DOCUMENT RESUME

ED 446 152 TM 031 924

AUTHOR Stone, Clement A.; Lane, Suzanne
TITLE The Relationship between Changes in MSPAP School Performance over Time and Teacher, Student, and School Factors.
SPONS AGENCY Department of Education, Washington, DC.
PUB DATE 2000-04-00
NOTE 28p.; Paper presented t the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2000). Supported by a grant from the U.S. Department of Education, Assessment Development and Evaluation Grants Program.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; *Institutional Characteristics; Middle Schools; Performance Based Assessment; *Scores; Structural Equation Models; *Teacher Attitudes; Teacher Surveys; *Teaching Methods; Test Results
IDENTIFIERS *Maryland School Performance Assessment Program

ABSTRACT

The relationship between changes in the scores from Maryland's performance assessment program, the Maryland School Performance Assessment Program (MSPAP) from 1993 to 1998 and classroom instruction and assessment practices, student learning and motivation, students' and teachers' beliefs about and attitudes towards the assessment, and school characteristics was studied. The final study sample consisted of 90 elementary and middle schools. Using growth models estimated within a structural equation modeling (SEM) framework, several factors from each of these dimensions were observed to explain a significant amount of the variability in school performance. These factors as well as the design of evaluations that hope to study the impact of assessment programs on students, teachers, and schools were discussed. Instruction-related variables were found to explain differences in MSPAP performance levels across the subject areas, and for some subject areas, to explain differences in rates of change in MSPAP performance over time. In addition, the perceived impact of MSPAP on instruction/assessment practices was also found to significantly explain either differences in MSPAP performance levels or rates of change over time across the subject areas. Findings suggest that the design of a validity or impact study could be improved by measuring the outcomes chosen for this study concurrently with assessment performance over time. (Contains 3 figures, 3 tables, and 27 references.) (SLD)

The Relationship Between Changes in MSPAP School Performance over Time and Teacher, Student, and School Factors

Clement A. Stone

Suzanne Lane

University of Pittsburgh

Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, April 2000. Preparation of this paper was supported by a grant from the U.S. Department of Education, Assessment Development and Evaluation Grants Program (CDFA 84.279-A), for the Maryland Assessment Project. Our deepest appreciation is extended to the teachers, principals and students in Maryland for their invaluable time and effort spent on this project.

Abstract

A number of states are implementing statewide assessment programs that are being used for high-stakes purposes such as holding schools accountable to state standards. These assessments often depend on performance-based tasks under the prevailing assumption that they not only serve as motivators in improving student achievement and learning, they encourage instructional strategies and techniques in the classroom that are more consistent with reform-oriented educational outcomes (e.g., instruction focusing on reasoning and communication skills). Given these high expectations, more comprehensive and direct evidence for the consequences of the assessments (both negative and positive) need to be addressed. The purpose of this paper is to explore the relationship between changes in the scores from Maryland's performance assessment program (MSPAP) from 1993 to 1998 and classroom instruction and assessment practices, student learning and motivation, students' and teachers' beliefs about and attitude towards the assessment, and finally, school characteristics. Using growth models estimated within a structural equation modeling (SEM) framework, several factors from each of these dimensions were observed to explain a significant amount of the variability in school performance. The paper discusses these factors as well as the design of evaluations that hope to study the impact of assessment programs on students, teachers, and schools.

The Relationship Between Changes in MSPAP School Performance over Time and Teacher, Student, and School Factors

Recently Linn (2000) reviewed the role that assessment and accountability has played in the various periods of educational reform during the past 50 years. One of these roles is as a motivational mechanism for change, and more recently, assessment programs have been used to hold schools accountable to state learning outcome standards through the use of rewards and sanctions. He cites several reasons for the "great appeal of assessment to policymakers as an agent of reform": relatively low cost compared to classroom changes (e.g., reducing class size, increasing instructional time, implementing curriculum changes), ease and speed with which assessment requirements can be mandated, and finally, the results of the assessment program are directly available. He further discusses aspects necessary to ensure the information provided by assessment and accountability programs is valid and the importance of understanding the impact of such programs on educational practices and student outcomes.

The nature of the assessment programs has also changed over the years to mirror the philosophy of the particular educational reform movement. In the most recent educational reform movement, a number of states are now implementing statewide assessment programs that involve performance-based tasks in response to arguments that assessments utilizing more traditional types of standardized tests have led to educational practices that over-emphasize basic skills (e.g., Resnick & Resnick, 1992). The prevailing assumption underlying the use of performance-based assessments is that they encourage the use of instructional strategies and techniques that foster reasoning, problem solving, and communication (National Council on Education Standards and Testing, 1992). One state implementing such an assessment program is Maryland. The Maryland State Performance Assessment Program (MSPAP) is a performance assessment program designed to measure school performance for grades 3, 5, and 8 and provide information for school accountability and improvement (Maryland State Board of Education, 1995). Implemented in the early 1990's, 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.

Given the 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, and 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)." Further, the consideration of potentially negative effects through the eyes of multiple stakepersons may help ensure a more comprehensive evaluation of the consequences (Cronbach, 1989).

The purpose of this paper is to examine the broader impact of the MSPAP assessment program and explore the relationship between changes in MSPAP test scores for schools and classroom instruction and assessment practices, student learning and motivation, professional development, students' and teachers' beliefs about and attitude towards MSPAP, and finally, school characteristics. The subject areas of MSPAP that are considered in this study include: Mathematics, Reading, Writing, Science, and Social Studies.

METHODOLOGY

School Sample

This study examined the effects of the assessment program in schools that reflected student populations with different SES backgrounds and in schools that differed in the amount of change that occurred in MSPAP performance. To accomplish this, a stratified random sampling procedure was used to select schools for the study, 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, and elementary and middle schools from each of the nine cells were randomly sampled. Note that a larger number of elementary schools were selected because, compared to the middle schools, they have fewer teachers per grade. The study collected information related to Mathematics and Reading/Writing during the 1996-97 instructional year, and collected information related to Science and Social Studies during the 1998-99 instructional year. Also, a number of additional schools were randomly selected to a pool of alternate schools that were identified as potential replacements for schools who chose not to participate. Finally, 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 as compared to the sample size for the 1998-99 instructional year.

