DOCUMENT RESUME

ED 434 927 TM 030 179

AUTHOR	Lane, Suzanne; Parke, Carol S.; Stone, Clement A.
TITLE	Consequences of the Maryland School Performance Assessment Program.
SPONS AGENCY	Department of Education, Washington, DC.
PUB DATE	1999-00-00
NOTE	56p.; Revised version of a paper presented at the Annual Meeting of the National Council of Measurement in Education (San Diego, CA, April 12-16, 1998).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Educational Assessment; Educational Change; *Educational Objectives; Educational Practices; Elementary Education; Mathematics Education; *Mathematics Tests; Middle Schools; *Performance Based Assessment; Questionnaires; State Programs; *Teacher Attitudes; Teacher Surveys; *Teachers; Testing Programs
IDENTIFIERS	*Maryland School Performance Assessment Program; Reform Efforts

ABSTRACT

The extent to which classroom instruction and assessment activities in mathematics are aligned to the State of Maryland's Learning Objectives and the Maryland School Performance Assessment Program (MSPAP) was studied by examining the underlying structure of the mathematics teacher questionnaire developed and administered to teachers in the 1996-97 instructional year. The existence of hypothesized dimensions of teacher attitudes and practices was studied, along with the relationship among MSPAP gains, percentage of reduced or free lunch students, and composite scores on dimensions of the teacher questionnaire. A stratified random sample procedure was used to obtain a sample of 72 elementary and 36 middle schools. Some students were also asked to respond to some of the items from the teacher questionnaire related to current mathematics instruction. Analyses indicate that elementary teachers were more likely than middle school teachers to report that they place greater emphasis on mathematics learning outcomes and reform-oriented problems and that their mathematics instruction has been influenced by the MSPAP. Analyses also suggest that only approximately 15% of instructional tasks reflect the majority of characteristics of the MSPAP tasks. (Contains 17 tables, 8 figures, and 27 references.) (SLD)

Consequences of the Maryland School Performance Assessment Program

by

Suzanne Lane, Carol S. Parke, and Clement A. Stone

University of Pittsburgh

An earlier version of this paper was presented at the annual meeting of the National Council of Measurement in Education, San Diego, 1998. Preparation of this paper was supported by a grant from the U.S. Department of Education, Assessment Development and Evaluation Grants Program (CFDA 84.279-A), for the Maryland Assessment System Project. Our deepest appreciation is extended to the teachers, principals, and students in Maryland for their invaluable time and effort spent on this project. We would also like to thank Jennifer O'Mara, Tracy Parnella, and James Ventrice for their help in coding the mathematics classroom instruction and assessment activities. Our appreciation is also extended to Yoon-A Chin for her assistance.

Objective

A number of states are implementing statewide assessment programs that depend heavily on performance-based assessments (e.g., Kentucky, Maryland). These assessments are considered critical tools in the educational reform movement (Linn, 1993) and are being used for high-stakes purposes such as holding schools accountable to state standards. A prevailing assumption underlying performance-based assessments is that they serve as motivators in improving student achievement and learning, and that they encourage instructional strategies and techniques that foster reasoning, problem solving, and communication (Frederiksen & Collins, 1989; National Council on Education Standards and Testing, 1992).

Given these high expectations for performance-based assessments, the consequences of the uses and interpretations of the assessments need to be addressed, including both negative and positive consequences, intended and plausible unintended consequences (Messick, 1989, 1992; Cronbach, 1988; Koretz, Barron, Mitchell, & Stecher, 1996; Linn, Baker, & Dunbar, 1991). As stated by Linn (1994), "If the argument that validation should include an evaluation of the consequences of the uses and interpretations of assessment results is accepted, then it is not sufficient to provide evidence that the assessments are measuring the intended constructs. Evidence is also needed that the uses and interpretations are contributing to enhanced student achievement and at the same time, not producing unintended negative outcomes (p. 8)." Messick (1992) suggests that "evidence should especially address both the anticipated consequences of performance assessment for teaching and learning as well as potential adverse consequences bearing on issues of bias and fairness (p. 35)".

Researchers are beginning to examine the consequences of assessment programs by using various methods such as surveys of principals and teachers (e.g., Koretz, Barron, Mitchell, & Stecher, 1966; Pomplun, 1997) and focus groups (e.g., Chudowsky & Behuniak, 1977). The purpose of this research program is to examine the impacts of the Maryland State Performance Assessment Program (MSPAP) and the Maryland Learning Outcomes (MLO's) on school curriculum, classroom instruction and assessment practices, student learning, professional development, and students, teachers, and principals beliefs about MSPAP. MSPAP is a performance assessment program for grades 3, 5, and 8 designed to measure school performance and provide information for school accountability and improvement so as to ensure quality education (Maryland State Board of Education, 1995). MSPAP was implemented in the early 1990's to assess student achievement and school performance with respect to the Maryland Learning Outcomes. MSPAP requires students to develop written responses to interdisciplinary tasks that require the application of skills and knowledge to real life problems, and is intended to promote performance-based instruction and classroom assessments.

The research questions are: (1) What are the effects of MSPAP on curriculum; classroom instructional and assessment practices; student learning; professional development activities; school-based decision-making; and student, teacher and principal beliefs and attitudes? and (2) How do the effects vary by content area (mathematics, reading, writing, science, social studies), grade level (on-grades: 3, 5, 8 and off-grades: 2, 4, 7), and school characteristics (percent free or reduced lunch and MSPAP performance)? This study described herein is limited to examining the impact of MSPAP for the 1996-97 instructional year for the mathematics content area in elementary and middle schools in Maryland.

Of particular interest was the relationship among school mathematics performance gains on MSPAP, the percentage of students who received a free or reduced lunch in the schools, which served as a proxy for socioeconomic level, and the effects MSPAP has had on instruction. The differences in the nature and the extent of the consequences of the assessment program for students in grade levels being tested, on-grades (3, 5, and 8), versus students in grade levels not being tested, off-grades (2, 4, and 7), was also of interest. If the intent of the assessment program is to improve student learning for all students regardless if they are being "tested", then it is necessary to examine the consequences for all students.

This study examines the underlying structure of the mathematics teacher questionnaire developed and administered to teachers in the 1996-97 instructional year. Confirmatory factor analyses (CFA;

Joreskgog, 1969; Joreskog & Sorbom, 1979) were conducted to examine the existence of the following hypothesized dimensions: teachers' support for MSPAP, teachers' emphasis on learning outcomes and on reform-oriented problem types in instruction, teachers' change in emphasis on learning outcomes and on reform-oriented problem types in instruction, MSPAP's impact on instruction, and MSPAP-related professional development activities. A multivariate analysis of variance was then conducted to determine the extent to which teachers of the on-grades and off-grades differed on these dimensions. Next, a growth model analysis (c.f., Meredith & Tisak, 1990; McArdle & Epstein, 1987; Muthen, 1994) was conducted to examine the relationship among MSPAP school performance gains, percent free or reduced lunch, and composite scores on the dimensions of the teacher questionnaire. Lastly, mathematics classroom instruction, assessment and test preparation activities were analyzed to provide more direct evidence of the nature of classroom instruction and assessment activities. Specifically, these analyses were done to determine the extent to which the classroom instruction and assessment activities were aligned to the Maryland Learning Objectives and MSPAP.

Method

School Sample

A stratified random sampling procedure was used to select the schools, with the strata being defined by three levels of each of the following: (a) percent free or reduced lunch according to the 1994-95 classification and (b) MSPAP performance gains (MSDE's 1993-95 change index). Schools were classified into one of the nine cells based on their rankings in the distributions for these two variables. Eight elementary schools from each of the nine cells were sampled and four middle schools from each of the nine cells were sampled. A total of 72 elementary and 36 middle schools were selected to participate in the study with alternate schools identified as potential replacements for schools who chose not to participate. A larger number of elementary schools were selected because, compared to the middle schools, they have fewer teachers per grade.

The final sample consisted of 59 elementary and 31 middle schools, with a total of 90 schools. Thus, the school participation rate was 82% for elementary schools and 86% for middle schools. There were an approximately equal number of schools within each of the nine classification cells. Of the 59 elementary schools, 42 were from the initial 72 that were sampled, and of the 31 middle schools, 22 were from the initial 36 that were sampled. The remaining schools were from the list of alternate schools for each cell. This represents schools from 19 systems/counties in Maryland. It should be noted that, because schools were unable to be contacted until January 1997 regarding their participation in the study, the sample size for the 1996-97 instructional year was reduced.

Instruments

To triangulate on the consequences of MSPAP, multiple measures were used. The data sources used for this study were questionnaires and samples of classroom instruction, assessment, and test preparation materials. Questionnaires were developed for principals, teachers, and students. The principal questionnaire was the same for both elementary and middle school principals. Separate mathematics questionnaires were developed for 2nd, 3rd, 4th, 5th, 7th, and 8th grade teachers. The teacher questionnaires did not vary substantially across on- and off- grades (i.e., tested and not tested grades, respectively). Mathematics questionnaires were developed for students in 4th, 5th, 7th, and 8th grades. The questionnaires for the 4th and 7th grade (i.e., off-grade) students contained a MSPAP public release task so that the students could examine the task prior to responding to questions pertaining to MSPAP-like tasks.

The questionnaires consisted of both likert and constructed response items. Some of the likert items were in the form of questions, and others were statements. In general, a four-point scale was used for the likert items. To triangulate on the consequential evidence, students, teachers, and principals responded to similar questions for areas in which it was deemed appropriate. The areas on the teacher questionnaire included the following: familiarity with MSPAP, support for MSPAP, beliefs about MSPAP, overall impact of MSPAP, the nature of instruction and classroom assessments, MSPAP's impact on instruction and classroom assessments, the nature of professional development activities, and MSPAP's impact on professional development activities. The principal and student questionnaires included items for areas that were deemed appropriate. Some of the ideas for questions pertaining to the support for MSPAP and the beliefs about MSPAP were based on a previous study examining the consequential evidence of state assessments (Koretz, Mitchell, Baron, & Keith, 1996). The instruments were piloted in the spring of 1996 in schools in Maryland and were reviewed by Maryland mathematics teachers.

Data collection forms were developed for a subset of the teachers in both the off-grades (2^{nd}, 4^{th}, 7^{th}) and on-grades (3^{rd}, 5^{th}, 8^{th}) who provided classroom materials. Teachers were asked to provide 10 instruction tasks and 10 assessment tasks that were representative of their classroom materials across the school year. They were also asked to provide an example scoring scheme and an example test preparation activity. The data collection forms asked teachers to indicate the nature of the students' ability levels in the mathematics class from which the materials were obtained (e.g., heterogeneous ability group, homogeneous ability group, exclusively special education). The forms also asked teachers to indicate the nature of the mathematics taught in the class (e.g., general math, pre-algebra, algebra).

Data Collection

Teachers and principals were asked to complete their respective questionnaires during February 1997. Students were administered the student questionnaire within the two weeks following the administration of MSPAP, that is, in either the 3^{rd} or 4^{th} week of May 1997.

Teachers were asked to send in approximately 5 mathematics instruction activities, 5 mathematics assessment activities, and 1 sample of a scoring scheme used from September to December 1996. In the spring they were asked to send in another set of 5 instruction activities, 5 assessment activities, and 1 sample of a scoring scheme used from January to June 1997. In addition, they were also asked to send a sample of a MSPAP test preparation activity used prior to the administration of MSPAP. If teachers taught more than one mathematics class, one of their classes was randomly selected for the collection of the materials.

Questionnaire and Classroom Materials Return Rate

Principal and Teacher Questionnaire. Of the 90 principals, 86 completed the principal questionnaire, resulting in a response rate of 96%. A total of 515 2nd, 3rd, 4th, 5th, 7th, and 8th grade teachers out of 594 completed the teacher questionnaires, resulting in a response rate of 87%. The number of mathematics teachers in each grade level that completed the questionnaires are 79 2nd grade teachers, 98 3rd grade teachers, 77 4th grade teachers, 99 5th grade teachers, 62 7th grade teachers, and 100 8th grade teachers.

Student Questionnaire. Each of the 4th, 5th, 7th, and 8th grade teachers participating in the study were asked to administer the student questionnaire to one of their classes. Overall, 115 of the 163 elementary classes (4th and 5th grades) that were identified for the administration of the mathematics student questionnaires actually administered the questionnaires, resulting in a return rate of 71%. In the middle school classes (7th and 8th grades), 95 of the 148 identified classes administered the mathematics student questionnaires (64%). Table 1 indicates the number of students and classes in 4th, 5th, 7th, and 8th grades who completed the mathematics questionnaires. It should be noted that each of the questionnaires was divided into 3 forms and a student received only one form. The forms were randomly distributed within each of the participating classrooms. This sampling design was used to reduce the amount of time taken away from instruction.

