness of these approaches in raising MSPAP scores, however, showed a disturbing pattern. Only a small minority of educators reported the opinion that broad improvements in knowledge and skills, or even improvements in the knowledge and skills emphasized in MSPAP, contributed a great deal to score gains in their schools. A far larger percentage of

educators gave credit to factors that have the potential to inflate test scores without improving achievement: work with practice tests, increased familiarity with MSPAP, and the use of test-preparation materials.⁸

Methods for Preparing Students

Many of the methods for preparing students for an assessment fall on a continuum from instruction focused broadly on the assessment's domains to test preparation focused narrowly on the content and format of the assessment. For purposes of description, however, we have classified them into four overlapping categories. *Instructional approaches* include activities designed to teach students the underlying skills and abilities assessed by MSPAP. *Motivational approaches* include methods to encourage students to try hard, either on the assessment or in general. The methods in the category *direct test preparation* include the use of practice tests and similar materials and instruction in test-taking skills. *Questionable test preparation and administration* includes methods that are often considered inappropriate or even unethical, such as providing practice on secure (nonreleased) test items and providing hints during testing. The distinctions between these categories, however, are not always precise.

Instructional Approaches

In the view of our respondents, broad instructional changes were among the most prevalent responses to MSPAP. (These instructional changes are discussed in more detail in the preceding section, but several are noted here to place the narrower forms of test preparation noted below into context.) About three-fourths of principals reported encouraging their teachers “a great deal” to raise expectations for all students. Nearly all (88%) reported giving their teachers a great deal of encouragement to focus instruction more on higher order thinking skills, and nearly as high a percentage reported giving teachers a great deal of encouragement to “improve instruction generally.” A somewhat smaller majority (58%) said their schools’ emphasis on higher order thinking skills had actually increased greatly.

Similarly, a majority of teachers (61%) reported that they had focused a great deal on “improving instruction generally” in their efforts to improve scores

⁸ We use the term “inflate” to indicate increases in test scores that are not generalizable, i.e., they are not accompanied by increases in the skills and abilities that the assessment is intended to measure.

on MSPAP. Only about one-fourth of teachers, however, reported focusing a great deal on requiring more or harder work in school, and only 9% reported focusing a great deal on more homework.

A long-standing concern about assessment-based accountability is the potential for narrowed instruction tailored to raise scores rather than improve mastery more generally. We approached this issue with questions asking principals and teachers about their efforts to align instruction with the test and about changes in emphasis on untested material. Alignment of instruction with the assessment is of course an ambiguous category, since some degree of alignment is one of the goals of the program, but excessive alignment can nonetheless inflate scores (i.e., by narrowing instruction to topics specifically covered by the assessment rather than the broad domain of knowledge the assessment is designed to measure).

Educators reported efforts to align instruction with MSPAP less frequently than they reported the general instructional changes just noted. About half of the principals reported encouraging their teachers a great deal to focus instruction on “skills or content likely to be on MSPAP.” Most (85%) reported that their schools’ emphasis on material likely to be emphasized by MSPAP had increased, but only 34% said it had increased greatly. Slightly under half (44%) of teachers reported focusing a great deal on “increasing the match between the content of instruction and the content of MSPAP” in their efforts to raise scores, and 38% reported focusing a great deal on using “MSPAP-like tasks” in regular instruction.

Principals did not report lessened emphasis on untested material, but many teachers did. We asked principals whether there had been a change in their schools in the emphasis given to the pre-MSPAP curriculum, to untested subject areas, or to material unlikely to be tested even though it is in a tested subject area. In each case, 11% or fewer of the principals reported a decrease in emphasis. In contrast, 69% of fifth-grade teachers agreed that “MSPAP has caused some teachers to de-emphasize or neglect untested subject areas,” and 21% strongly agreed with this statement. Eighth-grade mathematics teachers were asked a narrower question: whether MSPAP had caused some math teachers to de-emphasize or neglect untested mathematics topics. The responses, however, were similar: 69% agreed, and 30% strongly agreed, that such a change in emphases had occurred.

Motivational Approaches

Educators reported a considerable use of motivational approaches to raising MSPAP scores, but fewer teachers than principals reported reliance on them. Seventy percent of principals reported encouraging teachers a great deal to improve students' motivation to do well on MSPAP. About half of teachers (49%) reported that they focused a great deal specifically on student motivation to do well on MSPAP in their efforts to raise scores, and about the same number (45%) reported focusing a great deal on student motivation generally.

Teachers were asked about the extent to which their schools relied on each of 8 specific incentives to encourage students to do well on MSPAP. Almost all reported that their schools placed some reliance on discussing the importance of good performance on MSPAP, and about half (53%) reported placing a great deal of reliance on this approach. The only other motivational approaches for which at least a fourth of teachers reported a great deal of reliance were praising or criticizing performance on practice tests (27% of teachers) and providing feedback to students on the current year's assessment tasks (28% of teachers).