School Sample for Mathematics and Language Arts. 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. 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 approximately equal numbers 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.

School Sample for Science and Social Studies. Prior to selecting schools for the Science and Social Studies area, those schools that participated in the data collection in the 1996-97 year were excluded. A total of 126 elementary and 63 middle schools were selected to participate in the study with alternate schools identified as potential replacements for schools who chose not to participate. The final sample consisted of 103 elementary and 58 middle schools, with a total of 161 schools. Thus, the school participation rate was 82% for elementary schools and 92% for middle schools. There were approximately equal numbers of schools within each of the nine classification cells.

Of the 103 elementary schools, 87 were from the initial 126 that were sampled, and of the 58 middle schools, 44 were from the initial 63 that were sampled. The remaining schools were from the list of alternate schools for each cell. This represents schools from 22 systems/counties in Maryland. In summary, across the two years, a total of 251 schools participated in the study.

Instrumentation and Data Collection

Means of equated MSPAP scaled scores for schools in the sample from 1993 to 1997 or 1998 were provided by personnel within the Maryland State Department of Education. In the present study, changes in MSPAP test scores for schools were examined in relation to classroom instruction and assessment practices, student learning and motivation, beliefs about the impact of and attitude towards MSPAP, and finally, the school characteristic, percent free or reduced lunch which served as a proxy for socioeconomic status (SES). Percent free or reduced lunch data for the schools were also provided by the Maryland State Department of Education.

To triangulate on the impact of MSPAP, multiple data sources and measures were used to examine the changes in MSPAP performance. The data relevant to the present study was obtained from questionnaires that were developed for principals, teachers and students. The questionnaire for principals was the same for both elementary and middle school principals. Questionnaires specific to the different subject areas were developed for elementary (3^{rd} and 5^{th} grade) and middle school (8^{th} grade) teachers and students. Teachers and principals completed questionnaires in prior to the administration of MSPAP, whereas students completed questionnaires within the two weeks following the administration of MSPAP. The questionnaire return rates ranged from 68-87% across the subject areas for teachers, 64-78% across the subject areas for students, and were greater than 90% across the subject areas for principals.

The questionnaires consisted of both Likert (generally 4 point scales) and constructed response items. To triangulate on the impact evidence, students, teachers, and principals responded to similar questions for areas in which it was deemed appropriate. Sets of items on the teacher questionnaire were combined and validated through factor analytic methods to reflect the following dimensions:

- Familiarity with MSPAP - purpose, format, and subject of MSPAP; how to interpret, use, and explain MSPAP results
- Support for MSPAP - extent of and change over time; holding schools accountable; use for instructional purposes; and beliefs about MSPAP
- Current Classroom Instruction and Assessment - degree to which instruction and assessment reflected each of the state-defined learning outcome standards; and extent to which instruction and assessment reflected reform-oriented problem types
- Change in Instruction and Classroom Assessment – extent to which changes in the classroom occurred from the 1992 school year to the 1997 or 1998 school year
- MSPAP's Impact on Classroom Instruction and Assessment - extent to which MSPAP influenced changes in the classroom
- Nature of MSPAP-related professional development activities, and support for making changes in the classroom

It should be noted that 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 teachers.

The student questionnaires paralleled the teacher questionnaire where appropriate. Thus, students were also asked about the nature of the instruction and classroom assessment activities. In addition, they were asked about the extent of MSPAP preparation activities (e.g., practice exercises) and the extent to which students worked on tasks like those on MSPAP during the year. But, students were also asked questions related to the interest level of MSPAP tasks and how motivated they were to do well on MSPAP.

For other studies, teachers were also asked to provide a sample of instruction and 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. This data could not be incorporated into the present study because a much smaller sample of schools was targeted for this data source. Incorporating this data into the present study would have reduced the size of the samples prohibitively. Readers interested in the analyses of the classroom artifacts are referred to Cerrillo, Hansen, Parke, Lane, & Scott (2000).

RESULTS

Table 1 summarizes the mean MSPAP performance across the set of schools in the sample. As can be seen, the general trend indicates one of increasing mean performance over time, although the change from 1993 to 1998 is modest given the metric of the score scale. Except for the case of Writing, there appears to be larger gains in the early years, followed by a leveling of the scores during the period from 1995 to 1996, at which point there is again an increase in mean performance for schools. Additionally, a decrease in mean performance was noted for Science and Social Studies from the 1995 to 1996 administrations of MSPAP. It might be reasonable to ask whether some of these results are due to sampling error. However, an analysis of the mean MSPAP performance across all schools (n=962) revealed similar patterns to those in Table 1, with the same decrease in performance being found for Science and Social Studies from 1995 to 1996. Note that the sample sizes for the present study reflect a listwise deletion of missing questionnaire data. The samples for Science and Social Studies were reduced proportionately more than for Math, Reading, and Writing since both teacher and student questionnaire data were included in the analyses for Science and Social Studies.