[Insert Table 1]

Classroom Activities. A subset of schools was asked to participate in the collection of the classroom activities. Overall, 51 schools were asked to participate in this aspect of the study. Some or all of the teachers from 44 of the schools participated, resulting in a school participation rate for classroom activities of 86%. This represented schools from 15 different system/counties in Maryland. Of the 332 mathematics teachers that were asked to participate, 250 provided the materials (75%).

Description of Principals, Teachers, and Students who Completed Questionnaires

Principals. Principals were asked to indicate the number of years they had served as an administrator in a Maryland elementary or middle school. Of the elementary principals, approximately 28% had 1-5 years, 58% had 6-15 years, and 14% had 26 years or more of experience as an administrator. Of the middle school principals, approximately 30% had 1-5 years, 40% had 6-15 years, and 30% had 26 years or more of administrative experience in Maryland.

Teachers. Teachers provided information regarding the total number of years they had taught in a school in Maryland. Overall, approximately 34% of the teachers had 1-5 years, 29% had 6-15 years, and 37% had 16 or more years of experience teaching in Maryland. An examination of the results at each grade level indicated slightly larger percentages of new teachers in the middle school grades.

Students. Approximately 50% of the students responding to the mathematics questionnaires were female and 50% were male. This was similar across all grade levels. Students were also asked to indicate their ethnicity. The majority of students (about 70%) indicated Caucasian, approximately 20% indicated African-American, and a very small percentage indicated Hispanic, Asian American, or other.

Description of Classes and Teachers who Collected Classroom Activities

A total of 250 mathematics teachers sent in a sample of their mathematics classroom activities used during the 1996-97 school year. Teachers were asked to indicate the type of math class from which their sample of classroom activities was selected Ninety-eight percent of the elementary classes were "general math" classes, while only 42% of the middle school classes were "general math" classes. The remaining middle school classes were either prealgebra classes (39%) or algebra classes (15%).

On average across the entire school year, approximately 16 classroom instruction and assessment activities were collected per teacher. In the fall, 236 mathematics teachers sent in 10 classroom activities on average, and in the spring, 163 mathematics teachers sent in 10 classroom activities on average. For each grade, Table 2 indicates the number and percentage of teachers who sent in classroom activities and also the total number and percentage of all classroom activities received. For example, 39 2nd grade teachers sent in a total of 591 classroom activities. The percentages across grades for the number of teachers and the number of activities are somewhat similar, although a slightly smaller percentage of off-grade teachers (2, 4, and 7) than on-grade teachers (3, 5, and 8) sent in classroom activities.

[Insert Table 2]

Teachers were provided with labels to attach to each activity indicating the type of activity (e.g., instruction, assessment, test preparation, scoring scheme). Table 3 shows the number and percentage of activities for each type. Across all grades (2, 3, 4, 5, 7, and 8), there was a total of 1940 instruction activities, 1388 assessment activities, and 332 scoring schemes. For grades 3, 5, and 8 there was a total of 125 MSPAP test preparation activities. The table also includes a category called "not coded". These were activities that were not coded for one of two reasons. One reason for not coding an activity was because it pertained strictly to another content area such as social studies or science. Another reason an activity was not coded was because it consisted only of teacher notes or general lesson plans, and it was difficult to discern what the students were required to do. The percentages across grade levels for each of the types of activities were somewhat similar, although slightly more on-grade teachers than off-grade teachers sent in classroom activities.

[Insert Table 3]

Teachers were also asked to indicate the source of each activity. Over half of the instruction activities (57%) were from textbook or commercial resources and 25% were teacher-developed. Approximately equal percentages of the assessment activities were from textbook/commercial resources or were teacher-developed (36% and 38% respectively). While the percentage of activities that were county-developed was quite small for instruction and assessment activities, there was a slightly larger percentage of assessment activities (15%) than instruction activities (8%). The percentage of instruction and assessment activities obtained from state-level materials, such as MSPAP Release Tasks, Maryland Consortium Tasks, and Maryland Performance-Based Exemplars was very small. The results across grades were similar.

When examining the MSPAP test preparation activities, the sources were somewhat different than for the instruction or assessment activities. The percentage of teacher-developed activities was similar (33%), however, there was a larger percentage that were county-developed (26%), MSPAP Release

Tasks (4%) and other state-level materials (10%). The sources for the scoring schemes were similar to the test preparation activities. About 35% were teacher developed, 20% were county-developed, and 9% were state-level materials.

Rater agreement for coding the classroom activities. A total of four raters coded the classroom activities. A formal training session was conducted to familiarize the raters with the coding scheme using a sample set of pre-coded activities. Then, the raters coded another set of sample activities independently and their codes were compared and discussed by the group. After the formal training was complete, pairs of raters individually coded a set of classroom activities from a school (elementary or middle) for a certain collection period (fall or spring). The pair of raters met to discuss their discrepancies and reached a consensus on the codes for each activity. This was done to ensure that all raters shared a common understanding of the coding scheme. Thus, for a small percentage of classroom activities (7%), one set of codes, agreed upon by two raters, was obtained.

After it was determined that the raters reached a shared understanding of the coding scheme and were proficient in applying it to a variety of classroom activities, each rater individually coded sets of classroom activities. Approximately 20% of the sets of classroom activities (an elementary or middle school teacher's activities from either fall or spring) were coded individually by two raters. The overall adjusted rate of agreement between the raters was then calculated[1]. The adjusted rate of agreement was found to be 84% for the instruction, assessment, and test preparation activities and 81% for the scoring schemes. In addition to examining the agreement between rater pairs, the accuracy of raters' codes was examined for 23% of the sets of classroom activities. This was accomplished by comparing a rater's set of codes with codes obtained by the lead rater who had been involved in the conceptualization and

[1] Percent agreement was considered to be too lenient of an index of rater agreement because for a number of the categories to be coded there were a range of options that could be selected. As an example, for the content learning outcome, one to eight content outcomes could be selected for an activity. However, the majority of the activities had between one to three content outcomes coded. A simple percent agreement based on each of the eight outcomes would have inflated the index for rater agreement. Thus, an adjusted percent agreement was used.

development of the coding scheme. The adjusted rate of agreement was 87% for the instruction, assessment, and test preparation activities, and 74% for the scoring schemes.

Confirmatory Factory Analysis for the Teacher Questionnaire

Confirmatory factor analyses (CFA; Joreskog, 1969; Joreskog & Sorbom, 1979) were used to examine an hypothesized structure underlying the teacher questionnaire. The teacher questionnaire was designed to provide information about six dimensions. The six dimensions are teachers' familiarity with MSPAP, teachers' support for MSPAP, teachers' instruction and assessment practices, change in teacher's instruction and assessment practices, MSPAP's impact on instruction, and professional development support for teachers. Subsets of items were grouped according to the following 11 areas (i.e., measures) to reflect the six dimensions:

(1) MSPAP Familiarity – General (teachers' general familiarity with MSPAP),

(2) MSPAP Familiarity – Results (teachers' familiarity with MSPAP results),

(3) Support MSPAP – General (teachers' general support for MSPAP),

(4) Support MSPAP – Instruction (teachers' support for MSPAP for instructional purposes),

(5) Current Math Instruction/Assessment - LO (emphasis on learning outcomes in instruction and assessment),

(6) Current Math Instruction/Assessment - PT (emphasis on reform-oriented problem types in instruction and assessment),

(7) Change Math Instruction/Assessment - LO (change in emphasis on learning outcomes in instruction and assessment),

(8) Change Math Instruction/Assessment - PT (change in emphasis on reform-oriented problem types in instruction and assessment),

(9) MSPAP's Impact (MSPAP's impact on instruction and assessment),

(10) Professional Development Support - MSPAP (professional development activities related to MSPAP), and

(11) Professional Development Support - Amount (amount of professional development activities).

Teacher mean scores were obtained for each of these eleven subsets of items in order to minimize the number of parameters to be estimated. The majority of the items on the questionnaire had a four-point Likert scale. For those items that had more than a four-point scale, the responses were recoded to a four-point scale. Teacher data were excluded for those cases in which teachers had left blank more than 25% of the items on any one of the eleven subsets of items. Based on the intercorrelations among the items and the item-to-total score correlations, a small number of items were deleted from their respective subsets. For example, a few items were deleted due to low item-to-total score correlations. Figure 1 provides the final set of items for each of the subsets and the hypothesized dimension underlying each of the subset of items. Coefficient alpha reliability estimates for these 11 subsets (i.e., measures) for both on- and off-grade data sets ranged from .74 to .93.

[Insert Figure 1]

Maximum likelihood estimates for parameters of three hierarchical models were obtained using AMOS (Arbuckle, 1997). Two sets of analyses were conducted. The first set excluded the two teacher mean scores, Change Math Instruction-Learning Outcomes and Change Math Instruction-Problem Type, whereas, the second set of analyses included these two scores. Teachers answered the questions with respect to instructional change only if they taught in Maryland since the 1992-93 school year. Thus, the first set of analyses is based on a smaller sample size than the second set of analyses.

For the analyses excluding the instructional change measures, the first model that was estimated provided a test for the hypothesis that one factor accounted for the interrelations among the teacher mean scores for the nine measures. The second model that was estimated provided a test for the hypothesis that four factors accounted for the interrelationships as specified in Figure 2. The third model that was estimated, the hypothesized model, provided a test for the hypothesis that five factors accounted for the interrelationships as specified in Figure 2. For the analyses including the instructional change measures similar models were estimated as shown in Figure 2; however, the third model included six factors so that one factor would reflect the two instructional change measures.

[Insert Figure 2]

The analyses were done for the on-grade levels (3, 5, 8) combined and the off-grade levels (2, 4, 7) combined to determine whether the structure differed for on- and off-grade teachers. The sample sizes for the analyses excluding the instructional change measures were 254 for the on-grade and 172 for the off-grade. The sample sizes for the analyses including the instructional change measures were 178 for the on-grade and 112 for the off-grade.

Analyses Excluding the Instructional Change Measures

For the on-grade analyses excluding the instructional change measures, the one-factor model and the four-factor model did not fit the data as evidenced by the significant chi-square statistics presented in Table 4. The five-factor model, the hypothesized model, fit the data as evidenced by the nonsignificant chi-square statistic. Only one covariance among the factors was not significant and it was for the relationship between Support MSPAP and Current Math Instruction.

[Insert Table 4]

These analyses were also conducted for the off-grade levels (2, 4, 7), combined. Three similar models, excluding the instructional change measures, were estimated to determine whether the underlying structure of the teacher questionnaire was similar for the on- and off- grades. The five-factor model for the off-grade levels, which excluded the instructional change measures fit the data as evidenced by the nonsignificant chi-square statistic in Table 4. All the covariances among the factors were significant.

A third set of analyses was conducted to determine whether the parameters could be constrained across the on- and off-grades for the five-factor model. The results are provided in Table 4. The difference chi-square of 36.407 with 27 df was not significant ($p = .107$), indicating that the additional parameters estimated under the unconstrained model did not improve on model data fit as offered by the constrained model. Thus, the parameters could be constrained across the two groups. Table 5 provides the unstandardized regression coefficients, their standard errors, and the significance tests for the five-factor model with the parameters constrained across the on- and off-grades. The 1's in the column for the unstandardized regression coefficients denote the necessary constraints to attain model identification.

[Insert Table 5]

Analyses Including the Instructional Change Measures

Similar results were found for the on-grade analyses that included the instructional change measures. The one-factor model and the four-factor model did not fit the data as evidenced by the significant chi-square statistics in Table 6. The six-factor model, the hypothesized model, fit the data as evidenced by the nonsignificant chi-square statistic. The only covariance among the factors that was not significant is for the relationship between Support MSPAP and Current Math Instruction.

[Insert Table 6]

These analyses were also conducted for the off-grade levels (2, 4, 7) combined. Three similar models, including the instructional change measures, were estimated to determine whether the underlying structure of the teacher questionnaire was similar for the on- and off-grades. Similar to the on-grade levels, the one- and four-factor models for the off-grades did not fit the data as evidenced by the significant chi-square statistic in Table 6. The six-factor model for the off-grade levels did fit the data as evidenced by the nonsignificant chi-square statistic in Table 6. All of the covariances among the factors were significant.

Another set of analyses was conducted to determine whether the parameters could be constrained across the on- and off-grades for the six-factor model, including instructional change. The results are provided in Table 6. The difference chi-square of 59.107 with 36 df was significant (p=.009), indicating that the additional parameters estimated under the unconstrained model improved on model data fit. Thus, the parameters cannot be constrained across the two groups. Table 7 provides the unstandardized regression coefficients, their standard errors, and the significance tests for the six-factor model for the on-grade levels and the off-grade levels.