Direct Test Preparation

We also questioned principals and teachers about direct test preparation, such as giving students practice on old assessment items or practice tests and giving them instruction on test-taking skills. This practice, like curriculum alignment, is inherently ambiguous. A certain amount of practice can be desirable—for example, to provide students with the necessary degree of familiarity with test formats and to illustrate concretely the types of knowledge and skills that are expected of them. Here again, however, excessive reliance on direct test preparation runs the risk of inflating scores (and siphoning limited instructional time away from other activities).

Principals reported more emphasis on the use of test-preparation materials than on test-taking skills. About three-fourths of principals reported encouraging their teachers a great deal to use test-preparation materials. (Nearly all reported encouraging teachers to some degree to do so.) In contrast, only about half of the principals (45%) reported encouraging teachers a great deal to teach test-taking skills.

Teachers' responses suggested that many relied substantially on practice tests and instruction on test-taking skills. One set of questions asked of teachers was similar to those asked of principals: Teachers were asked how much they focused on "improving test-taking skills" and "using practice tests and other test-preparation materials" in trying to improve MSPAP scores. When asked in this way, half of the teachers reported focusing a great deal on test-taking skills, while slightly fewer (41%) reported focusing a great deal on the use of practice tests. However, nearly all teachers reported focusing at least a moderate amount (more than "a small amount") on practice tests (86%) and test-taking skills (91%). In addition, over half of the teachers (56%) reported that students were occasionally or frequently given practice with items "highly similar to the previous year's nonreleased MSPAP items."

In our mail surveys, we also asked teachers how much time they devoted to 6 types of test-preparation activities during the previous school year. Eighth-grade mathematics teachers were asked how many partial or full class periods they devoted to each; fifth-grade teachers were asked the number of days during which they did each. All teachers were also asked to convert these answers to a number of minutes in the entire year.⁹

Fifth-grade teachers' responses varied greatly: Some reported very little use of any of the test-preparation activities about which we asked, while others reported allocating considerable time to them. The median teacher reported allocating 5 hours over 5 days on sample or publicly released MSPAP tasks (Table 9), but a fourth of the teachers reported 2 hours or less, while another fourth reported 9.3 hours or more. (In Tables 9 and 10, the percentiles refer to a ranking of teachers based on the time they report allocating to each activity. Thus, for example, 75% of teachers reported allocating 9.3 hours or less to the use of released MSPAP tasks.) Tasks developed by the Maryland Assessment Consortium (MAC) or other assessment groups were used a bit less; the median teacher reported 4 hours, again over 5 days. The median teacher allocated 2.5 hours to practice tests developed at the local (school or district) level and about an

⁹ For example, the question asked of eighth-grade teachers read as follows: "Last year (1993-94), how much did you use each of the following methods to prepare students for MSPAP? EXAMPLE: if you devoted about 5 minutes per day for approximately 10 days, you would enter 10 under 'Partial or full class periods' and 50 under 'Total minutes in entire year'."

Table 9

Reported Hours Devoted to Test-Preparation Activities, Fifth-Grade Teachers

	25th Percentile	50th Percentile	75th Percentile
Used sample MSPAP tasks	2.0	5	9.3
Used MAC tasks	2.0	4	8.7
Used local practice tests	0	2.5	5.3
Used own practice tests	0	1.1	5.0
Used student MSPAP work	0	0	0
Taught test-taking strategies	0.8	1.7	2.5

hour to practice tests developed specifically to prepare for MSPAP. Very few teachers allocated any time to using copies of student work on MSPAP.

Despite the fact that about half of the teachers reported focusing a great deal on test-taking skills, teachers generally reported allocating much less time to instruction in test-taking strategies than to the use of practice test items. The typical teacher reported allocating less than 2 hours to test-taking skills, and 75 percent reported 2.5 hours or less. It is possible, however, that they focused on test-taking skills in the context of other activities (such as using practice tests) the primary focus of which was not test-taking skills *per se*.

The total time allocated to all of these forms of practice tests together is of course considerably larger than the time spent on any one, but it cannot be estimated precisely from the survey. The median teacher reported allocating 12 hours in total to all five of the types of practice tests about which we asked (excluding instruction in test-taking skills), and a fourth of the teachers reported 23 hours or more. These totals may overstate their allocation of time to practice tests, however, because it is possible that some teachers reported the same time more than once. (For example, a teacher who gave a practice test comprising both MSPAP and MAC items might have mistakenly reported the total time for that practice test for both categories.) These totals, even if overstated somewhat, represent a sizable amount of time but only a modest share of the total instructional time available during the year. For example, if one assumes that a typical fifth-grade teacher has available 825 instructional hours per year (5 hours

per day for 165 days),¹⁰ the 23 hours reported by the teacher at the 75th percentile represents roughly 2.4% of available instructional time.