Table 1: Means and Standard Deviations of MSPAP Scale Scores 1993-1998*

	Math		Writing		Reading		Science		Social Studies	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
93 Adm	510.6	24.1	504.6	19.3	503.8	20.6	509.6	25.3	503.3	18.5
94 Adm	513.9	23.0	503.8	18.8	505.6	19.5	514.6	23.0	508.4	18.7
95 Adm	518.1	22.8	506.1	20.1	512.4	17.1	519.0	21.4	512.8	17.0
96 Adm	518.7	22.9	511.3	20.0	513.0	17.6	518.3	23.8	511.5	19.0
97 Adm	521.8	24.1	513.9	22.5	517.1	17.5	518.9	24.9	514.4	20.0
98 Adm							523.6	22.9	518.7	19.4

* Note: The sample sizes for the subject areas were as follows: Math (86), Writing (86), Reading (86), Science (116), and Social Studies (111)

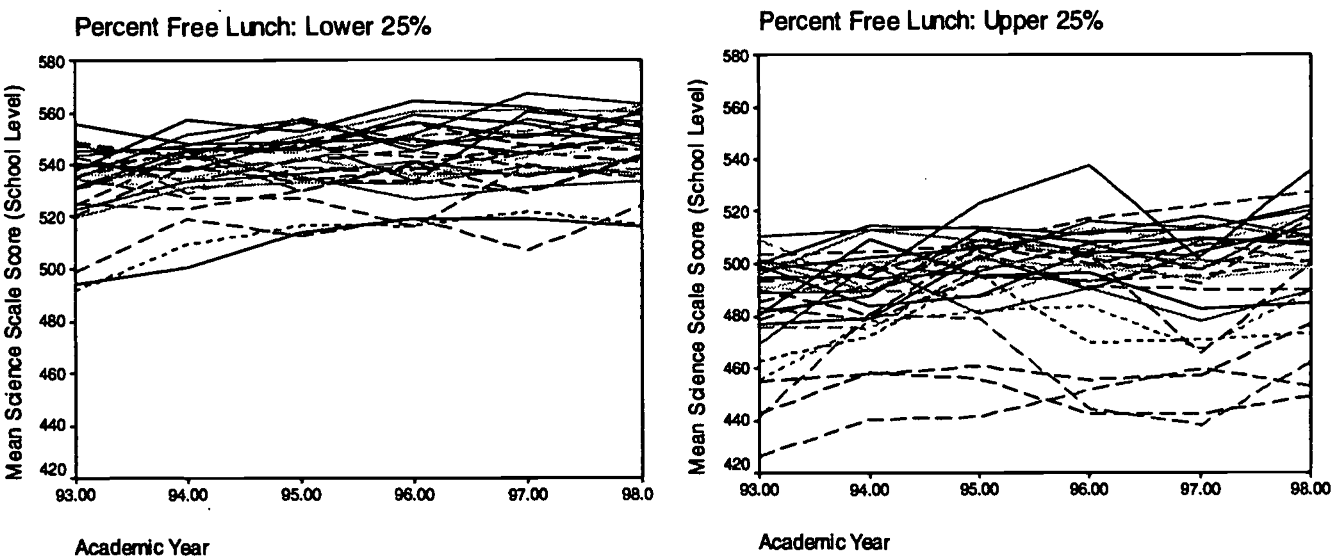
Modeling Differences in School Performance Over Time

Of more interest than mean performances are the differences among the schools in terms of their initial performance (1993) and their rates of change over time and whether these individual differences can be explained by factors relevant to the MSPAP assessment program and factors relevant to characteristics of the schools. Random coefficient or growth models were used to examine MSPAP performance from 1993 to 1998 in relation to variables derived from the teacher and student questionnaires, and the school characteristic, percent free or reduced lunch which served as a proxy for socioeconomic status. 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 change). Further, these methodologies allow for studying individual differences in the patterns and identifying factors that affect the patterns of change. This type of analysis can not be modeled by time-specific comparisons involving group-level (e.g., means) differences.

Variables from questionnaires administered to teachers and students from the schools in the sample were hypothesized to explain individual differences in school performance over time. Due to the relatively small number of schools in the sample, the present study focused on data collected from teachers and students. The data for principals was excluded as it reflected more indirect knowledge of classroom practices, and as it turned out was not very variable. In addition, a subset of variables from the questionnaires was used that were considered to be more relevant than other dimensions for examining the relationship between change and teachers' and student's perceptions. From the teacher questionnaire, two dimensions were examined: MSPAP Impact and Current Classroom Instruction and Assessment. From the student questionnaire, the Current Instruction dimension and two Likert-scaled items were analyzed: 1) In class this year, how often did you work on tasks like those on MSPAP? And, 2) How important is it for you to do well on MSPAP? Readers interested in more detailed analyses of the questionnaires in relation to differences between principals, teachers, and students for all the dimensions as well as differences across grades and subject areas are referred to Lane, Stone, Parke, Hansen, & Cerrillo (2000).

Figure 1 illustrates the typical differences that were observed between schools in regards to their initial mean MSPAP performance and changes in mean MSPAP performance over time in the various subject areas. The figure presents the mean MSPAP Science performance from 1993 to 1998 for the sample of schools. Since percent free or reduced lunch was found to correlate significantly with MSPAP performance, the plots are presented for two subgroups of this variable (i.e., lower and upper quartiles) 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 Science performance and their change over time. Schools in the lower quartile (Higher SES) were concentrated in the range of 520-550 in 1993 whereas schools in the upper quartile (Lower SES) were concentrated in the range of 480-500 in 1993. In addition, the rate of change for schools in the lower quartile exhibited a more consistent increase over time whereas considerably more variability was observed for schools in the upper quartile. In both cases, the rate of change appears modest from 1993 to 1998.

Figure 1: Change in Mean MSPAP Science Scores Over Time by Percent Free Lunch Percentiles



The models used to capitalize on the information contained in multiwave data appear in the literature under a variety of labels, including random-effects models or random coefficient models (e.g., Laird & Ware, 1982) and hierarchical linear models (Bryk & Raudenbush, 1992). 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 analyze the repeated measurements of test scores, analyze the relationship between time (year of adminsitration) and test score levels, and estimate a reference status (intercept) and rate of change (slope) for each school. It would be expected that schools would differ with regard to their initial levels MSPAP performance (measured at time 1), their rates of change over time, and the shape or pattern of change (e.g., linear, nonlinear).