[Insert Table 7]

In general, these results suggest that the underlying structure of the teacher questionnaire items for the off-grade levels is similar to the structure for the on-grade levels when excluding the instructional change measures. When including the instructional change measures, the factor structure for the on- and off-grade levels is similar, however the relationship between the measures and the factors differ across the on- and off-grades to some extent.

Results

Multivariate Analysis of Variance for the Questionnaire Data

Results for the Teacher Questionnaire. The teacher questionnaire data were analyzed with a one-way multivariate analysis of variance, with the between-subjects effect being the grade and the dependent measures being the teacher composite mean scores on the dimensions. The dimensions are MSPAP Familiarity, Support MSPAP, Current Math Instruction, Change Math Instruction, MSPAP Impact on Instruction, and Professional Development Support. Descriptive data for the dependent measures are provided in Table 8. The range on the questionnaire item scale is 1 - 4, with the more positive responses being at the upper end of the scale. Overall, the mean scores were at the upper end of the score scale.

[Insert Table 8]

The multivariate test was significant at $p < .001$ (Wilkes' Lambda, F (18, 795) =3.568, $p < .001$.). Table 9 provides a summary of the results of the univariate analyses. As indicated in the table, there were significant grade differences for five of the dimensions: MSPAP Familiarity, Current Math Instruction, Change Math Instruction, MSPAP Impact on Instruction, and Professional Development Support.

[Insert Table 9]

Tukey HSD post-hoc analyses were conducted to determine, for each of the five dependent measures, which differences between composite mean scores were significant. Table 10 provides the results of the post-hoc analyses. In general, an examination of the table indicates that composite mean scores for elementary on-grade teachers were significantly greater than composite mean scores for middle on- and off-grade teachers. For example, elementary on-grade teachers, as compared to middle on- and off-grade teachers, were more likely to indicate that they place a greater emphasis in their mathematics classrooms on the learning outcomes and reform oriented problem types as evidenced by the composite mean differences for the dimension, Current Math Instruction. Elementary on-grade teachers, as compared to middle on-grade teachers, were also more likely to indicate that their emphasis on the learning outcomes and reform oriented problem types is greater than what it was a few years ago as evidenced by the mean differences for the variable, Change Math Instruction. Further, elementary on-grade teachers, as compared to middle on- and off-grade teachers, were more likely to indicate that MSPAP had a greater impact on their mathematics instruction and that they had received more professional development support regarding MSPAP as evidenced by the mean differences for the dimensions, MSPAP Impact and Professional Development Support, respectively. As indicated in Table 9, however, the adjusted $\underline{r}^2$ value is relatively small for each of the significant variables indicating that grade accounts for only a small percentage of the variance.

[Insert Table 10]

There were few differences between mean scores for elementary on- and off-grades and when these differences occurred they were small. For example, elementary on-grade teachers, as compared to elementary off-grade teachers, were more likely to indicate that their emphasis on the learning outcomes and reform oriented problem types is greater than what it was a few years ago, as evidenced by the mean differences for the variable, Change Math Instruction. However, the mean difference was small (e.g., .151 , p = .049).

In summary, elementary on-grade teachers as compared to middle on- and off-grade teachers indicated that their instruction was more aligned to the content and format of MSPAP and that they have had more professional development support related to MSPAP. Further, there were only a few differences between elementary on- and off-grade teacher results and no difference between middle on- and off-grade teacher results. Thus, although there were differences between elementary and middle school teachers, within school type, teachers who taught grades that were not administered MSPAP responded similarly to teachers who taught grades that were administered MSPAP.

Results for the Principal and Student Questionnaire. Elementary and middle school principals were asked to respond to some of the same items as in the teacher questionnaire. Table 11 provides elementary and middle school principal mean scores on four of the dimensions discussed above: MSPAP Familiarity, Support MSPAP, MSPAP Impact, and Professional Development Support. This table also provides corresponding mean scores for the teachers. It should be noted that the mean scores for the teachers in this table are somewhat different than the mean scores provided in Table 8. This is because the scores in Table 11 are based on only the items that were the same for the principals and the teachers. For the dimensions, MSPAP Familiarity and Support MSPAP, the items were the same for both teachers and principals. For the dimensions, MSPAP Impact and Professional Development Support, the principals had fewer items than the teachers and consequently the teacher means in Table 11 are based on a smaller number of items than those reported in Table 8.

[Insert Table 11]

A one-way multivariate analysis of variance was conducted on the principal data, with the between-subjects effect being the school type and the dependent measures being the composite mean scores on the four dimensions of the principal questionnaire. The multivariate test was not significant (Wilkes' Lambda, $F(4, 77) = 2.245$, $p = .072$). This result suggests that elementary and middle school principals are similar with respect to their familiarity with MSPAP, their support of MSPAP, the extent to which they think MSPAP has had an impact on instruction, and the extent to which they think their teachers received professional development support related to MSPAP.

In general, the principal composite mean scores were higher than the teacher composite mean scores on the dimensions as indicated in Table 11. A one-way multivariate analysis of variance was conducted, with the between-subjects effect being teacher/principal and the dependent measures being the composite mean scores on the four dimensions of the principal questionnaire. The multivariate test was significant (Wilkes' Lambda, $\underline{F}$ (4, 367) = 16.510, p < .001). Table 12 provides a summary of the results of the univariate analyses. All univariate tests were significant. Both elementary and middle school principals, as compared to elementary and middle school teachers of mathematics, indicated that they were more familiar with MSPAP, that they were more supportive of MSPAP, that MSPAP had a greater impact on classroom instruction, and that teachers received more professional development support related to MSPAP. It should be noted, however, that the adjusted $\underline{r}^2$ values were relatively small.

[Insert Table 12]

Students in 4^{th}, 5^{th}, 7^{th}, and 8^{th} grade were also asked to respond to some of the same items as in the teacher questionnaire related to the dimension, Current Math Instruction. Class composite mean scores for each of the grades were obtained on this dimension and are provided in Table 11. A one-way univariate analysis of variance, with the between-subjects effect being the grade level was conducted on the class data. The univariate test was significant ($\underline{F}$ (3) = 7.841, p < .000, $\underline{n}$ = 189). Tukey HSD post-hoc analyses were conducted to determine which differences between mean scores were significant. Table 13 provides the results of the post-hoc analyses. As indicated in the table, elementary on-grade students (5^{th}) and off-grade students (4^{th}) were more likely to indicate that a greater emphasis was placed on the learning outcomes and reform-oriented problems than off-grade students (7^{th}). Further, on-grade elementary school students (5^{th}) were more likely to indicate that a greater emphasis was placed on the learning outcomes and reform-oriented problems than middle on-grade students (8^{th}). It should be noted, however, that the mean differences, although significant, are relatively small given the 4-point scale.

Further, the adjusted r^2 value of .089 is relatively small, indicating that approximately 9% of the variance in the Current Math Instruction variable is accounted for by grade level.

[Insert Table 13]

In general, the composite mean scores for classes on this dimension were consistently lower than the teacher composite mean scores. A one-way univariate analysis of variance, with the between-subjects effect being the class/teacher was conducted on the data. The univariate test was significant, $F(1, 376) = 25.367$, $p < .000$. This suggests that teachers, as compared to students, were more likely to indicate that their mathematics classrooms had a greater emphasis on the learning outcomes and reform oriented problem types. Similar to the previous results, the adjusted r^2 value of .061 is relatively small, indicating that approximately 6% of the variance in the Current Math Instruction variable is accounted for by the type of respondent (teacher vs. class of students).

Modeling Differences in School Performance Over Time

Random coefficient or growth modeling was used to examine mathematics performance on MSPAP from 1993 to 1997 in relation to two dimensions from the teacher questionnaire and the school characteristic, percent free or reduced lunch. Only two dimensions, MSPAP Impact and Current Math Instruction, were used because of the relatively small school sample size. In addition, these two dimensions were considered to be more relevant than the other dimensions for examining the relationship between change and teachers' perceptions.

The advantages of using growth curve methodologies to analyze change has been discussed in the literature (c.f., Rogosa & Willet, 1985; Willet & Sayer, 1994; Rogosa, 1987). These methodologies are particularly well suited for studying processes that consider change as continuous with individual differences in the pattern of change (e.g., initial level and rate of growth). Further, these methodologies allow for studying individual differences and identifying factors that affect the trajectory of change. This type of analysis can not be modeled by time-specific comparisons involving group-level (e.g., means) differences.

Figure 3 illustrates the differences in initial mean MSPAP performance and changes in mean MSPAP performance from 1993 to 1997 for the sample of schools in the present study. Since percent free or reduced lunch was found to correlate significantly with 1993 MSPAP math performance, the plots are presented for three subgroups of this variable (i.e., lower 3^{rd}, middle 3^{rd}, and upper 3^{rd}) to reduce the number of lines in any one graph. As can be seen, there are differences among the schools in terms of their initial MSPAP math performance and their change over time. For example, schools in the lower quartile were concentrated in the MSPAP math performance range of 520-540 in 1993 whereas schools in the upper quartile were concentrated in the range of 480-500 in 1993. In all cases the rate of change appears modest.

[Insert Figure 3]

In order to model individual differences in change and assess the correlates or predictors of change, two levels of statistical modeling are required: Level 1 - within individual schools, trends across the repeated measurements are modeled; and Level 2 - across schools, the parameters from the model of individual differences in change at Level 1 are modeled in relation to other factors. At Level 1, growth models were used to analyze the repeated measurements of test scores, analyze the relationship between time (year) and test score levels, and estimate a reference status (intercept) and rate of change (slope) for each school. For example, from Figure 3, it would be expected that schools would differ with regard to their 1993 MSPAP performance (intercept) and their rates of change over time. At Level 2, the parameters from the model at Level 1 (intercepts and slopes) were then modeled in relation to factors that were introduced to explain variation in the intercept and slope parameters across schools (MSPAP Impact, Current Math Instruction, Percent Free or Reduced Lunch).

Growth models can be estimated using a variety of software. Recently, Singer (1999) illustrated the estimation of such models in SAS PROC MIXED. Specialized software is also available (e.g., HLM: Bryk & Raudenbush, 1992). In addition, several researchers have discussed how growth models can be estimated within a structural equation modeling (SEM) framework by considering the intercept and slope factors as latent variables (e.g., McArdle & Epstein, 1987; Meredith & Tisak, 1990; Muthen, 1991; Willet & Sayer, 1994). Muthen and Curen (1997) have further discussed the flexibility in modeling that is afforded by estimating growth models using SEM. In the present study, the growth models were estimated using the SEM program AMOS (Arbuckle, 1997).

Figure 4 presents the Level 1 unconditional latent variable growth model for the present study. This model involves the outcome variable, MSPAP mathematics standard score, measured at five timepoints. In order to translate the growth model into the framework of structural equation modeling, the school-specific random coefficients (intercepts and slopes from Level 1) are each modeled using two latent factors: 1) a factor representing a reference status of MSPAP math performance (intercept), and 2) a factor which corresponds to the rate of change in MSPAP math performance over time (slope). The mean of these factors represent group level estimates (Level 2) of the intercepts and slopes, respectively, and the variance of these factors reflects the school differences or random effects that exist around these group level parameters. Larger variances reflect increased variability or less similarity in intercept and slopes among the schools.

[Insert Figure 4]

As can be seen from the figure, the Level 1 model has the format of a measurement or confirmatory factor analysis model in structural equation modeling with restrictive loadings: $Y = \Lambda\eta + \varepsilon$, where Y are the original measurements over time, η is a vector of latent variables (intercept and slope parameters), Λ is a matrix of regression coefficients relating the slope and intercept factors to the Y measurements, and ε is a vector of residuals representing variance not accounted for due to time specific factors not included in the model or random error. In addition, an association between the intercept and slope factors is assumed and indicated by the curved bi-directional arrow.

The meaning of the intercept factor depends on the scaling of the time variable for the slope factor, and the scaling of the slope factor is determined by the factor loadings or regression coefficients relating the slope factor to the observed measurements. For example, to reflect a simple linear pattern in 1993

MSPAP performance to 1997 MSPAP performance, the regression coefficients could be constrained to be 0, 1, 2, 3, and 4 for the variables. Under this scaling, the intercept could be interpreted as MSPAP initial status of schools since time 0 corresponds to 1993 performance. However, it is also possible to estimate coefficients or constrain the parameters to some other pattern. In Figure 4, the pattern is 4, 3, 2, 1, and 0. Since time 0 is associated with 1997 MSPAP performance, the intercept factor is interpreted as 1997 MSPAP status and a decrease in performance would be expected from 1997 to 1993. This scaling was adopted because other school related information was collected in 1997 and introduced into the analysis to explain variations in the 1997 MSPAP performance and rates of change among schools. The intercept factor will be referred to as 1997 MSPAP performance hereafter.