Eighth-grade mathematics teachers reported spending considerably fewer hours on most of these test-preparation activities, but if teachers in other subject areas behaved similarly, the total time allocated to test preparation across subjects would have been considerably greater for eighth-grade students than for fifth-grade students. As in fifth grade, sample and released MSPAP tasks were allocated the most time: 3.3 hours by the median teacher, and 6 hours or more by 25% of the teachers (Table 10). Local practice tests were used a bit less than MSPAP tasks by the typical teacher (2.5 hours) but about as much by teachers who allocated relatively large amounts of time to these activities. MAC tasks were used less, and practice tests developed by the teachers themselves were used least, with less than an hour allocated to such tests by the typical teacher. With the exception of locally developed tests, more than a fourth of the eighth-grade mathematics teachers reported no use of each of the types of practice tests about which we asked. As in Grade 5, teachers reported allocating relatively little time to instruction in test-taking skills—less than an hour in the case of the median teacher.

The greater reported allocation of time to test preparation by eighth-grade mathematics teachers becomes apparent when these results are compared to

Table 10

Reported Hours Devoted to Test-Preparation Activities, Eighth-Grade Mathematics Teachers (Mathematics Only)

	25th Percentile	50th Percentile	75th Percentile
Used sample MSPAP tasks	0	3.3	6.0
Used MAC tasks	0	1.5	4.2
Used local practice tests	0.8	2.5	6.7
Used own practice tests	0	0.8	2.0
Used student MSPAP work	0	0	0
Taught test-taking strategies	0.8	0.8	2.5

¹⁰ This assumes that a total of 10 days of a typical 175-day year are spent on non-instructional activities such as testing, field trips, half-days off for professional time, etc.

available instructional time. The median eighth-grade mathematics teacher reported allocating 8.3 hours to the five types of practice tests together. This corresponds to 10 50-minute class periods, or perhaps 6% of annual instructional time.¹¹ A fourth of teachers reported more than 14 hours, corresponding to 16.8 class periods, or nearly 10% of instructional time. Even allowing for some overstatement from double-counting of time, as explained above, this suggests that many eighth-grade mathematics teachers are allocating a sizable share of available instructional time to practice tests.

About two-thirds of the teachers in both grades reported that they carried out most of this test-preparation activity throughout the year. Only one-fourth of the teachers reported that they did it mostly in the month preceding MSPAP, and only a handful reported doing it mostly during the two weeks before the assessment.

The interpretation of this test preparation is ambiguous. Because MSPAP, like many performance assessments, is intended to model improved instruction, it is not clear how much of this reported use of practice tests should be seen as coaching for the assessment rather than as instruction. On the other hand, it is also plausible that activities that many observers would consider test preparation, such as the use of tasks “highly similar” to MSPAP, might not have been counted in some teachers’ responses to our very specific questions—particularly given the blurring of the distinction between instruction and assessment by some current reformers. Additional information, such as detailed descriptions of the instruction and test-preparation activities and empirical data on the generalizability of students’ gains on MSPAP, would be needed to reach a firm conclusion about the value and effects of these activities.

Questionable Test Preparation and Administration

Because we feared that teachers would find it difficult to provide us with information about questionable test-preparation activities and inappropriate test administration, we took several steps to make our questions about these topics (which were asked by telephone) less threatening. Teachers were asked about practices throughout their schools, rather than their own practices. They were told that people disagree about which practices are desirable. They were reminded that their responses were confidential and that our intent was to describe

¹¹ This assumes 8,250 instructional minutes per year: 165 50-minute class periods.

practices throughout the state, not to judge individual schools. Nonetheless, pilot interviews suggested that the questions made some respondents uncomfortable. For example, even though the questions were asked about the school as a whole, one pilot respondent paused after one of the questions and then responded several times that *she* had never engaged in the practices in question. These factors raise the risk of “social-desirability bias” in the results, which in this case could entail underreporting the actual incidence of the activities in question. We have no evidence, however, that underreporting occurred.

Teachers reported an appreciable incidence of only two of the questionable test practices about which we asked. Forty-two percent reported that some testing rooms had “posters on the walls or instructional materials in view that might help students during the assessment” (Table 11). About a fourth reported occasional or frequent rephrasing of test questions during MSPAP administration. The reported incidence of other questionable practices was much lower: fewer than 15% of teachers reported knowing of these practices in their schools (Table 11).

Table 11

Percent of Teachers Reporting Incidence of Questionable Test-Preparation and Administration Practices

	Percent
Potentially helpful materials in view in some testing rooms ^a	42
Questions rephrased during testing time	27
Revisions recommended during or after assessment	14
Questions about content of assessment answered during testing time	13
Items read for non-accommodated students	13
Students given practice on last year’s nonreleased items	12
Additional time given to non-accommodated students	8
Writing responses for non-accommodated students	7
Hints provided on correct answers	6
Copied student responses to nonreleased items ^a	2
Changing or editing answers in assessment booklets	2
Copied or obtained nonreleased items ^a	1

^a These items allowed only “yes” or “no” answers. In all other cases, respondents were allowed to answer “never,” “occasionally,” or “frequently,” and the percent answering either “occasionally” or “frequently” is presented.

Moreover, because any instance would be cause for a “yes” answer, the extent of the practices may be very limited even in schools in which they are known to have occurred.¹²

Perceived Causes of Gains on MSPAP

We asked educators whether their schools’ MSPAP scores had increased. Those who answered positively (both principals and teachers) were asked seven questions about the factors to which they ascribed their gains in scores.