A linear growth model with a single outcome variable y measured for each school at each timepoint is:

$$y_{it} = \alpha_i + \beta_i x_{it} + \varepsilon_{it}, \qquad (1)$$

where α_i is an intercept parameter for each ith school, x_{it} is the time-related variable for the i^{th} school at time t, β_i is a slope parameter reflecting the linear rate of change over time for the i^{th} school, and ε_{it} is a residual reflecting both random measurement error and unspecified time-specific effects.

The parameters from the model at Level 1 (intercepts and slopes) are then modeled in relation to factors that are introduced to explain variation in the parameters across schools (Level 2). For example, the school-specific parameters, α_i and β_i from the Level 1 model, are incorporated into the Level 2 model with one school-specific explanatory variable (z_i) as follows:

$$\begin{aligned} \alpha_i &= \mu_\alpha + \gamma_\alpha z_i + \varepsilon_{\alpha i} \\ \beta_i &= \mu_\beta + \gamma_\beta z_i + \varepsilon_{\beta i} \end{aligned} \qquad (2)$$

where μ_α and μ_β are parameters reflecting group-level means of the intercepts and slopes, respectively, and the variance of these factors reflects the individual differences or random effects that exist around these group level parameters (e.g., larger variances reflect increased variability or less similar patterns in intercepts and slopes); z_i is a time-invariant covariate introduced to explain variation in these parameters (e.g., SES level); γ_α and γ_β are regression parameters reflecting the effects of the covariate on the Level 1 intercept and slope parameters; and, $\varepsilon_{\alpha i}$ and $\varepsilon_{\beta i}$ are residual terms. It is assumed that the ε_{it} are uncorrelated with $\varepsilon_{\alpha i}$ and $\varepsilon_{\beta i}$, but $\varepsilon_{\alpha i}$ and $\varepsilon_{\beta i}$ may be correlated. It should be noted that it is straightforward to increase the number of explanatory variables in the Level 2 model and consider time-varying covariates as well as non-linear growth rates in the Level 1 model. In the present study, various dimensions from the teacher and student questionnaires, and the variable percent free or reduced lunch were introduced to explain variation in the intercepts and slopes.

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).

Modeling the Changes in MSPAP Performance Over Time - Level I Growth Analyses

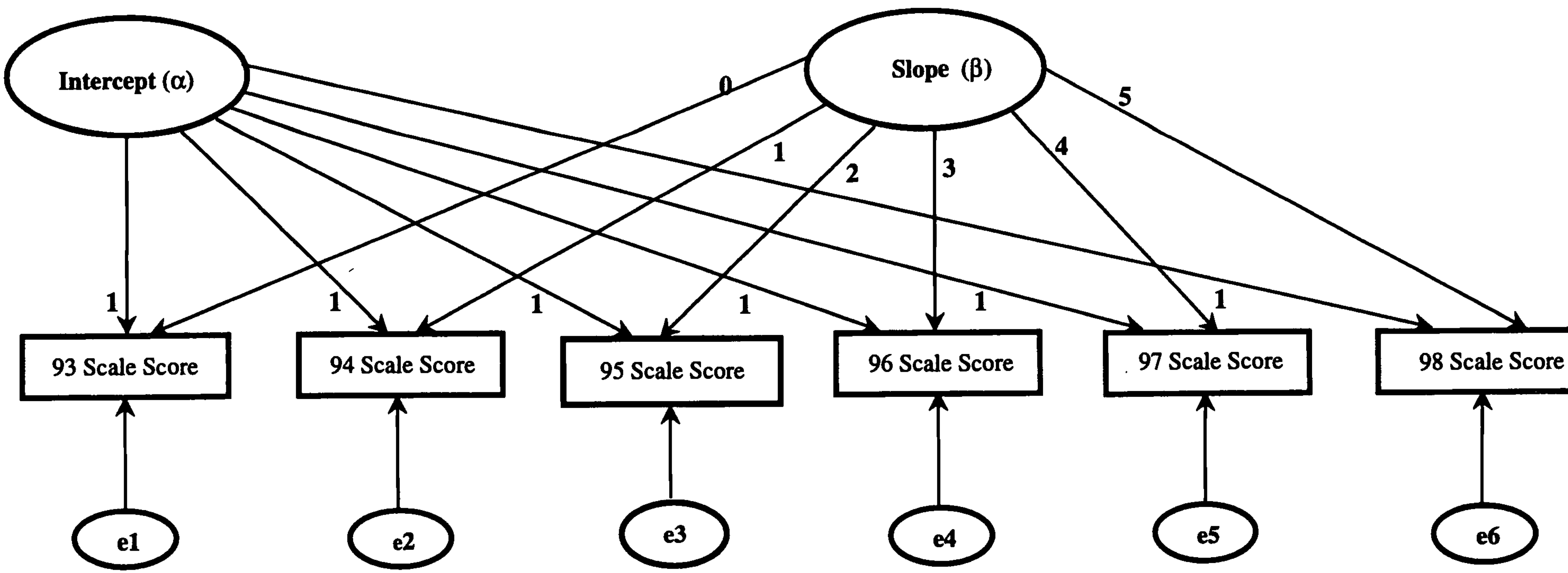
Figure 2 presents a Level 1 (Unconditional) growth model for the present study. This model involves the outcome variable, MSPAP Scale Score, measured at six timepoints. In order to translate the growth model into the framework of structural equation modeling, the school-specific random coefficients (intercepts and slopes from Level 1) are each modeled using two latent factors: 1) a factor representing a reference status of MSPAP performance (intercept or α), and 2) a factor which corresponds to the rate of change in MSPAP performance over time (slope or β). 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.