The structure or distribution of the residuals (Level 1 error model) is defined through constraints on the parameters of the error variance-covariance matrix. The classical assumption of homoscedastic independent errors can be defined by constraining the diagonal elements (variances) of the error variance covariance matrix to be equal over time and off-diagonal elements (covariances) fixed at 0. This assumption can be relaxed by allowing the variances to vary over time and/or estimating a certain pattern to the error variances and covariances (e.g., compound symmetry or adjacent error covariances estimated). In addition, all error variances and covariances can be estimated as in a fully parameterized or unstructured error matrix. In Figure 4, independent but unequal error variances are assumed.

In order to estimate group level estimates of the intercept and slope latent variables for the Level 2 model, means for the latent variable intercepts and slope factors must be estimated. The general covariance structure model accommodates such a parameterization and is often used when analyzing longitudinal data or multiple populations. In order to estimate these types of models, the general covariance structure model includes an intercept term as follows: $Y = \tau + \Lambda\eta + \varepsilon$, where τ is a vector of intercepts and is the E[Y] when $\eta = 0$, and all other model parameters are defined as before. Note that $\tau = 0$ when deviations from means are analyzed.

Table 14 presents the results from estimating the Level 1 model in Figure 4 for 86 schools (1 aberrant pattern of performance over time was detected and deleted for the growth curve analyses). The chi-square statistic for model-data-fit was 8.16 with 9 *df* (p=.52) indicating that the null hypothesis that the variance-covariance matrix implied by the model in the table equals the observed variance-covariance matrix could not be rejected. As can be seen, the 1997 MSPAP performance (intercept factor) across the schools was 521.61 with a significant mean rate of change (slope factor) of –2.70, although the rate of change was modest given the scale of the test scores. Recall that the rate of change is associated with a decrease in performance from 1997 to 1993. Thus, this result suggests that there was a significant increase in performance from 1993 to 1997. The variances for 1997 MSPAP performance and rate of change also indicate significant variability in these parameters across the schools. In addition, the covariance between 1997 MSPAP performance and rate of change was not significant (r = -.05). In order to investigate this last finding further, an analysis in which 1993 MSPAP performance was the reference point was examined. This analysis revealed a significant negative covariance between 1993 MSPAP performance and rate of change (r = -.4[illegible]) indicating that higher rates of change were associated with lower initial performance in 1993. This suggests that the rate of change is more similar for schools in 1997 than in 1993 and this may be due to the observed decrease in variability in 1997 school performance as compared to 1993.

[Insert Table 14]

It should be noted that a non-linear rate of change was estimated in the model. The chi-square difference between a model assuming linear change and the non linear rate of change model described in Table 17 was 3.11 with 1 *df* (p<.10) and the RMSEA (root mean square error of approximation, Browne and Cudeck, (1993) was reduced by .03. Therefore, a non-linear rate of change in the Level 1 model was assumed. The pattern in the regression coefficients in the table indicate that a larger than average change occurred between 1994 and 1995 (estimated coefficient of 1.39 versus a fixed coefficient of 2), followed by a corresponding smaller than expected change from 1995 to 1996.

The structural component of the structural equation model is used to reflect factors that are hypothesized to explain the variability in 1997 MSPAP performance (intercepts) and rates of change (slopes): $\eta = \alpha + \beta\eta + \zeta$; where, η is defined as above, α is a vector of population means for the latent variables, β is a matrix of structural slopes for the effects among endogenous and exogenous η variables (e.g., variables included to explain individual differences in intercepts and slopes), and ζ are structural residuals.

Figure 5 presents the Level 2 (conditional) growth model for the present study. Two dimensions from the teacher questionnaire and the variable, percent free or reduced lunch were introduced into the growth model and paths are included from these variables to the latent variables (1997 MSPAP performance and rate of change). The structural residuals are specified by d1 and d2 in the figure, and the relationship between 1997 MSPAP performance and rate of change is estimated through these two residual parameters. As indicated previously, only two dimensions were introduced since the school sample size is relatively small (n=86). Note that, in theory, it would be possible to incorporate the confirmatory factor analysis model for the teacher questionnaire directly with the growth model rather than use the derived variables for the two dimensions. However, given the sample size in the present study, such a model was overly complex to be estimated.

[Insert Figure 5]

Table 15 presents the regression coefficients for the variables introduced to explain variation in 1997 MSPAP performance and changes in performance over time. The chi-square statistic for model-data-fit was 24.989 with 18 *df* (p=.125) indicating that the null hypothesis that the variance-covariance matrix implied by the model in Table 20 equals the observed variance-covariance matrix could not be rejected. The RMSEA statistic was .068, which is within the acceptable range (Browne and Cudeck, 1993). As can be seen, the variable Percent Free Lunch is significantly related to 1997 MSPAP performance. Thus, increases in the percentage of students receiving free or reduced lunch is associated with lower levels of MSPAP performance in 1997. The only factor that was found to significantly explain variability in rates of change was the teacher questionnaire dimension, MSPAP Impact. This indicates that higher levels of teacher reports of MSPAP having a direct impact on instruction are associated with greater rates of decrease in performance from 1997 to 1993 or higher levels of rate of change in MSPAP school performance. Finally, it is interesting to note that, although increases in the percentage of students receiving free lunch is associated with lower levels of MSPAP performance in 1997, corresponding increases were not significantly associated with rate of change in MSPAP performance over time.

[Insert Table 15]

Mathematics Classroom Activities Results

Each of the teachers' classroom instruction, assessment, and MSPAP test preparation activities were analyzed using a coding scheme designed to provide information about the format of the activities, the extent to which they reflect the Maryland Learning Outcomes (MLO's), and other features of the activities (e.g., response type required of student, integration with other subject areas, etc.). They were also analyzed with respect to how similar they were to MSPAP in general. The Maryland Learning Outcomes and the format and content of MSPAP served as the basis for the coding schemes that were developed for the analysis of the classroom instruction, assessment, and MSPAP test preparation activities. The only results that will be reported herein are those based on the analysis of the classroom activities with respect to their similarity to MSPAP.

Each of the classroom instruction, assessment, and MSPAP test preparation activities were coded with respect to their similarity to MSPAP tasks. In particular, the level of problem solving and reasoning required, the type of responses required of students (e.g., explanations, solution processes), and the format and length of the responses were considered in order to classify the activities according to one or more MSPAP-like levels. The first two levels include those activities that were considered "not at all like MSPAP": 1) computations, estimations, and equations, and 2) traditional textbook-like word problems. The first category reflects those problems that solely ask students to do a computation or estimation, or to solve an equation. The problems in the second category reflect traditional word problems in which students need to provide or select a numerical answer based on their computations. Thus, the first two categories do not require the same level of problem solving and/or reasoning as defined by the MLO's and MSPAP. Although some of the skills required in problems of these types may also be required by MSPAP tasks, overall the problems themselves are not considered to be similar to MSPAP tasks.

The other four levels include activities that are similar to MSPAP tasks to some extent: MSPAP-like 1, MSPAP-like 2, MSPAP-like 3, and MSPAP-like 4. Activities at the MSPAP-like 1 level only require students to develop or complete a graph, table, pattern, or to physically measure an object. In these types of activities, students are not required to provide any interpretation or explanation of their work, and the activity does not require the same level of problem solving and/or reasoning as defined by the MLO's and required by the MSPAP tasks. MSPAP-like 2 activities require some problem solving and/or reasoning, but not to the same extent as required by MSPAP tasks. They also require students to show their work, provide an explanation, and/or interpret tables or graphs, and they can be completed in about five minutes.

MSPAP-like 3 and 4 activities require a similar level of problem solving and/or reasoning as required by the MSPAP tasks. MSPAP-like 3 tasks also require at least two short explanations or one long explanation (i.e., about a paragraph), and consist of approximately 3-5 items related to the same problem situation. Many of them also ask students to develop graphs, tables, or charts. MSPAP-like 3 tasks are considered to be similar to MSPAP tasks in terms of the processes being measured and the format, but not as extensive in length. MSPAP-like 4 tasks are considered to be similar to MSPAP tasks in terms of the processes being measured as well as the format in which they are measured and their length. These tasks require students to show their work and/or to develop graphs, tables, or charts; they require at least 3 short explanations and/or one or more long explanations; and they require students to respond to 6 or more items related to the same situation.

Each activity, regardless if it was one task or a set of distinct items, could be coded in more than one of the six MSPAP-like levels. For example, as indicated in Table 16, of the 1,940 instruction activities, 83% (1,617) were coded solely for one MSPAP-like level, 14% were coded for two MSPAP-like levels, and 2% were coded for three MSPAP-like levels. The other 1% were coded for four or more MSPAP-like levels. The test preparation activities were similar in this regard, about 89% were coded solely in one MSPAP-like level. A smaller percentage of assessment activities were coded in only one level (62%), and approximately 27% were coded in two levels, 9% in three levels, and 2% in four levels.

[Insert Table 16]

All Grades. Table 16 indicates the percentage of times an activity was coded for each MSPAP-like level when one, two, and three levels were coded per activity. The last column in the table, labeled 'overall', indicates the percentage of times an activity was coded for each level regardless if one, two, or three levels were coded per activity. For example, for those instruction activities in which only one level was selected, computation/equation was selected for 39% of them; for those activities in which two levels were selected, computation/equation was selected for 79% of the tasks; and for those activities in which three levels were selected, computation/equation was selected for 98% of the tasks. Overall, regardless of how many MSPAP-like levels were coded for an activity, computation/equation was selected for 46% of the instruction activities.

As indicated in the overall column in the table, the most common type of instruction and assessment activity required the student to perform computations or estimations, or to solve equations. This level was selected for 46% of the instruction activities and 66% of the assessment activities. In general, the MSPAP-like 2 was the next most commonly coded level for the instruction (34%) and assessment (32%) activities followed by traditional word problems (14% for instruction and 31% for assessment activities). The MSPAP-like 4 level was one of the least frequently coded categories for instruction (5%) and assessment (4%) activities.

As might be expected, the 3rd, 5th, and 8th grade MSPAP test preparation activities, as compared to the instruction and assessment activities, are more similar to MSPAP tasks as indicated by the 'overall' column. The most frequently coded levels for the test preparation activities were MSPAP-like 2 (38%) and MSPAP-like 4 (37%) task types. The next most frequently coded task type for test preparation activities was the MSPAP-like 3 task type (27%). The computation/equation level was selected for only 15% of the MSPAP test preparation activities.

Differences Across Grades. Table 17 provides the overall results for each grade level. The overall percentages in the table reflect the percentage of times each MSPAP-like level was coded for an activity regardless of the number of codes per activity. Differences across grades were rather small. For instruction, slightly more elementary activities were coded as MSPAP-like 1 and slightly more middle school activities were coded as MSPAP-like 3 activities. Also, for instruction and assessment activities, there was a slight increase in the percentage of MSPAP-like 3 and MSPAP-like 4 activities for the on-grade levels when compared to the off-grade levels. As an example, the percentage of MSPAP-like 3 and 4 instruction activities for the on-grades (3rd, 5th, and 8th) range from 14% to 20% depending on grade, whereas the percentages for the off-grades range from 8% to 13%. With regard to MSPAP test preparation, 58% of the 5th grade activities were coded as MSPAP-like 2, whereas only 33% of the 3rd and 8th grade activities were coded at this level. More activities in the 3rd and 8th grades (57% and 53%) were coded as either MSPAP-like 3 or MSPAP-like 4 compared to a smaller percentage of activities in the 5th grade (40%).

[Insert Table 17]

Summary

Performance-based assessments are being used by a number of states to promote instructional practices that foster critical thinking and reasoning skills. They are also being used for high-stakes purposes such as to hold schools accountable to state standards. Given the intentions of performance-based assessments and the stakes associated with them, it is imperative that the consequences of such assessments be examined (Linn, 1994; Koretz, Barron, Mitchell, & Stecher, 1996). This study is part of a larger, comprehensive research program designed to examine the consequences of the Maryland School Performance Assessment Program (MSPAP). The primary focus of the present study was to examine the consequences of MSPAP on mathematics instruction and assessment. In particular, it examined the differences among on-grade and off-grade mathematics teachers' composite scores on a number of dimensions reflected in the teacher questionnaire. Moreover, it examined the relationship among MSPAP mathematics school performance gains, mathematics teacher composite scores on the dimensions reflected in the teacher questionnaire, and the variable percent free or reduced lunch that served as a proxy for SES. Lastly, an analysis of classroom instruction and assessment materials was conducted to examine the extent to which the classroom materials reflect the Maryland Learning Outcomes and the goals of MSPAP. The intention of this latter analysis was to provide more direct evidence of the consequences of MSPAP on instruction as compared to self-report data obtained through the questionnaires.