Although substantial majorities of principals and teachers reported placing a great deal of emphasis on broad instructional changes in their efforts to raise scores, their explanations of gains in their own schools paint a more pessimistic picture. “Broad improvements in knowledge and skills” was cited *least* frequently by teachers as having contributed a great deal to their MSPAP gains; only 15% gave that response (Table 12). In contrast, the factor teachers cited most frequently (57%) as having contributed a great deal to their MSPAP gains was “work with practice tests and preparation materials,” followed by “increased familiarity with MSPAP” (45%) and “improved test-taking skills” (29%). Even “improvements in knowledge and skills emphasized in MSPAP” was cited relatively infrequently (by 20%) as having contributed a great deal. Principals

Table 12

Percent of Teachers and Principals Reporting That Each Factor Contributed “A Great Deal” to MSPAP Gains

	Teachers	Principals
Work with practice tests and preparation materials	57	51
Increased familiarity with MSPAP	45	55
Improved test-taking skills	29	15
Differences between cohorts	21	13
Improvements in knowledge and skills emphasized in MSPAP	20	32
Increased student motivation	16	8
Broad improvements in knowledge and skills	15	25

¹² Note that at least one of the questions may have been misunderstood by some of the respondents. Twelve percent reported that students were given practice on the previous year’s nonreleased items, but that is inconsistent with the negligible percentage who reported that someone had copied or obtained the previous year’s nonreleased items.

were somewhat more positive, in that a slightly higher percentage cited improvements in knowledge and skills as having contributed a great deal. Principals too, however, were more likely to cite work with practice tests (51%) and familiarity with MSPAP (55%) than improvements in skills (25% and 32%) as having contributed a great deal to their MSPAP gains (Table 12). These opinions about gains in educators' own schools were consistent with their opinions about the value of MSPAP as a tool for evaluating educational improvement in Maryland more generally; as noted earlier, more than 65% of our respondents said that some schools have found ways to raise scores without improving education.

These responses raise concerns about the validity of gains in MSPAP scores. The goal of the accountability program is improvements in students' knowledge and skills. If educators are correct that factors such as work with practice tests have contributed more to gains than have improvements in knowledge and skills, it is likely that gains in scores are misleading as indicators of success. (These issues are discussed in more detail in the following section.) Survey data, however, are not a sufficient test of this hypothesis. To ascertain with confidence whether gains are substantially inflated would require empirical data about the generalizability of student performance gains on the MSPAP assessment to other, comparable measures of mastery of the domains it is designed to assess.

DISCUSSION

Principals' and teachers' responses to our surveys provide a complex view of Maryland's assessment-based reform. Both positive and negative opinions were expressed by sizable majorities of respondents, and opinions on some issues were divided. On the positive side, about three-fourths of principals and half of teachers expressed some degree of global support for the program, and principals reported that they have become more positive in the last few years. Although almost all principals (91%) reported that MSPAP imposes more than a minor burden on their schools, almost half reported that the benefits outweighed the burden, and fully two-thirds reported that the benefits of the program at least balanced the burden. A large majority of teachers believe MSPAP tests a broader range of skills than do multiple-choice tests, and a smaller majority responded that the assessment more accurately mirrors their teaching. A majority also reported that MSPAP provides at least somewhat accurate information on student performance in specific areas.

Improving instruction is a primary goal of MSPAP, and a large majority of educators believe that it has met with some success. Over 80% of principals said that the program has been useful as a tool for encouraging resistant teachers to make positive changes in their instruction. Almost all teachers reported that MSPAP has had a positive effect on instruction in their schools in their grade or subject, and about three-fourths reported that instruction had improved by more than a small amount. Sixty-one percent of teachers reported that they have focused a great deal on “improving instruction generally” in their efforts to improve scores on MSPAP, and 71% agreed that holding schools accountable for meeting MSPAP standards causes teachers to increase their efforts to improve learning.

Many of the specific changes in instruction reported by teachers are consistent with the goals of the Maryland reform program. For example, almost all (89%) of fifth-grade teachers reported placing more emphasis on writing for a variety of purposes, and about 80% reported increasing their emphasis on math communication, meaningful problem solving, and math applications. About three-fourths of eighth-grade mathematics teachers reported placing greater emphasis on data analysis, communicating mathematical ideas and solutions, and problem solving and reasoning. Large numbers of teachers have made use of professional development opportunities to prepare for MSPAP, and many express positive opinions of the activities in which they have participated. Few teachers reported any instances of most of the clearly inappropriate test preparation or testing practices about which we asked.

On the other hand, many of the principals and teachers raised serious concerns about the current functioning of the program. Most teachers (85% in fifth grade and 71% in eighth-grade mathematics) reported at least a small negative effect of MSPAP on instruction. Open-ended comments about negative effects were diverse, but among the most common were comments on time taken away from instruction to prepare for or administer MSPAP and perceived pressure to teach to the test. In addition, some teachers complained of a lack of integration between MSPAP and their broader curriculum. Although most teachers reported an increase in expectations for students, their responses suggest that the program may be increasing rather than shrinking inequities in this respect. While over 40% of fifth-grade teachers reported that expectations had increased greatly for high-achieving students since MSPAP began, only about

half as many said that expectations had increased greatly for low-achieving students. Eighth-grade mathematics teachers painted a more pessimistic picture yet: Only 9% reported such a change for low-achieving or average students, in contrast to 28% for high-achieving students.