As can be seen from the figure, the Level 1 model has the format of a measurement or confirmatory factor analysis model in SEM with restrictive loadings: $Y = \Lambda\eta + \varepsilon$

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \\ Y_{i5} \\ Y_{i6} \end{pmatrix} = \begin{pmatrix} 1 & x_t \\ 1 & x_t \\ 1 & x_t \\ 1 & x_t \\ 1 & x_t \\ 1 & x_t \end{pmatrix} * \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \quad (3)$$

where Y is a vector of 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 may be specified and indicated through a curved bi-directional arrow in the figure. Note that, in order to specify these models in SEM, it is necessary to assume that $x_{it} = x_t$, which means that all individuals are measured at the same point in time at each time-point. In this study as well as other state-wide testing situations, tests are typically administered at the same time.

Figure 2. Level 1 Unconditional Growth Model



The regression coefficients relating the intercept factor to the measurements are fixed at 1 since the intercepts reflect a constant contribution to the measurements over time. The scaling of the slope factor is determined by the pattern in the x_t coefficients that relate the time variable to the observed measurements. To reflect a simple linear growth pattern with one unit of change between time points the coefficients (x_t) would be set to 0, 1, 2, 3, 4, and 5. Note that in the framework of SEM, it is possible to freely estimate coefficients or constrain parameters to any other specified pattern. Thus, there is no constraint that time points be equally spaced or that all x_t be specified.

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. Under the scaling in Figure 2, 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 this study, a decreasing pattern such as 5, 4, 3, 2, 1, and 0 was adopted. Since time 0 is associated with 1998 MSPAP performance, the intercept factor is interpreted as 1998 MSPAP status and a decrease in performance would be expected from 1998 to 1993. Note that analyses for Math, Reading, and Writing occurred over the period from 1993 to 1997. Thus, the intercept factor for these subject areas is interpreted as 1997 MSPAP status. The decreasing scaling pattern was adopted because other school related information was collected in 1997 (Math, Reading, and Writing) or 1998 (Science and Social Studies), and this information was introduced into the analyses to explain variations in the 1997 or 1998 MSPAP performance and rates of change among schools.

The structure or distribution of the residuals (Level 1 error models) 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 2, 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 2 presents results from estimating the Level 1 model for the various MSPAP subject areas. Included in the table are the regression coefficients relating the slope latent variable to the MSPAP scale scores over time. Note that the regression coefficients related to the intercept latent variable are fixed at 1 as indicated in Figure 2. Also included are means, variances, and covariances for both the intercept and slope latent variables. Fit statistics are presented in the context of the full models (Level 2 models) that are discussed later. It should be noted that for all but the Math subject areas, the 1995 time-point was deleted from the analyses in order to attain an acceptable model-data-fit. For unexplainable reasons, data from this time-point were found to be problematic. However, it is interesting to note that for four subject areas (Math, Reading, Science and Social Studies), this time-point was associated with the "leveling off" period in performance that was observed in Table 1.

Table 2: Results for the Level 1 Growth Model

	Math		Writing		Reading		Science		Social Studies	
	r	SE	r	SE	r	SE	r	SE	r	SE
Regression Coefficients:										
93 Score ← Slope	4		2.8*	.22	4		5		5	
94 Score ← Slope	3		3		3		3.3*	.31	3.4*	12.8
95 Score ← Slope	1.4*	.34								
96 Score ← Slope	1		1		1		2	.25	2	
97 Score ← Slope	0		0		0		1.8*		1	
98 Score ← Slope							0		0	
Latent Variable Means:										
97 or 98 Performance	521.6*	2.47	514.1*	2.4	516.6*	1.8	523.6*	2.1	518.0*	1.8
Slope	-2.7*	.26	-3.4	.42	-3.4*	.28	-2.8*	.24	-3.0*	.22
Variances/Covariances:										
97,98 Perform, Slope			-36.9*	10.6	8.0*	4.9			-11.8*	4.7
97,98 Performance	496.7*	79.4	458.8*	73.9	264.7*	43.7	473.6*	65.7	349.5*	50.3
Slope	2.4*	1.0	9.1*	2.6	3.9*	1.2	3.2*	1.6	2.7*	.88
e1	31.5*	15.2	36.0*	9.1	24.0*	10.1	44.2*	15.2	38.3*	12.0
e2	32.7*	12.6	29.2*	9.4	41.7*	8.6	75.5*	12.6	58.3*	9.7
e3	71.0*	9.8								
e4	23.3*	13.8	42.3*	9.3	23.9*	5.9	52.4*	9.8	44.3*	7.5
e5	47.3*	14.3	24.9*	13.5	33.6*	9.1	73.9*	13.8	42.5*	7.9
e6							45.2*	14.3	42.5*	9.8

From the table, it can be seen, the mean 1997 or 1998 MSPAP performance (intercept factor) across the schools was comparable across the subject areas (range - 514.1 to 523.6). In addition, significant mean rates of change (slope factor) over time across the schools that were also comparable across the subject areas (range –2.77 to –3.4) were also found, although the rates of change were modest given the scale of the test scores. Recall that the rate of change is associated with a decrease in performance from 1997,98 to 1993. Conversely, this result suggests that there was a significant increase in performance from 1993 to 1997,98. The results in the table also indicate that a non-linear rate of change was estimated in the model for most of the subject areas. For example, the pattern of coefficients for Science indicate that a larger than average change was apparent between 1993 and 1994 (estimated coefficient of 3.3 versus a fixed coefficient of 4), followed by a leveling off, and then a larger than average change between 1997 and 1998 (estimated coefficient of 1.8 versus a fixed coefficient of 1). The chi-square difference between a model assuming linear change and the non-linear rate of change model was significant and this result was consistent with the results in Table 1.