The results of a multivariate analysis of variance and associated post hoc analyses indicated that elementary teachers, as compared to middle school teachers, were significantly more likely to report that (1) they place a greater emphasis on the mathematics learning outcomes and reform-oriented problems in their instruction, (2) their emphasis on the mathematics learning outcomes and reform-oriented problems has increased to a greater extent, (3) their mathematics instruction has been influenced by MSPAP to a greater extent, and (4) they have received greater professional development support with respect to MSPAP. There was not a significant difference among the grades with respect to teacher support for MSPAP. There were few significant differences between the on-grade and off-grade teachers. In general, when differences did exist they were between the elementary on-grade and off-grade teachers; however, the differences were relatively small. This implies that the consequences of MSPAP on these dimensions, as reported by teachers, are similar for both on-grades and off-grades. Additional analyses indicated that principals had higher composite mean scores than teachers with respect to (1) their support for MSPAP, (2) their belief that MSPAP has had an impact on classroom instruction, and (3) their belief that teachers have had adequate professional development activities related to MSPAP. There were no significant differences between elementary and middle school principals with respect to these dimensions. Further, students had relatively lower composite mean scores than teachers on the dimension regarding the extent to which their classrooms emphasized the mathematics learning outcomes and reform-oriented problems. The 5th grade students had a significantly higher composite mean score on this dimension than the 4th, 7th, and 8th grade students; whereas, the 7th grade students had the lowest composite mean score.

A latent variable growth model analysis (c.f., Meredith & Tisak, 1990; McArdle & Epstein, 1987; Muthen, 1991) examined MSPAP mathematics performance from 1993 to 1997 in relation to the teacher questionnaire dimensions and the variable, percent free or reduced lunch, which served as a proxy for SES. The following is a summary of the results from this analysis:

(1) Teachers in schools that had higher MSPAP mathematics scores in 1993 reported higher levels of Current Math Instruction (emphasis on learning outcomes and reform-oriented problem types) as compared to teachers in schools that had lower MSPAP scores in 1993.

(2) Teachers in schools that had lower MSPAP mathematics scores in 1993 reported higher levels of MSPAP Professional Development Support than teachers in schools that had higher MSPAP scores in 1993. This may imply that schools who initially performed poorly on MSPAP are providing teachers with more professional development support than schools who performed well on MSPAP.

(3) Schools that had lower MSPAP math scores in 1993 have higher rates of MSPAP math performance change as compared to schools with higher MSPAP math scores in 1993.

(4) Schools with lower mathematics scores on MSPAP in 1993 were schools with a higher percentage of free or reduced lunch (i.e., lower SES).

(5) There was no relationship, however, between percent free or reduced lunch and change in MSPAP mathematics score over time. This implies that the amount of free or reduced lunch that a school receives is not related to MSPAP mathematics performance gains.

(6) Higher levels of teacher reported MSPAP Influence on Instruction were associated with higher levels of rate of change in MSPAP mathematics performance over time. Thus, the schools for which teachers reported that MSPAP had a greater influence on their instruction had greater MSPAP mathematics performance gains.

It should be noted that although the latent growth model fit the data and the results suggest several positive consequences of MSPAP, the sample size used in the analysis was relatively modest (i.e., the number of schools used in the analysis was 82).

An important aspect of this study was the analysis of the mathematics classroom instruction and assessment materials. These data provided more direct evidence of the consequences of MSPAP on mathematics instruction. The results from this analysis indicated that approximately 50% of the mathematics instruction and assessment tasks consisted solely of computations, equations, or traditional word problems; whereas, the other 50% of the tasks reflected one or more characteristics of MSPAP tasks. However, there was only approximately 15% of the tasks that were very similar to MSPAP in terms of the level of problem solving and reasoning required, explanations required, and format of responses. It is important to note, however, that MSPAP tasks are set in a realistic context, are interdisciplinary, and have a number of extended items related to the same problem situation. Further, they require a high level of problem solving and reasoning and require students to provide explanations for their thinking. Thus, the finding that only approximately 15% of the classroom tasks reflected the majority of the characteristics of MSPAP tasks may not be that unreasonable. However, it would be important to conduct such an analyses in several more years to determine the extent to which classroom materials are changing over time.

References

Arbuckle, J.L. (1997). AMOS User's Guide Version 3.6. Chicago: SmallWaters Corporation.

Bentler, P.M. & Chou, C.-P. (1987). Practical issues in structural equation modeling. Sociological Methods and Research, 16, 78-117.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. (Eds.). Testing structural equation models. Newbury Park, California: Sage, 136-162.

Bryk, A.S., & Raudenbush, S.W. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. (Eds.). Testing structural equation models. Newbury Park, California: Sage, 136-162.

Chudowsky, N. & Behuniak, P. (1997). Establishing the consequential validity for large-scale performance assessments. Paper presented at the annual meeting of the National Council of Measurement, Chicago.

Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer (Ed.), Test validity (pp. 3-17). Hillsdale, NJ: Erlbaum.

Cronbach, L.J. (1989). Construct validation after thirty ears. In R.E. Linn (Ed.), Intelligence: Measurement, theory and public policy (pp. 147-171). Urbana: University of Illinois Press.

Frederiksen, J.R., & Collins, A. (1989). A districts approach to educational testing. Educational Researcher, 18(9), 27-42.

Joreskog, K.G., & Sorbom, D. (1994). LISREL 8 Users Reference Guide. Chicago: Scientific Software.

Koretz, D. M., Barron, S., Mitchell, K. J., & Stecher, B.M. (1996). Perceived effects of the Kentucky instruction results information district. MR-792-PCT/FF . Santa Monica, CA: RAND.

Koretz, D. M., Mitchell, K., Barron, S., & Keith, S. (1996). Final report: Perceived effects of the Maryland School Performance Assessment Program. (CFDA No. 84.117G). National Center for Research on Evaluation, Standards, and Student Testing, LA.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. Educational Evaluation and Policy Analysis, 15(1), 1-16.

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. Educational Researcher, 23(9), 4-14.

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 State Board of Education (1995). Maryland school performance report: State and school systems. Baltimore, MD.

McArdle, J.J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Development, 58, 110-133.

Meredith, W. & Tisak, J. (1990). Latent curve analysis, Psychometrika, 55, 107-122.

Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments (ETS RR-92-39). Princeton, NJ: Educational Testing Service.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (3rd ed.) (pp. 13-104). New York: American Council on Education.

Muthen, B.O. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds), Best methods for the analysis of change. Recent advances, unanswered questions, future directions (pp. 1-17). Washington, D.C.: American Psychological Association.

Muthen, B.O. & Curran, P.J. (1997). General growth modeling in experimental designs: A latent variable framework for analysis and power estimation. Psychological Methods, 2, 371-402.

National Council on Education Standards and Testing. (1992). Raising standards for American education. Washington, DC: Author.

Pomplum, M. (1997). State assessment and instructional change: A path model analysis. Applied Measurement in Education, 10(3), 217-234

Rogosa, D.R. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. Journal of Educational Statistics. 12. 185-195.

Rogosa, D.R. & Willet, J.B. (1985) Understanding correlates of change by modeling individual differences. Psychometrika, 50, 203-228

Singer, J.D. (1999). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth curve models. Journal of Educational and Behavioral Statistics, 23, 323-356.

Willet, J.B. & Sayer, A.G. (1994) Using covariance structure analysis to detect correlates and predictors of change. Psychological Bulletin. 116. 363-381.

Table 1

Student Questionnaire Return Rate

Grade	Number of Students	Number of Classes	Number of Students Per Form
4^{th}	1076	48	359
5^{th}	1442	67	481
7^{th}	845	37	282
8^{th}	1207	58	402

Table 2

Number of Teachers and Classroom Activities by Grade Level

Grade	Teacher		Activities	
	Number	Percentage	Number	Percentage
2	39	15%	591	15%
3	49	19%	854	22%
4	31	12%	454	11%
5	45	18%	698	18%
7	37	15%	639	16%
8	52	21%	712	18%
Total	253*	100%	3948	100%

* 3 teachers changed the grade taught from fall to spring

Table 3

Type of Classroom Activity

	Activities		Teacher		Mean Number of Activities Per Teacher
	Number	Percentage	Number	Percentage	
Instruction	1940	49%	245	98%	7.92
Assessment	1388	35%	214	86%	6.49
MSPAP Test Preparation (3, 5,8)	125	3%	51	35%	2.45
Scoring Schemes	332	8%	141	56%	2.35
Not Coded	163	4%	90	36%	1.81

Table 4

Confirmatory Factor Analysis Excluding Instructional/Assessment Change Measures – Teacher Questionnaire

	χ^2	df	P	RMSEA	NFI
On-grade (n=254)					
1-factor model	258.407	27	.000	.184	.686
4-factor model	63.053	21	.000	.089	.923
5-factor model	18.859	18	.401	.014	.977
Off-grade (n=172)					
1-factor model	150.488	27	.000	.164	.747
4-factor model	40.702	21	.006	.074	.931
5-factor model	18.362	18	.432	.011	.969
On and off grade					
5-factor model					
Constrained	73.634	63	.169	.020	.948
Unconstrained	37.227	36	.412	.009	.974

Table 5

Regression Coefficients and Significance Tests for Confirmatory Factor Model with Five Factors - On and Off grade levels (Parameters constrained) – Teacher Questionnaire

Dimension and Measure	Unstandardized Regression Coefficients	SE	t
MSPAP Familiarity			
General	1.000		
Results	1.546	.115	13.484*
Support MSPAP			
General	1.000		
Instruction	1.304	.174	7.488*
Current Math Instruction/ Assessment			
Learning outcomes	.835	.056	14.965*
Problem types	1.000		
MSPAP Impact	1.000		
Professional Dev. Support			
MSPAP	1.000		
Amount	.777	.094	8.305*

Note: *p < .01

Table 6

Confirmatory Factor Analysis Including Instructional/Assessment Change Measures – Teacher Questionnaire

	χ^2	df	p	RMSEA	NFI
On-grade (n=178)					
1-factor model	319.030	44	.000	.188	.605
4-factor model	183.384	38	.000	.147	.773
6-factor model	34.777	30	.251	.030	.957
Off-grade (n=112)					
1-factor model	232.182	44	.000	.196	.629
4-factor model	140.088	38	.000	.156	.776
6-factor model	41.682	30	.076	.059	.933
On and off grade					
6-factor model					
Constrained	135.607	96	.005	.038	.905
Unconstrained	76.500	60	.074	.031	.947

Table 7

Regression Coefficients and Significance Tests for Confirmatory Factor Model with Six Factors – On-grade and Off-grade Levels – Teacher Questionnaire

Dimension and Measure	Unstandardized Regression Coefficients		SE		t	
	On-grade	Off-grade	On-grade	Off-grade	On-grade	Off-grade
MSPAP Familiarity						
General	1.000	1.000				
Results	1.509	1.946	.204	.269	7.406*	7.228*
Support MSPAP						
General	1.000	1.000				
Instruction	1.200	1.303	.201	.277	5.958*	4.703*
Current Math Instruction/ Assessment						
Learning outcomes	.830	.628	.080	.080	10.412*	7.858*
Problem types	1.000	1.000				
Change Math Instruction/ Assessment						
Learning outcomes	1.088	.770	.097	.094	11.240*	8.225*
Problem types	1.000	1.000				
MSPAP Impact	1.000	1.000				
Professional Dev. Support						
MSPAP	1.000	1.000				
Amount	.679	.929	.144	.151	4.727*	6.175*

Note: *p<.01

Table 8

Descriptive Data for the Six Dimensions– Teacher Questionnaire

Dimension		Off-Elem ($2^{nd}/4^{th}$) (n=81)	On-Elem ($3^{rd}/5^{th}$) (n=120)	Off-Middle (7^{th}) (n=31)	On-Middle (8^{th}) (n=58)
MSPAP	mean	3.230	3.393	2.930	3.175
Familiarity	sd	.572	.566	.673	.562
Support	mean	2.639	2.549	2.544	2.508
MSPAP	sd	.603	.604	.550	.610
Current Math	mean	3.140	3.296	2.916	2.993
Instruction/ Assessment	sd	.493	.393	.360	.486
Change Math	mean	3.029	3.181	2.962	2.945
Instruction/ Assessment	sd	.401	.459	.309	.351
MSPAP	mean	2.964	3.255	2.628	2.818
Impact	sd	.605	.586	.509	.692
Professional	mean	2.866	3.080	2.427	2.704
Dev Support	sd	.621	.575	.616	.756