Support for some of the accountability uses of MSPAP data is low, particularly among teachers. Further, majorities of both principals and teachers maintained that estimates of progress and comparisons among schools using MSPAP may be distorted by a variety of irrelevant factors, such as differences in student characteristics. Many teachers, particularly in the fifth grade, expressed concerns about the nature of MSPAP itself, including the developmental appropriateness of MSPAP tasks and the time allotted to perform them.

Educators' responses to many of our questions raised the issue of potentially inflated scores. Many teachers reported reliance on various forms of test preparation, such as use of practice tests and instruction on test-taking skills, that run the risk of inflating scores without producing comparably large real gains in achievement. Roughly 40% to 50% reported placing a great deal of reliance on such techniques, and a considerable number reported allocating substantial time to them. About two-thirds of teachers reported that MSPAP has led some teachers to de-emphasize or neglect untested material. Nearly a third of principals reported changing staff assignments across grades to place their more capable teachers in the grade in which MSPAP is administered, which raises the risk that scores from the target grades may overstate the performance of the schools in which that occurred. A majority of educators reported that some schools have found ways to raise scores without improving education. Even more striking are educators' explanations of the MSPAP gains in their own schools. Almost 60% of teachers responded that work with practice tests and preparation materials had contributed a great deal to their gains on MSPAP, while only 15% said that broad improvements in knowledge and skills had contributed in this way. Principals were substantially more optimistic in this regard, but nonetheless, they were nearly twice as likely to point to practice tests than to broad improvements in knowledge and skills as sources of their schools' score gains.

Implications

That these surveys generated a mix of favorable and unfavorable responses from teachers and principals is not surprising. The Maryland reforms call for very

large increases in student performance and fundamental changes in instruction. Any reform and accountability program of this scope can be expected to produce unintended as well as intended changes in schooling and to generate dissatisfaction among educators even when it is working as intended. Moreover, the reform program is still young. Missteps are inevitable, and considerable time will be required for the program to evolve, for staff to be retrained, and for educators' and students' behaviors to change in response.

Nonetheless, the results reported here suggest important concerns that MSDE might address to build upon the initial successes of the program and ameliorate its early weaknesses. In addition, these findings point to the need for additional research to monitor the program and its effects. Surveys can provide only a first look at the changes effected by reforms of this sort; they raise many questions that can be answered with confidence only by additional investigation.

Lack of Support for Accountability

The limited support for accountability aspects of the program voiced by teachers is one finding that warrants further exploration. Our results suggest that support may be restricted by specific problems that many teachers perceive in the assessment or its use, but there may be other important causes as well, some of which may indicate positive effects of MSPAP. Some teachers may simply be displeased at being pressured to change; recall that many of our principals reported that they have found MSPAP a useful tool for working with such teachers. Similarly, some of the "undue pressure" reported by many teachers might be desirable in the eyes of program advocates, while some may indicate substantial problems. Some teachers may object to an accountability system that is focused so much on test scores, regardless of the nature of the assessment, and yet others may feel that they are caught between the demands of the program and obstacles impeding their efforts to meet its demands. Thus, in some instances, a lack of support might be seen by program advocates as a symptom of success—for example, reluctant teachers feeling pressure to change instruction—while in other cases, the lack of support might point to important and remediable weaknesses in the program.

The Perceived Burdens Imposed by MSPAP

Educators' reports of the burdens imposed by the program may also signal both successes and deficiencies. For example, the reports of many principals that the need for staff retraining was burdensome may be an indication that the program is succeeding in causing widespread retraining as a path to instructional improvement. On the other hand, reports that time demands and unclear achievement targets are perceived as burdensome may point to possible need for modifications of the program. Information more detailed than that provided here would help determine whether program modifications designed to reduce burdens would be appropriate.

Approaches for Providing Student-Level Scores

The fact that MSPAP does not provide scores for individual students is currently a hotly debated issue in Maryland. Consistent with that, almost all teachers and principals responded that MSPAP results would be more useful if individual scores were provided. Because of the design of the assessment, however, the material administered to any individual student is often not a reliable measure of the broader domains or constructs that MSPAP is intended to measure.¹³ Routine reporting of individual scores therefore requires either accepting the lower quality of student scores for this purpose or changing the assessment's design to make individuals' scores better estimates of the domains assessed by MSPAP. Individual scores could be improved by increasing the number of tasks each student is administered, either by increasing testing time or by changing the mix of formats to include some, such as multiple-choice items, that require much less time per task. Most principals were unwilling to add testing time; most teachers and middle-school principals, but only half of the elementary principals, said they would accept a change in the mix of task formats in order to obtain individual scores. This suggests that if MSDE decides to provide individual scores, a reasonable option might be to experiment with alterations of the assessment's

¹³ The Maryland State Department of Education and CTB/Macmillan/McGraw-Hill have reported internal-consistency reliability estimates for MSPAP (e.g., Yen et al., 1992). Although these vary substantially across subject areas and grades, many of them are quite high. However, these internal consistency estimates reflect only the tasks included in the particular task clusters each student takes. The clusters are not designed to be strictly comparable, and these internal consistency estimates do not take into account variation *among* the clusters included in the assessment. Therefore they are not an indication of the reliability of individual students' scores as measures of the constructs assessed by MSPAP.

mix of formats while monitoring educators' responses to ensure that instructional incentives are not eroded.