The variances for 1997 or 1998 MSPAP performance and rates of change for the various subject areas indicate significant variability in these parameters across the schools. In addition, the covariance between 1997,98 MSPAP performance and rates of change was significant for Writing, Reading, and Social Studies. The corresponding correlation coefficients were -.57, -.60, and -.39. The direction of the covariance or correlation (negative) indicated that higher rates of 1997,98 performance were associated with less negative rates of decline from 1997,98 to 1993, or, lower rates of change were associated with higher performance levels in 1997,98. Conversely, this indicated that higher rates of change were associated with lower performance levels in 1997,98. However, the covariance between 1997,98 MSPAP performance and rates of change was not significant for Math and Science and was thus fixed at 0. 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 covariances between 1993 MSPAP performance and rates of change with correlations of -.40 and -.46 for Math and Science, respectively. This indicated that higher rates of change were also associated with lower initial performance in 1993. The finding of a significant correlation between 1993 performance and rates of change may suggest that the rate of change in Math and Science is more similar for schools in 1997,98 than in 1993. Finally, note that although a fully unstructured error model was not required, two covariances between errors for the 1997 and 1998 MSPAP scores for Science were significant and required estimation. For the other subject areas, a model with independent but unequal error variances was assumed.

Explaining Differences in Performance Levels and Rates of Change- Level 2 Growth Analyses

The structural component of the structural equation model was used to reflect factors which were hypothesized to explain variability in 1997 and 1998 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 variablility in intercepts and slopes), and ζ are structural residuals.

Figure 3 presents the Level 2 (Conditional) growth models for Science and Social Studies in the present study. A school variable (percent free lunch), and a limited number of variables from the teacher and student questionnaires were introduced into the growth model to explain variability in 1997,98 MSPAP performance and rates of change across the schools. For the Math, Reading, and Writing subject areas, no student variables were introduced into the analysis due to the reduced sample sizes. In addition, for Reading and Writing, an additional variable that was constructed from the teacher questionnaire was used: Use of Reform-Oriented Problem Types (e.g., use of essays, journals, etc.). This variable was imbedded in the Current Instruction variables for the other subject areas, but evaluated separately for Reading and Writing given the nature of the combined Reading/Writing (Language) teacher questionnaire. The structural residuals were specified by the latent variables d1 and d2 in the figure, and the relationship between 1997,98 MSPAP performance and rate of change was estimated through these two residual parameters. Note that, in theory, it would be possible to incorporate the confirmatory factor analysis model for the questionnaires directly within the growth model rather than use the derived variables. However, given the sample size in the present study, such a model was overly complex to be estimated.

From the figure it is clear that the predictors of growth reflect information collected at one time-point (1997 or 1998) and therefore reflect a static snapshot of these variables. It is very reasonable to believe that with the introduction of an assessment program we would not only expect changes in performance on the assessment over time, but we would also expect corresponding changes over time in classroom instruction and assessment practices, professional development activities, students' and teachers' beliefs about and attitude towards MSPAP, and possibly students' motivation. Ideally, the type of information that is collected to evaluate the various impacts of an assessment program should be introduced with a new assessment program and collected concurrently with the assessment data. However, the present study examined the impact of an assessment program that has already been in place for several years. Thus, this study focused on current perceptions by teachers, principals, and students as well as retrospective perceptions by teachers and principals since the inception of MSPAP.

With regard to school characteristics such as percent free lunch, it would also be reasonable to expect that these variables change over time and that these changes could be related to changes in schools' performance over time. Since the percent free lunch data was available from 1993 to 1998 for schools in Maryland, an unconditional growth model for this variable that was similar to that for the MSPAP mean school scale scores was examined. The means from 1993 to 1998 increased gradually from 32.4 in 1993 to 37.2 in 1998 (n=962) with an average change from year to year of less than one percent. The variance components from the growth analysis were 608.6, 2.1, 2.0, and 13.1 for 1993 levels of percent free lunch, changes over time in percent free lunch, covariance between initial levels and change, and error, respectively. Thus, approximately 97% of the total variability in the percent free lunch variable over time was due to differences in initial levels. Since so little variance was attributable to changes in the percent free lunch variable over time, the data for a single time-point was used to be consistent with the other predictors that were studied.

Table 3 presents the unstandardized regression coefficients (r) and standard errors for the effects for the variables introduced to explain variability in 1997,98 MSPAP performance and changes in performance over time. Significant effects at α=.05 are asterisked. Non-asterisked effects represent effects that bordered on significance and the parameter for any effect that was not significant or borderline significant was fixed at 0.

The remaining variance in the intercept (1997,98 MSPAP Performance) and slope (Rate of Change) latent variables after introducing the explanatory variables is also provided along with the percent of variance accounted for by the predictors of the latent variables. These are derived by taking one minus the ratio of the variance component remaining in the conditional or Level 2 model to the variance component in the unconditional or Level 1 model. For example, the amount of variance in 1997 MSPAP Math performance accounted for by the predictors was equal to 1 – 151.2/496.7.

Finally, the following fit statistics for the models are also included in the table: chi-squared statistic, root mean squared error of approximation (RMSEA) and the normed fit index (NFI). The chi-squared statistic can be used to test the null hypothesis that the variance-covariance matrix implied by the model equals the observed variance-covariance matrix. Significant chi-squared statistics indicate a significant lack of model-data-fit. The RMSEA statistic reflects the discrepancy per degree of freedom in approximating the population covariance matrix. Values less than .05 in the RMSEA statistic or within the range of .05 to .08 are acceptable (Browne and Cudeck, 1993). NFI reflects the ratio of the fit function for the estimated model in comparison with the fit function for the independence model (uncorrelated variables). Values greater than .9 are desirable (Bentler & Bonnet, 1980). As can be seen, the fit statistics indicated acceptable model-data-fit for all but Reading. Although the model for Reading as described in the table did not fit the data adequately, it was possible to estimate a model with acceptable model-data-fit by removing the Current Instruction variable from the analysis. As can be seen in the table, this variable did not significantly explain variability in 1997 Reading Performance or rates of change over time across the schools. After removing this variable from the analysis, the chi-square statistic was 19.5 with 12 *df* (p=.08) indicating a model that did not significantly reject the null hypothesis.