Table 9

Univariate ANOVA's for the Six Dimensions– Teacher Questionnaire

Dimension	df	F	p	r^2
MSPAP Familiarity	3	5.956	.001	.049
Support MSPAP	3	.623	.601	.004
Current Math Instruction/ Assessment	3	9.850	.000	.084
Change Math Instruction/ Assessment	3	5.730	.001	.047
MSPAP Impact	3	12.702	.000	.108
Professional Dev. Support	3	10.818	.000	.092

Table 10

Tukey HSD Post-Hoc Analyses – Teacher Questionnaire

Dimension	Contrast	Mean Difference	SE	p
MSPAP Familiarity	3/5 vs 7	.463	.117	.000
Current Math Instruction/	3/5 vs 7	.380	.089	.000
Assessment	3/5 vs 8	.303	.070	.000
Change Math Instruction/	3/5 vs 2/4	.151	.059	.049
Assessment	3/5 vs 7	.218	.082	.040
	3/5 vs 8	.236	.065	.002
	2/4 vs 7	.336	.128	.043
MSPAP Impact	3/5 vs 2/4	.291	.087	.005
	3/5 vs 7	.627	.122	.000
	3/5 vs 8	.437	.097	.000
	2/4 vs 7	.438	.133	.006
Professional Development Support	3/5 vs 7	.652	.127	.000
	3/5 vs 8	.376	.101	.001

Table 11

Descriptive Data for the Dimensions - Teacher, Principal, and Student Questionnaire

		Teacher				Principal		Class (Students)			
Dimension		Off-Elem (2nd/4th) (n=81)	On-Elem (3rd/5th) (n=120)	Off-Middle (7th) (n=31)	On-Middle (8th) (n=58)	Elem (n=55)	Middle (n=27)	Off-Elem (4th) (n= 41)	On-Elem (5th) (n=60)	Off- Middle (7th) (n=35)	Off- Middle (8th) (n=53)
MSPAP	mean	3.230	3.393	2.930	3.175	3.613	3.475				
Familiarity	sd	.572	.566	.673	.562	.400	.369				
Support	mean	2.639	2.549	2.544	2.508	3.109	3.147				
MSPAP	sd	.603	.604	.550	.610	.557	.423				
Current Math	mean	2.8569	3.147	2.652	2.714			2.672	2.759	2.490	2.603
Instruction/ Assessment	sd	.486	.472	.481	.554			.314	.271	.227	.302
MSPAP	mean	3.031	3.308	2.734	2.879	3.464	3.191				
Impact	sd	.655	.645	.577	.701	.535	.438				
Professional	mean	2.784	3.008	2.581	2.560	3.196	3.080				
Dev Support	sd	.869	.791	.776	.937	.546	.657				

Note: The teacher sample sizes for the dimension, Current Math Instruction and Assessment, are 41, 60, 35, and 53 for off-elementary, on-elementary, off-middle, and on-middle, respectively. This is because only teachers with corresponding class (students) data were considered for this analysis.

Table 12

Univariate ANOVA's for the Four Dimensions– Teacher vs. Principal

Dimension	df	F	p	r^2
MSPAP Familiarity	1	20.310	.000	.049
Support MSPAP	1	58.581	.000	.134
MSPAP Impact	1	12.783	.000	.031
Professional Dev. Support	1	11.848	.001	.028

Table 13

Tukey HSD Post-Hoc Analyses – Teacher and Student Questionnaire

	Teacher				Class (Students)			
Dimension	Contrast	Mean Diff	SE	p	Contrast	Mean Diff	SE	p
Current Math Instruction/ Assessment	3/5 vs 2/4	.239	.071	.004	4 vs 7	.177	.061	.019
	3/5 vs 7	.500	.099	.000	5 vs 7	.261	.057	.000
	3/5 vs 8	.422	.079	.000	5 vs 8	.159	.050	.008

Table 14

Results for the Level 1 Growth Model

Measure and variable	Estimates	SE	T
Regression Coefficients:			
Math93<- 1997 Performance	1		
Math94<- 1997 Performance	1		
Math95<- 1997 Performance	1		
Math96<- 1997 Performance	1		
Math97<- 1997 Performance	1		
Math93<- Rate of Change	4		
Math94<- Rate of Change	3		
Math95<- Rate of Change	1.39	.34	4.10
Math96<- Rate of Change	1		
Math97<- Rate of Change	0		
Latent Variable Means:			
1997 Performance	521.61	2.47	211.04
Rate of Change	-2.70	.26	-10.22
Variances/Covariances:			
1997 Perform-Rate of Change	-1.75	6.13	-0.28
1997 Performance	496.66	79.44	6.25
Rate of Change	2.43	1.03	2.34
e1	31.54	9.27	3.40
e2	32.70	6.96	4.70
e3	71.02	12.23	5.80
e4	23.25	5.75	4.05
e5	47.28	10.39	4.55

Table 15

Results for the Level 2 Growth Model - Factors Introduced to Explain MSPAP 1997 Performance and Rate of Change

Measure and Variable	Estimates	SE	t
Regression Coefficients			
Effects on 1997 Perform.			
Current Math Instruction	6.93	5.67	1.22
MSPAP Impact	.39	4.16	.09
Percent Free Lunch	-.78	.06	-13.13
Effects on Rate of Change			
Current Math Instruction	1.21	1.02	1.26
MSPAP Impact	-1.58	.75	-2.10
Percent Free Lunch	-.01	.01	-.72

Table 16

MSPAP-like Levels for Mathematics Classroom Activities - All Grades

	Number of Levels Selected			
	One	Two	Three	Overall
Instruction	n=1617 (83%)	n=278 (14%)	n=41 (2%)	
Not at all like MSPAP				
Computation/Equation	39%	79%	98%	46%
Traditional Word Problems	6%	49%	73%	14%
MSPAP-like Levels				
MSPAP-like 1	10%	19%	37%	12%
MSPAP-like 2	31%	45%	88%	34%
MSPAP-like 3	9%	7%	2%	9%
MSPAP-like 4	5%	<1%	2%	5%
Assessment	n=857(62%)	n=388(27%)	n=129(9%)	Overall
Not at all like MSPAP				
Computation/Equation	49%	91%	98%	66%
Traditional Word Problems	4%	67%	92%	31%
MSPAP-like Levels				
MSPAP-like 1	6%	13%	24%	12%
MSPAP-like 2	27%	26%	83%	33%
MSPAP-like 3	9%	3%	2%	6%
MSPAP-like 4	6%	0%	0%	4%
MSPAP Test Preparation	n=115 (92%)	n=7 (6%)	n=2 (2%)	Overall
Not at all like MSPAP				
Computation/Equation	9%	100%	50%	15%
Traditional Word Problems	0%	14%	50%	2%
MSPAP-like Levels				
MSPAP-like 1	3 %	0%	50%	4%
MSPAP-like 2	33%	86%	100%	38%
MSPAP-like 3	23%	0 %	50%	27%
MSPAP-like 4	32%	0 %	0%	37%

Table 17

MSPAP-like Levels for Mathematics Classroom Activities -- For Each Grade

	All Grades	Grade					
Instruction		2nd	3rd	4th	5th	7th	8th
Not at all like MSPAP							
Computation/Equation	46%	51%	44%	50%	41%	45%	46%
Traditional Word Problems	14%	6%	12%	17%	17%	15%	17%
MSPAP-like Levels							
MSPAP-like 1	12%	17%	12%	11%	15%	6%	9%
MSPAP-like 2	34%	28%	35%	38%	38%	36%	30%
MSPAP-like 3	9%	5%	8%	8%	8%	11%	13%
MSPAP-like 4	5%	3%	6%	1%	7%	2%	7%
Assessment							
Not at all like MSPAP							
Computation/Equation	66%	64%	54%	72%	62%	79%	67%
Traditional Word Problems	31%	34%	26%	38%	35%	34%	23%
MSPAP-like Levels							
MSPAP-like 1	12%	16%	10%	12%	12%	9%	11%
MSPAP-like 2	32%	31%	40%	31%	35%	34%	24%
MSPAP-like 3	6%	4%	8%	3%	6%	5%	11%
MSPAP-like 4	4%	1%	5%	1%	7%	2%	4%
MSPAP Test Preparation							
Not at all like MSPAP							
Computation/Equation	15%	--	10%	--	11%	--	21%
Traditional Word Problems	2%	--	3%	--	4%	--	2%
MSPAP-like Levels							
MSPAP-like 1	4%	--	5%	--	4%	--	3%
MSPAP-like 2	38%	--	33%	--	58%	--	33%
MSPAP-like 3	27%	--	26%	--	11%	--	24%
MSPAP-like 4	37%	--	31%	--	29%	--	29%

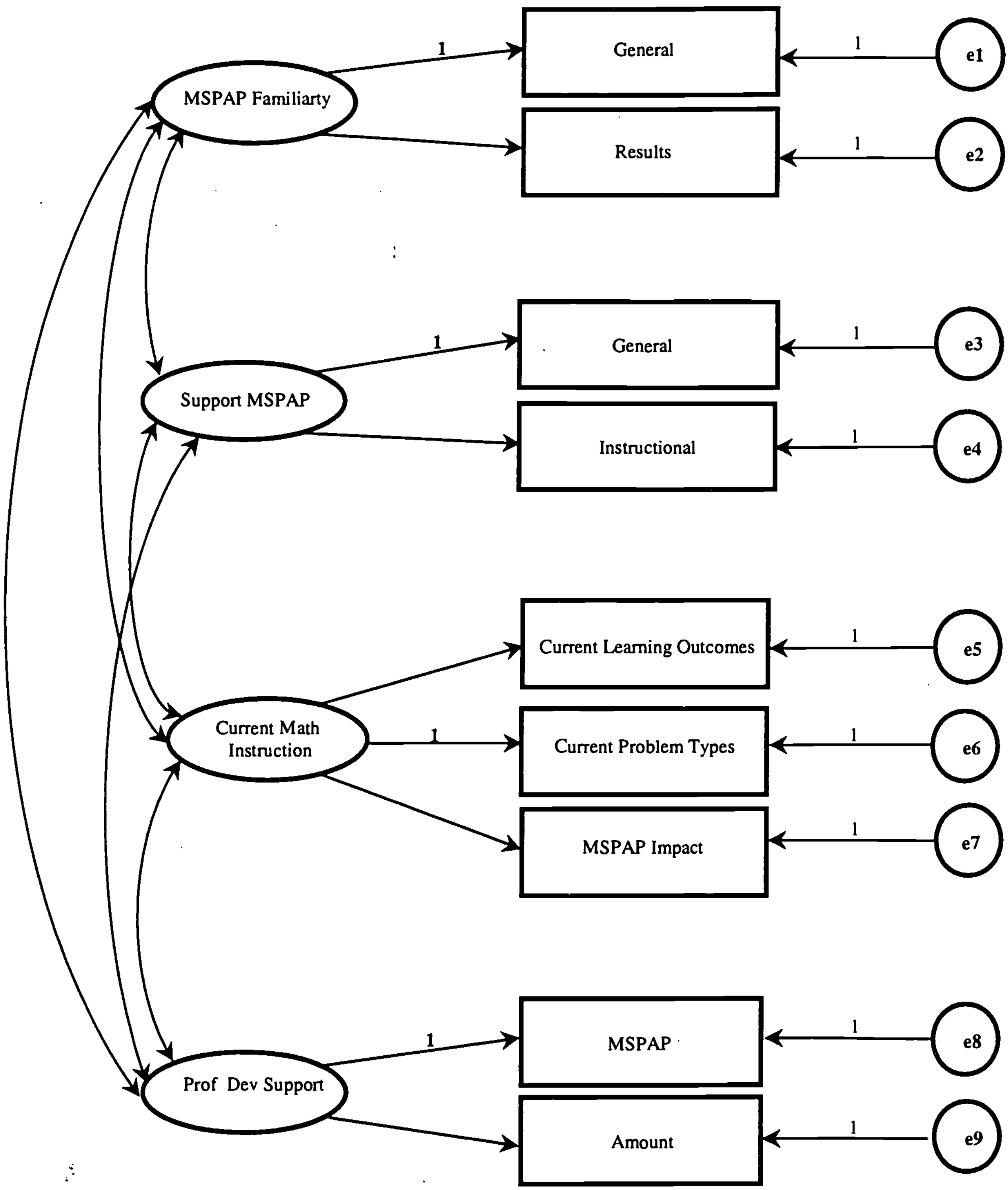
Figure 1. Hypothesized Dimensions, Measures, and Teacher Mathematics Questionnaire Items