Specificity of Curriculum Frameworks

The response of some teachers that the curriculum content for the assessment was too poorly defined to permit them to prepare students is noteworthy, but additional information is needed to determine whether this signals a problem. Some degree of dissatisfaction on this count is to be expected when teachers are being asked to make fundamental changes in their instruction, and indeed it might be seen as a positive sign. Whether the degree of concern noted in these surveys signifies a substantially negative outcome remains an open question; recall that although about three-fourths of teachers agreed to some extent that content was too poorly specified to permit them to prepare students for MSPAP, only 20% strongly agreed.

Moreover, the optimal level of specificity of content standards and curriculum frameworks in assessment-driven reform remains controversial and has been the focus of substantial debate in the context of other, similar reforms (e.g., Hambleton et al., 1995). The optimal balance between breadth and specificity may also vary across subject areas and grades. On the one hand, reformers often want the standards to be broad and general in order to focus educators' attention on major goals of education rather than narrow, decontextualized outcomes. On the other hand, if teachers perceive the frameworks as providing less specific information than is provided by the assessment itself, they might respond by using the assessment as a surrogate for a curriculum framework as a guide for instruction. Research has shown similar behaviors in another state using assessment as a lever for reform (Stecher & Mitchell, 1995). This in turn might increase the risk of both inflated test scores and inconsistent instructional changes across schools.

Despite this uncertainty, the results reported here suggest reexamining the state's curriculum frameworks to determine whether some might warrant greater specificity. In this respect, it is suggestive that fifth-grade teachers were almost twice as likely as eighth-grade mathematics teachers to agree strongly that the curriculum content for the assessment was too poorly defined to permit them to prepare students. Fifth-grade teachers, unlike eighth-grade mathematics teachers, have to concern themselves with all subjects tested by MSPAP, and

frameworks in some other subjects appear considerably less specific and detailed than the mathematics frameworks (see Maryland School Performance Program, 1990).

Effects on Equity

An underlying theme of the current reform movement is a desire for greater equity, and therefore the fact that Maryland teachers report large increases in expectations less frequently for low-achieving student than for high-achieving students warrants concern and further investigation. Expectations are only one aspect of equity, however, and additional investigation would be needed to determine whether teachers' perceptions of changes in expectations are mirrored by changes in actual opportunities to learn. If MSPAP is in fact not sufficiently improving opportunities for low-achieving students, additional research could help guide policy by clarifying the reasons. For example, the high level of Maryland's lowest performance standard may have induced some educators to focus their efforts on higher performing students at the expense of those far below that standard. Alternatively, perhaps higher achieving students on average responded favorably to the new and more difficult work, while many low-achieving students became yet more discouraged; this might lead staff to raise expectations more for higher achieving students. If the former is found to be true, MSDE could modify the system to give schools credit for raising the performance of students who nonetheless fail to reach that standard, perhaps by instituting a "partially satisfactory" standard at a lower level. If the latter is true, professional development could help teachers design instruction for low-achieving students that is consistent with the goals of the program but not so difficult as to be demoralizing.

Writing in Instruction and Assessment

Consistent with the goals of the program, MSPAP appears to have caused a substantial increase in the amount of writing students do in the classroom. Teachers reported not only that students spend more time writing, but also that they are better able to explain their answers. However, teachers' responses also suggest there is a negative side to this change: Many maintained that there is too much emphasis on writing in MSPAP and that instruction has suffered because of the amount of time consumed by students' writing. In addition, virtually all of the

surveyed teachers believe that MSPAP's emphasis on writing makes it difficult to judge the mathematical competence of some students.

These findings raise the question of whether the format of the assessment places too much reliance on writing, in terms of both test validity and instructional impact. This question is a matter of policy as well as empirical evidence, but the results of this survey suggest an apparent need to revisit the issue. With respect to validity, the question of whether the emphasis on writing obscures the mathematical competence of some students depends in part on Maryland policy makers' definition of the domain of mathematics achievement. For example, if mathematical communication is weighted very highly in that definition relative to, say, knowledge of algebraic techniques, designing the assessment to place heavy emphasis on writing may improve the overall validity of inferences about mathematical performance even if measurement of some specific aspects of mathematics suffers in the case of some students. With respect to instructional impact, whether the current emphasis on writing is excessive depends in part on a policy judgment about the value of both the marginal time allocated to writing and the time taken away from other activities to make way for it. Finally, there may be trade-offs *between* impact and validity as well; policy makers may decide that the need for students to develop greater skills in writing in all subjects warrants some decrement in the validity of certain inferences based on MSPAP. The responses of teachers noted here suggest an apparent need to conduct other forms of empirical research to clarify the present trade-offs in terms of both validity and impact and to determine how well the system is meeting the intent of Maryland policy makers.