Figure 3. Level 2 Growth Model with School Level Covariates

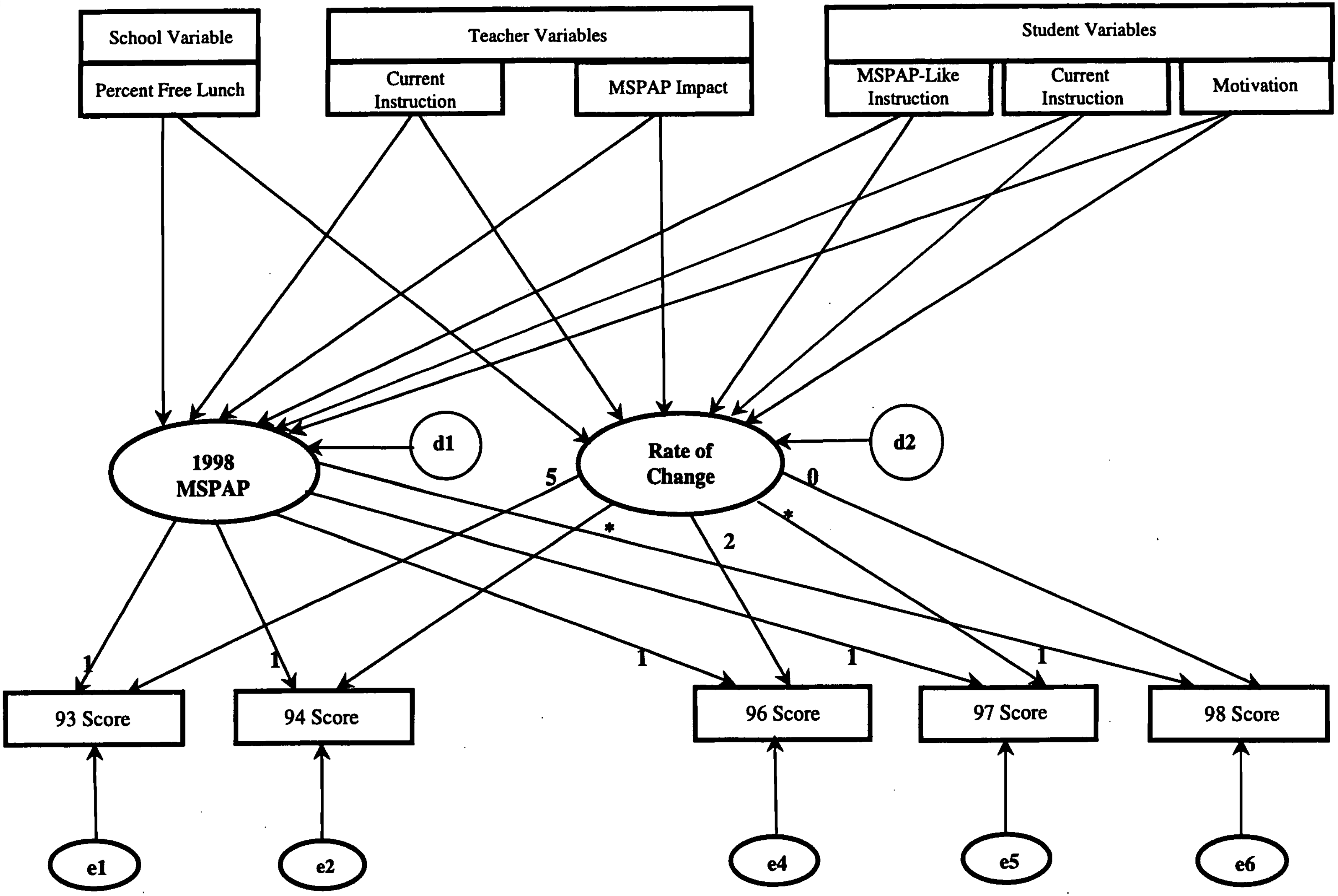


Table 3: Results for the Level 2 Growth Model with School Level Covariates

	Math		Writing		Reading		Science		Social Studies	
	r	SE	R	SE	r	SE	r	SE	r	SE
Effects: 1997 or 1998 Performance										
Percent Free Lunch	-.80*	.05	-.71*	.06	-.60*	.04	.79*	.05	-.71*	.04
Current Instruction	9.2*	4.1	0		0		5.9	3.1	0	
MSPAP Impact	0		8.5*	3.4	4.4*	2.1	0		0	
MSPAP-like Instruction							-4.0*	1.6	-2.2	1.2
Student Motivation							0		-5.7*	2.8
Reform-Oriented Tasks			10.5*	4.5	5.5	2.9				
Effects: Rate of Change										
Percent Free Lunch	0		0		-.03*	.01	0		0	
Current Instruction	0		0		0		0		0	
MSPAP Impact	-1.1*	.54	0		0		-1.1*	.40	-.62	.34
MSPAP-like Instruction							0		0	
Student Motivation							-2.1*	.63	1.0	.6
Reform-Oriented Tasks			-2.4*	1.1	-1.7*	.73				
Variances:										
97,98 Performance	151.2*	26.5	153.6*	27.9	53.2*	10.3	97.9*	17.3	95.8*	16.2
Variance Accounted For by Predictors	70%		67%		80%		80%		73%	
Rate of Change	2.3*	1.0	8.3*	2.3	2.6*	.89	.97*	.81	2.5*	.82
Variance Accounted For by Predictors	5%		8%		33%		70%		8%	
Fit Statistics:										
Chi-Square Statistic	27.0		10.5		28.3		37.1		38.6	
Df	21		17		17		26		29	
Probability	.17		.88		.04		.07		.11	
RMSEA	.06		1.00		.09		.99		.06	
NFI	.99		0		.99		.06		.99	