Dimension/ Measure	Teacher Mathematics Questionnaire Item
Support MSPAP	
General	To what extent do you support or oppose MSPAP?
	To what extent has your support or opposition changed over the last few years?
	To what extent do you support or oppose the reporting of MSPAP results?
	To what extent do you support or oppose holding schools accountable for meeting the performance standards on MSPAP?
Instruction	MSPAP is a useful tool for helping me make positive changes in my instruction.
	MSPAP is a useful tool for making positive changes in instruction for those teachers who are resistant to change.
	Results of MSPAP provide useful information for making inferences about school improvement.
Current Math Instruction (1996-97)	
Learning Outcomes	How much emphasis have you placed on each of the following learning outcomes in your mathematics instruction this year?
	problem solving
	communication
	reasoning
	connections
Problem Type	How often have you used each of the following types of problems in your mathematics classroom this year?
	open-ended problems
	problems that take a few days or more to complete
	problems using manipulatives
	problems emphasizing relationships among mathematics concepts
	problems that integrate other subject areas in math
	problems that apply math to real-life situations
	How often do you ask your students to solve math tasks similar to MSPAP?
Change Math Inst. (1992-1997)	
Learning Outcomes	How has the emphasis on each of the following learning outcomes in your mathematics classroom changed from 1992-93 to 1996-97?
	problem solving
	communication
	reasoning
	connections
Problem Type	How has the emphasis on the use of the following types of problems in your mathematics classroom changed from 1992-93 to 1996-97?
	open-ended problems
	problems that take a few days or more to complete
	problems using manipulatives
	problems emphasizing relationships among mathematics concepts
	problems that integrate other subject areas in math
	problems that apply math to real-life situations

Figure 1. Hypothesized Factors, Measures, and Teacher Mathematics Questionnaire Items - Continued

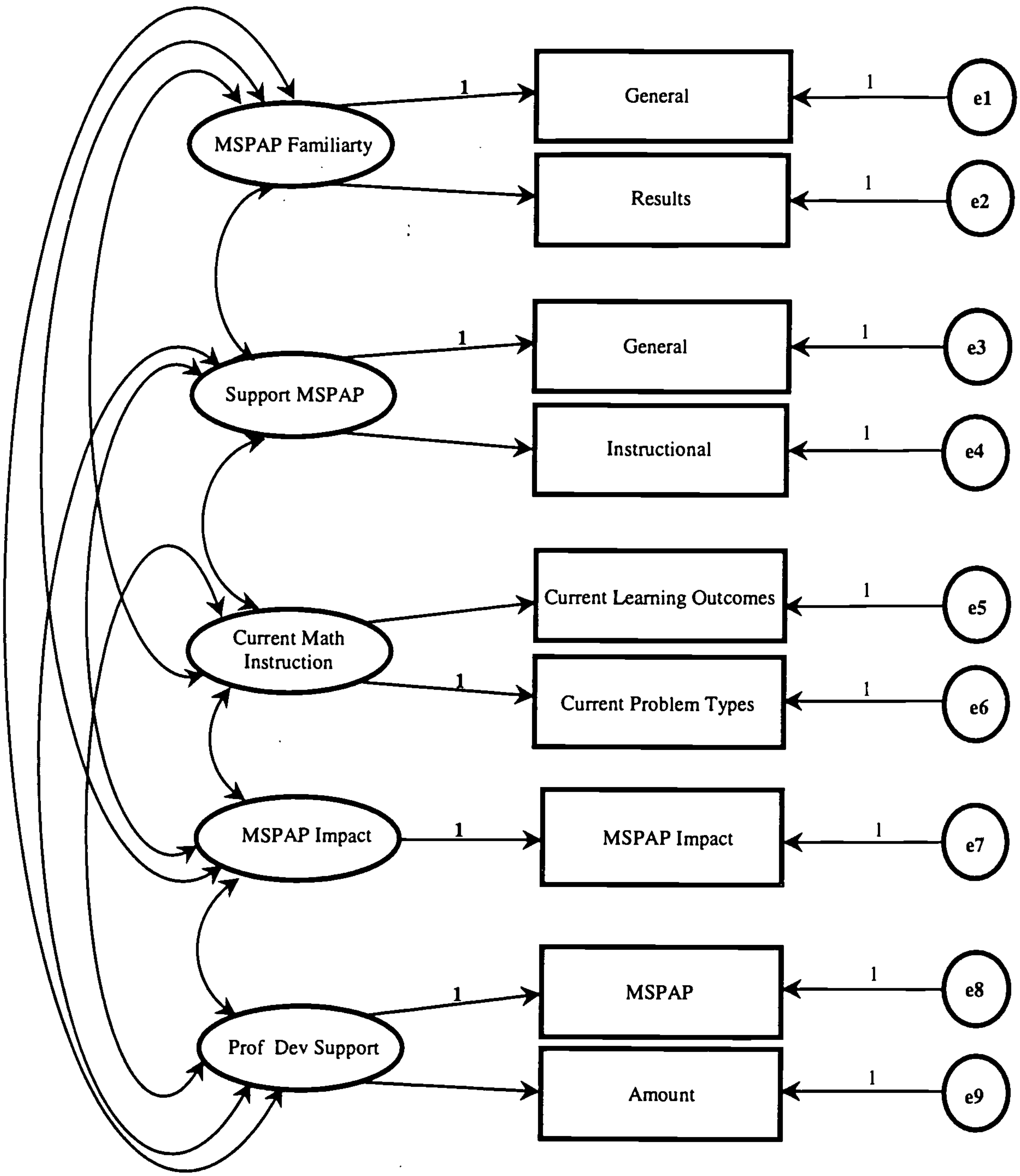
Dimension/ Measure	Teacher Mathematics Questionnaire Item
MSPAP Influence	
	To what extent has MSPAP influenced you to make positive changes in your mathematics instruction?
	To what extent have you focused on the following strategies in preparing your students for MSPAP?
	increasing the use of MSPAP-like tasks in instruction
	increasing the match between the content of instruction and the content of MSPAP
	improving instruction throughout the year
Professional Development Support	
Focus on MSPAP	To what extent did staff development activities address the following?
	Maryland Learning Outcomes
	Maryland Curriculum Framework
	Purpose of MSPAP
	Format of MSPAP tasks
	Content and skills assessed by MSPAP
	How to prepare students for MSPAP
	How to interpret and use MSPAP results to improve instruction
	How to explain MSPAP results to students/parents
Amount of Support	To what extent have you had the necessary support to enable you to make changes in your instruction to better reflect what is expected of students in MSPAP?
	To what extent have you had the necessary support to enable you to make changes in your assessments to better reflect what is expected of students in MSPAP?
	To what extent have you had the following necessary support/resources to enable you to make changes in your classroom activities to better reflect what is expected of students in MSPAP?
	inservices/workshops
	new instructional materials aligned to MSPAP