Other Effects on Instruction

Apart from the increase in writing, these surveys suggest both positive and negative effects on instruction, and both warrant further exploration. For example, most observers will consider teachers' reportedly increased emphasis on problem solving as good news. "Problem solving" is a broad term, however; people are more likely to agree that it is desirable than about what it entails. These reports undoubtedly mask a wide variety of actual changes in instruction, some more desirable than others. More detailed investigation would be needed to pin down the nature of these instructional changes and how they vary among teachers and schools; that information would in turn help better hone professional

development efforts. On the negative side, more information is also needed about the trade-offs entailed by the changes teachers cite, such as decreases in time allocated to various activities and types of material.

The Potential for Inflated Scores

It is critically important, both for program improvement and as a matter of public accountability, to investigate further the findings that suggest the potential for inflated scores on MSPAP. MSPAP, like most assessments of achievement, is intended to represent students' mastery of broad domains of knowledge and skills, and gains on MSPAP are valuable only to the extent that they signal improved mastery of those domains. Such improvements would not be specific to MSPAP, but would generalize to other assessments developed to similar test specifications. If a sizable portion of MSPAP gains were limited to the specific assessment, the validity of the most important inferences based on MSPAP would be undermined. Educators, policy makers, and members of the public who drew the inference that the gains reflect improved outcomes would be misled, and schools' responses to the reforms could be misdirected.

Research has documented that excessive test preparation and severe inflation of scores sometimes results from using traditional tests for accountability (e.g., Koretz et al., 1991; Shepard & Dougherty, 1991). Indeed, that risk is now widely accepted and forms one basis for the current enthusiasm for replacing traditional tests with various forms of performance assessment. Some observers have maintained, however, that performance assessments used for accountability are vulnerable to the same problem (e.g., Koretz, forthcoming). Evidence on this point is scarce, but recent evidence from the Kentucky performance assessment program suggests that some of the initial rapid gains it produced were misleading (Hambleton et al., 1995).

Nonetheless, the survey data reported here can only raise a warning flag; the extent to which MSPAP score gains represent meaningful gains in performance remains an empirical question that can be confirmed only by other forms of research. Moreover, the effects of some amount of familiarization and test preparation may be desirable.

Sizable score gains caused by increased familiarity are common during the first years of testing programs; whether they should be considered "real" or "inflated" depends on the circumstances and the inferences the test is used to

support. For example, suppose a decision is made that certain types of complex problem solving are important outcomes of mathematics instruction and should be given much more weight on a new test than on an old one. If students and teachers learn what types of problem solving are valued as outcomes as a result of familiarity with the test and students score higher as a result, that increase in scores would be considered a real gain. In contrast, if scores improve because students and teachers learn which of several alternative formats or which subsets of valued outcomes are likely to be emphasized and focus their efforts on those particular things, the meaningfulness of the resulting gains is questionable. It was for this reason that our surveys asked about the impact of improvements in the knowledge and skills emphasized by MSPAP (as well as broad improvements in knowledge and skills). The fact that relatively few educators reported that even those more focused improvements contributed a great deal to score gains in their schools is a warning flag that suggests a need for further research.

Moreover, some score gains attributable to familiarity might represent an increase in validity even when they do not represent commensurate gains in mastery. Because MSPAP represented such a large change in the content and format of MSDE's assessments, it is plausible that scores in the first years of the assessment were misleadingly low because of students' lack of familiarity with aspects of the assessment largely unrelated to the constructs it is supposed to measure. Teachers' lack of familiarity could have had similar effects; for example, there were reports of proctors the first year of the assessment who were not sufficiently familiar with assessment materials. As students become more familiar with those aspects of the test, scores would be expected to rise and to become a more accurate indicator of what students know.¹⁴

On the other hand, familiarity may also enable teachers to engage in forms of test preparation that can inflate scores—for example, tailoring instruction so closely to the details of the assessment that the resulting gains are too specific to the test to represent meaningful improvements in the skills the test is supposed to measure. The potential for overemphasis on narrow test preparation that can substantially inflate scores is greater when stakes are high, because incentives to raise scores per se are stronger. In addition, when the initial effects of familiarity

¹⁴ Some workers in educational measurement use the term “familiarity” primarily in this way, that is, to refer to students' learning about construct-irrelevant aspects of a test. Therefore, they use “gains caused by familiarity” to refer primarily to the increase in validity that familiarity can cause.

have run their course, some teachers may increase their reliance on narrow forms of test preparation in an effort to maintain the gains they experienced in the first years of the assessment's use.

The question for both policy and research is therefore more difficult than whether some share of MSPAP gains can be attributed to familiarity or test preparation. Rather, the key question is what share of gains represents meaningful improvements that would generalize beyond the specifics of the MSPAP assessment and, conversely, what share is test-specific. In addition, it would be valuable to obtain information on variations in test preparation and score inflation. For example, if appreciable score inflation is present, it would be important to know whether it is more severe in certain types of schools (e.g., those with initially low achievement); whether scores are more accurate in some subjects than others;¹⁵ and which forms of instruction and test preparation are related to score inflation.