Four-Factor Model Excluding Instructional Change Measure



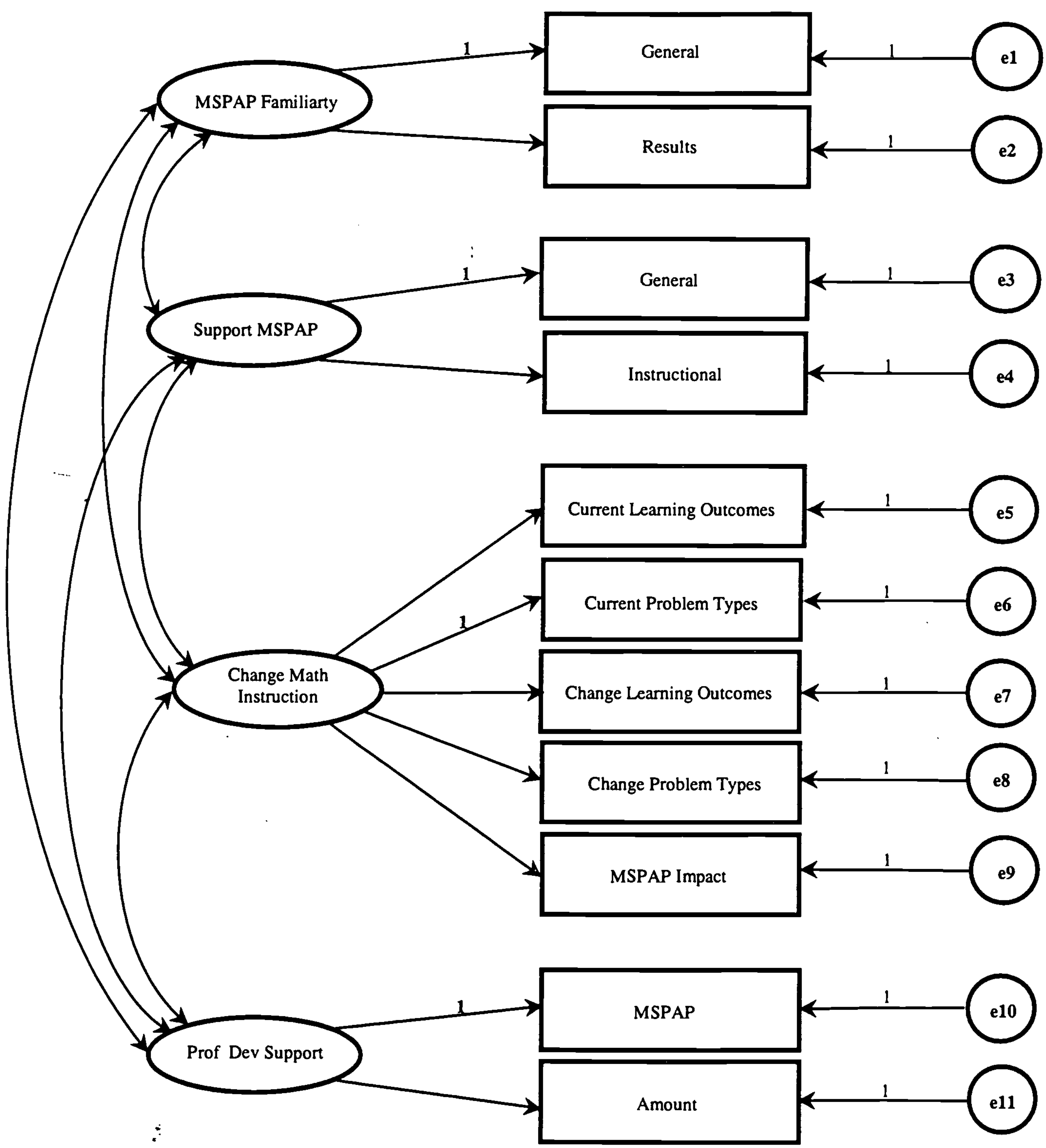
BEST COPY AVAILABLE

Five-Factor Model Excluding Instructional Change Measure



BEST COPY AVAILABLE

Four-Factor Model Including Instructional Change Measure



BEST COPY AVAILABLE

Six-Factor Model Including Instructional Change Measure

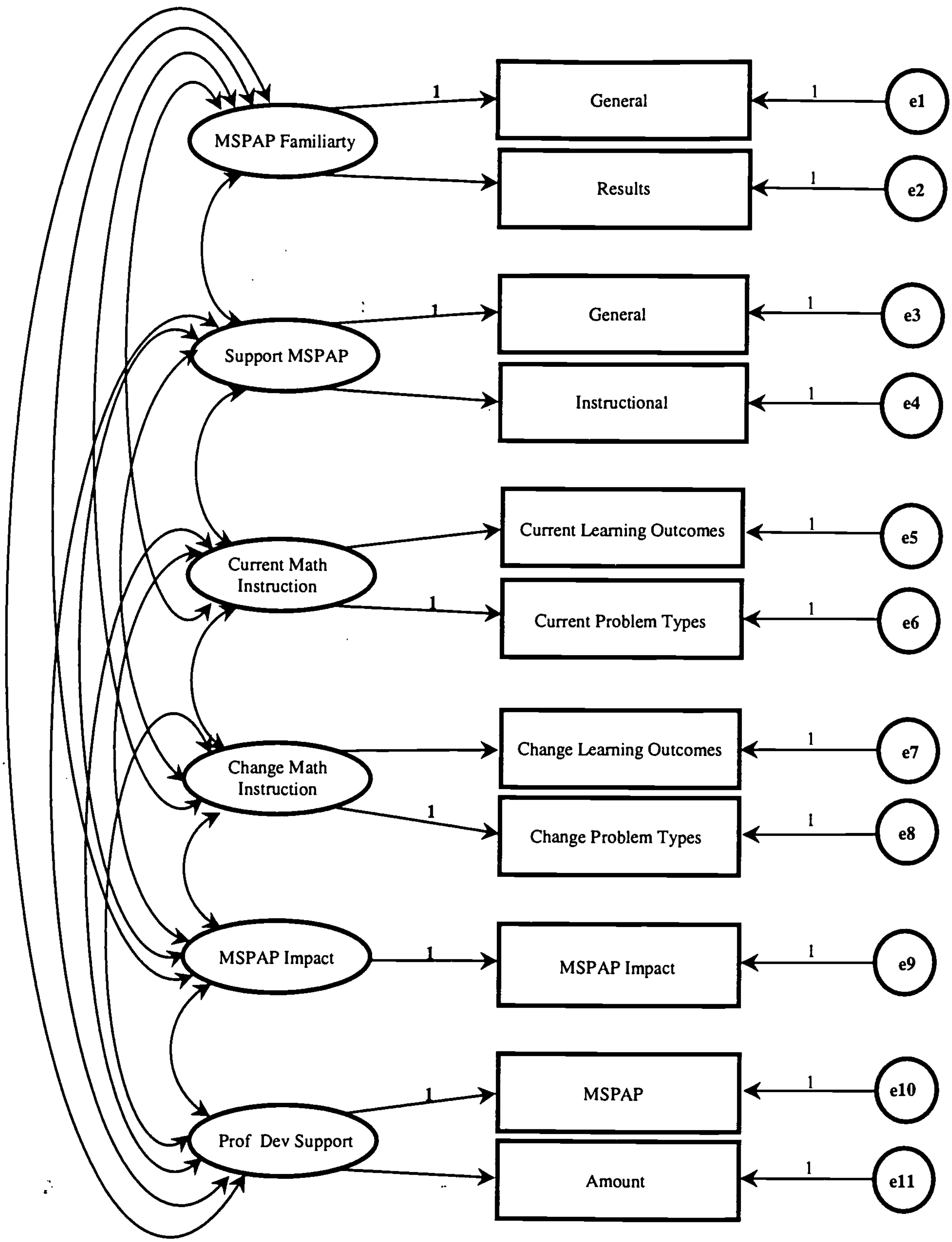
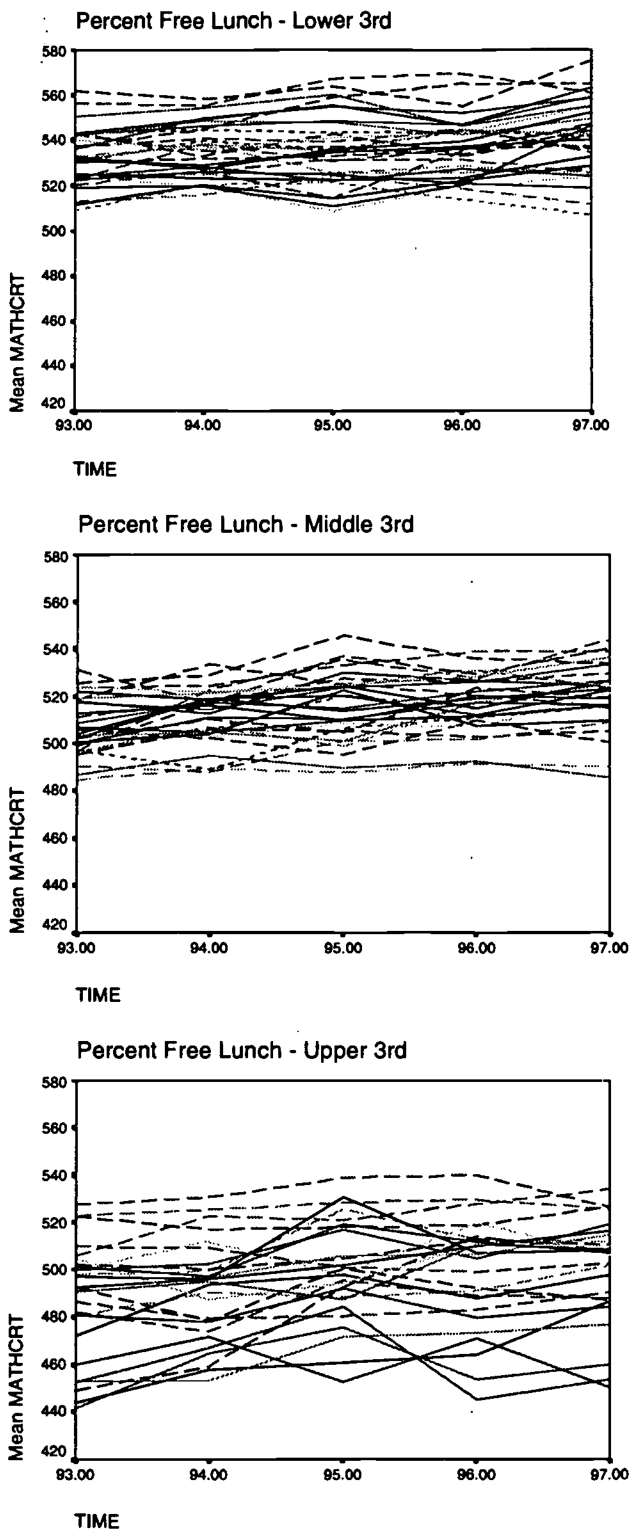
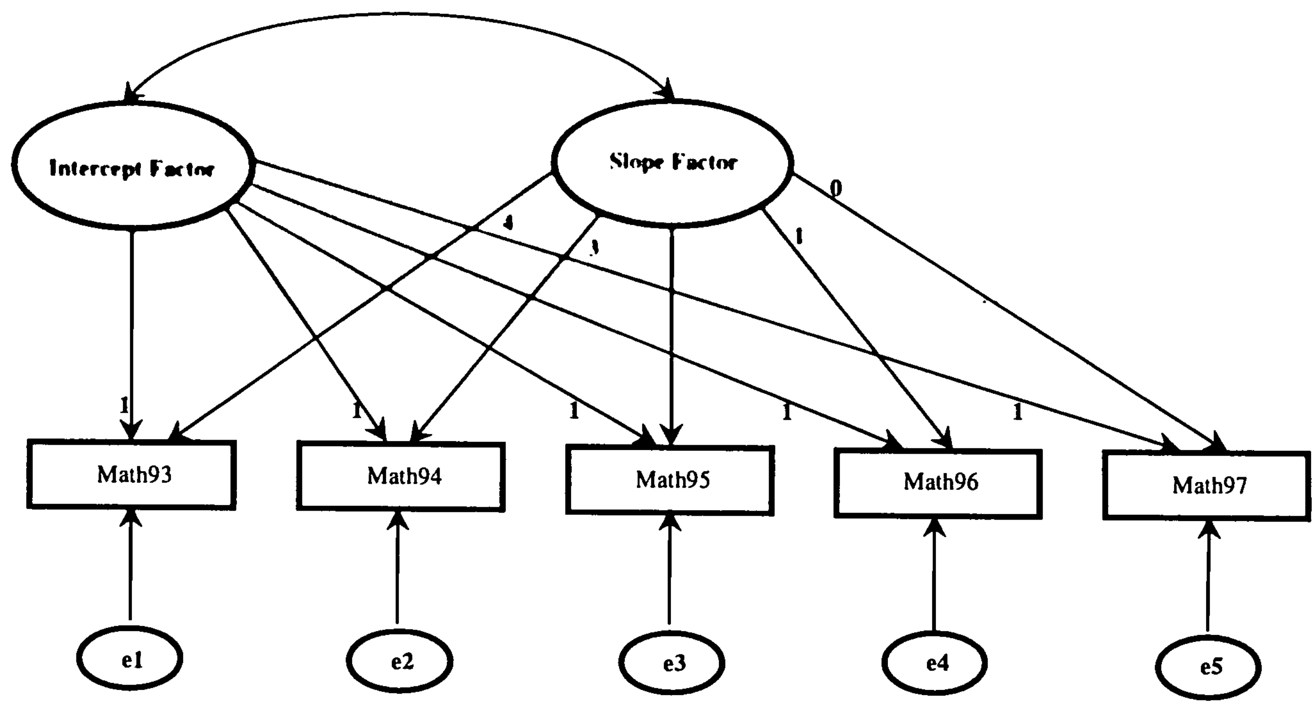


Figure 3. Change in Mean MSPAP Math Score Over Time by Percent Free Lunch Percentiles

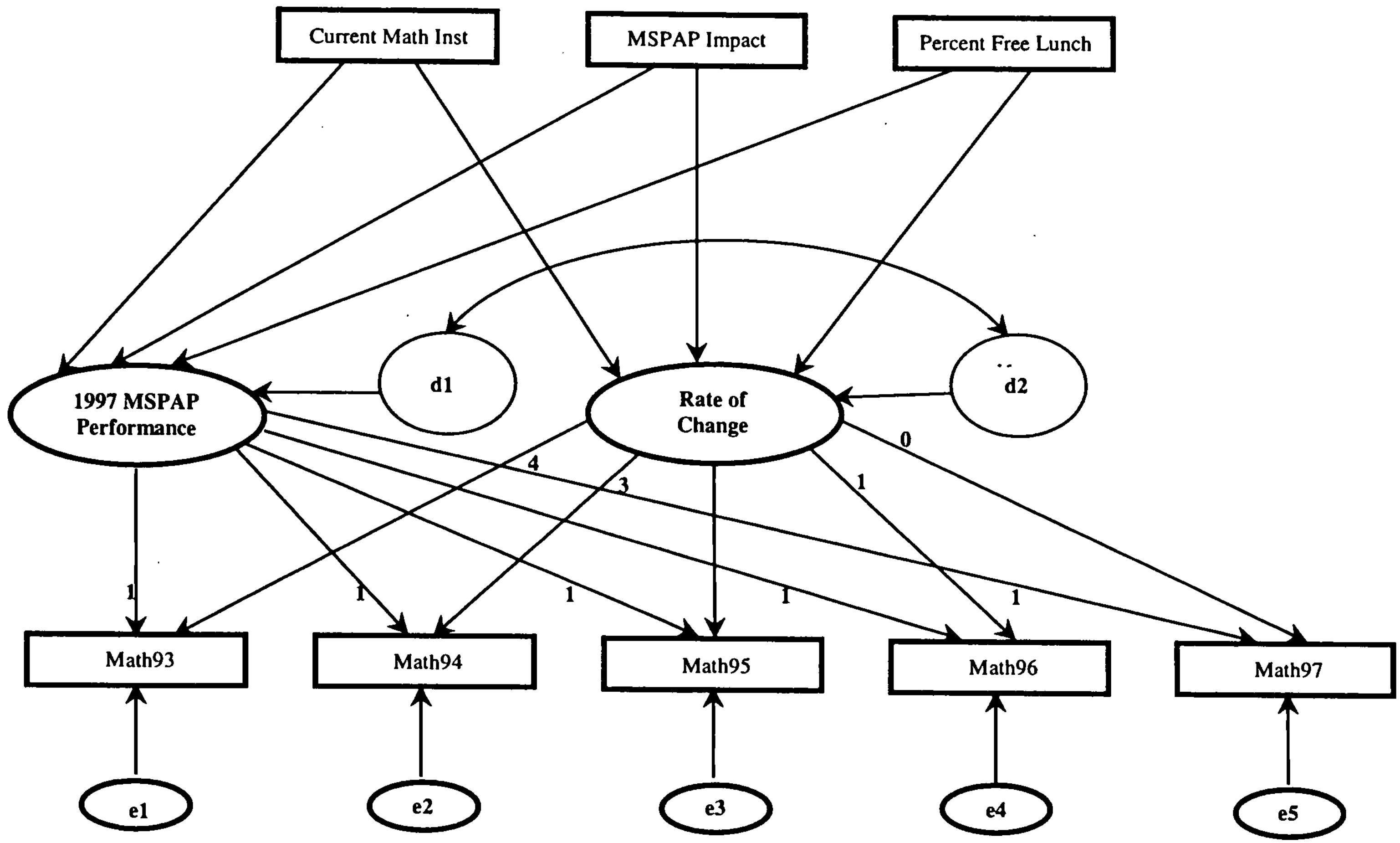


Level I Latent Variable Growth Model



BEST COPY AVAILABLE

Level II Latent Variable Growth Model



BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030179

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: 1) Impact of MSPAP: Evidence from P, T, & S Questionnaires . . . .
2) MSPAP Performance Gains and Their Relationship . . .
3) Impact of MSPAP: Evidence from Classroom Inst. & Asst Activities . . . .
4) Consequences of the MSPAP

Author(s): Lane et. al . . . .

Corporate Source: University of Pittsburgh

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1	2A	2B
Level 1	Level 2A	Level 2B
[✓]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]

Printed Name/Position/Title: Suzanne Lane Professor

Organization/Address: University of Pittsburgh
5C01 Forbes Quad
University of Pittsburgh, PA 15260

Telephone: 412-648-7095

FAX: 412-624-~~535~~ 7231

E-Mail Address: SL+@pitt.edu

Date: Sept 15, 1999

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com