Thus these findings appear to warrant further research and perhaps implementation of a routine audit mechanism to explore possible inflation of scores and its correlates. If such information revealed inflated scores, one could design program modifications in an effort to make undesirable test preparation less common or less effective. These might include, for example, changes in the specificity of curriculum frameworks, new guidelines distinguishing between appropriate and inappropriate test preparation, and perhaps modifications to the assessment itself, such as changes in content coverage or in the sampling of task types.

Other Issues of Validity

Several of the findings reported here point to a need for additional studies of the validity of MSPAP, apart from the essential validity question of possible score inflation. Respondents raised concerns about the characteristics of the assessment itself that could affect the validity of many important inferences. Examples include concerns about inconsistency of content representation and scoring standards over time and about tasks that may be developmentally inappropriate. They also raised concerns pertaining to the assessment's use rather than its characteristics. For example, many pointed to possible distortions

¹⁵ Previous research suggests that inflation of scores may be more prevalent in mathematics than in reading (Koretz et al., 1991).

of inferences about school performance and improvement stemming from irrelevant factors such as students' backgrounds and transience. (These distortions, if they are present, would be similarly germane if MSPAP were replaced with a traditional, standardized, multiple-choice test used for similar purposes.) Such responses also raise concerns about "consequential validity"—that is, the possibility that the changes in education induced by the program may be less consistently positive than intended.

Validation usually requires numerous types of information, and it is an even more complex task in the case of assessments such as MSPAP that use innovative performance-assessment formats, serve multiple functions, and are designed to change instruction (see, for example, Linn, Baker, & Dunbar, 1991; Messick, 1995). Traditional forms of validity evidence—for example, convergent and divergent evidence about the relationships between MSPAP scores and other measures—will be essential, but it may also be important to use less commonplace techniques. For example, protocol analysis may be helpful for ascertaining the developmental appropriateness of tasks. Ascertaining the validity of hybrid tasks performed partly in groups and partly alone may also require innovative techniques (see, for example, Webb, 1993).

Next Steps

Maryland is widely seen a leader in the use of innovative assessments as a lever for raising standards and reforming public education. These surveys suggest that its efforts are reaping rewards, but they also identified potentially serious obstacles and negative effects. The present results may help guide midcourse corrections, and they point to additional investigations that could prove invaluable as the state works to improve the effectiveness of the program and to minimize its unintended adverse effects.

REFERENCES

- Afflerbach, P., Guthrie, J., Schafer, W., & Almasi, J. (1994, April). *Barriers to the implementation of a statewide performance program: School personnel perspectives*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Druker, S. L., & Shavelson, R. J. (1995, April). *Assessment reform from both sides of the fence: Reformers' expectations and teachers' reports of impacts on classroom practice*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Guthrie, J. T., Schafer, W. D., Afflerbach, P., & Almasi, J. (1994, April). *District-level policies of reading instruction in Maryland and their relation to the statewide performance assessment*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Hambleton, R. K., Jaeger, R. M., Koretz, D., Linn, R. L., Millman, J., & Phillips, S. E. (1995). *Review of the measurement quality of the Kentucky Instructional Results Information System, 1991-1994*. Frankfort: Office of Education Accountability, Kentucky General Assembly.
- Koretz, D. (1996, forthcoming). Using student assessments for educational accountability. In R. Hanushek (Ed.), *Improving the performance of America's schools*. Washington, DC: National Academy Press.
- Koretz, D., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991, April). The effects of high-stakes testing: Preliminary evidence about generalization across conventional tests. In R. L. Linn (Chair), *The effects of high stakes testing*. Symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Maryland School Performance Program. (1990). *Learning outcomes in mathematics, reading, writing/language usage, social studies, and science for Maryland School Performance Assessment Program*. Baltimore: Maryland State Department of Education.
- Maryland School Performance Program Report, 1994*. Maryland School Performance Program Office. (1994). Baltimore, MD: Maryland State Department of Education.
- McLaughlin, M. W. (1990). The RAND change agent study revisited: Macro perspectives and macro realities. *Educational Researcher*, 19(9), 11-16.

- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.
- Shepard, L. A., & Dougherty, K. C. (1991, April). Effects of high-stakes testing on instruction. In R. L. Linn (Chair), *The effects of high stakes testing*. Symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10(4), 7-11.
- Stecher, B. M., & Mitchell, K. J. (1995). *Portfolio driven reform; Vermont teachers' understanding of mathematical problem solving* (CSE Tech. Rep. No. 400). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.
- Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Webb, N. W. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment*, 1(2), 131-152.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Yen, W. M., Candell, G. L., Ercikan, K., Green, D. R., Roudabush, G., & Baum, K. (1992). *Final technical report: Maryland School Performance Assessment Program, 1991*. Monterey, CA: CTB/Macmillan/McGraw-Hill.