Explaining Differences in 1997,98 Performance Levels. As can be seen in the table, a large percentage of the variance in 1997 and 1998 performance levels was accounted for by the predictors that were introduced (67-80%). The variable Percent Free Lunch was consistently related to 1997,98 MSPAP performance with similarly sized effects across the subject areas of MSPAP. Thus, increases in the percentage of students receiving free or reduced lunch were associated with lower levels of MSPAP performance in 1997 or 1998. The regression coefficients can be interpreted as any unstandardized regression coefficient. For example, in the case of the Percent Free Lunch variable for Science, one unit change in this variable corresponds to a decrease of .79 units in mean 1998 MSPAP Science scores for schools. With the exception of Social Studies, instruction-related variables as perceived by teachers (Current Instruction and Use of Reform-Oriented Problem Types) were also found to consistently explain differences in performance levels in 1997 or 1998. Not surprisingly, as teachers' current instruction reflected increasingly the reform-oriented problem types and learning outcomes of MSPAP, the greater the performance on MSPAP.

With regard to student level predictors for Science and Social Studies, students' indication of how often they worked on MSPAP-like tasks in class and how important it was to do well on MSPAP were negatively associated with performance levels in 1998. Thus, increases in how often students worked on MSPAP-like tasks and their motivational level were associated with decreased performance in 1998 Science and Social Studies scores. Note that Current Instruction as described by students was not represented in the analyses. Both teacher and student Current Instruction variables predicted significantly 1998 MSPAP performance when included separately in the model. However, when they were included simultaneously, the effects were attenuated. Therefore, the student level variable was excluded since it was not as inclusive with regard to the classroom instruction and assessment activities. In addition, this variable was not included in the Math, Reading, and Writing subject areas since the introduction of student level predictors prohibitively reduced the sample size.

The finding that students' perceptions of the degree to which they worked on MSPAP-like tasks was negatively related with 1998 MSPAP performance was interesting, and in the case of MSPAP Science performance, there was an apparent paradox between the direction of the relationship for Current Science Instruction and student's perception of MSPAP-Like Instruction. As teachers' instruction more closely reflected the Maryland Learning Outcomes and reform-oriented problem types, higher 1998 MSPAP performance was observed. On the other hand, student perception of the degree to which they worked on MSPAP-like tasks was negatively related with 1998 MSPAP performance. Given the question "...how often did you work on tasks like those on MSPAP", students may have been focusing on the format of MSPAP tasks and not on the learning outcomes reflected in the tasks. Thus, this may reflect a greater likelihood of lower performing schools using more MSPAP-like formatted tasks than higher performing schools. Schools performing at higher levels may be more successful at reflecting the science learning outcomes in a variety of reform-oriented problem formats. The finding of a similar negative effect for Social Studies provides corroboration of this result.

Finally, teacher's perception of the degree of MSPAP Impact on classroom reading and writing activities was found to explain significantly variability in 1997 Reading and Writing performance. The direction of the effect indicates that increased perceived impact of MSPAP was associated with increased levels in 1997 MSPAP Reading and Writing performance.

Explaining Differences in Rates of Change. In contrast to explaining differences in 1997 or 1998 MSPAP performance levels, considerably less of the variability in rates of change was explained by the predictors except for the case of Science and to some extent Reading. One might argue that, since rates of change among schools were modest and there was not a lot of variability in the rates of change across schools, it may be more difficult to account for variation in rates of change as opposed to variation in levels of 1998 performance levels. While this may be true, a significant amount of the variability in rates of change for MSPAP Science scores was accounted for by the predictors.

It is interesting to note that the variable percent free lunch did not significantly account for variability in rates of change except for Reading, and in the case of Reading, the effect was so small relative to other effects that it might be considered not of practical significance. However, if one were to interpret the effect, the direction of the effect (negative) indicates that a very modest decrease in the negative rate of change over time (i.e., more change) was associated with higher levels of the Percent Free Lunch Variable, or conversely, higher levels of the Percent Free Lunch variable were associated with slightly greater rates of change in the scores from 1993 to 1997.

With regard to instruction-related variables from the teacher questionnaire, only the use of reform-oriented problem types was found to significantly affect rates of change for Reading and Writing. Thus, increased use of these types of tasks was associated with decreases in the negative rates of changes over time (or, greater rates of change in MSPAP school performance from 1993 to 1997). Recall that this variable was separated out from the Current Instruction variable for these subject areas. Although not done in the present study, extracting this component from the Current Instruction variables for the other subject areas might yield similar results.

The variable MSPAP Impact was also found to explain differences in rates of change in Math, Science, and Social Studies, as was students' motivational level (How important is it for you to do well on MSPAP?) for Science and Social Studies. This indicates that higher levels of teacher reports of MSPAP having a direct impact on Math, Science, and Social Studies instruction were associated with greater rates of decrease in performance in these subject areas from 1998 to 1993 (or higher levels of rate of change in MSPAP school performance from 1993 to 1998). However, in the case of students' motivational level, the direction of the effect differed when comparing Science and Social Studies. For Science, greater levels of student motivation were associated with greater rates of increase from 1993 to 1998, whereas for Social Studies, the opposite effect was found (i.e., greater levels of student motivation were associated with smaller rates of change in Social Studies performance over time). It should be noted that a model with acceptable fit for Social Studies could also be obtained by fixing the two regression coefficients associated the variable Student's Motivation to 0.

DISCUSSION

The purpose of this paper was to explore the relationship between changes in MSPAP scores from 1993 to 1998 and classroom instruction and assessment practices, student learning and motivation, students' and teachers' beliefs about and attitude towards MSPAP, and finally, a school characteristic. Several factors from each of these dimensions were observed to explain a significant amount of the variability in 1997 or 1998 performance of schools and rates of change in performance over time. Thus, there is some correlational evidence for the impact of the assessment program on classroom instruction and assessment practices.

Instruction-related variables were found to explain differences in MSPAP performance levels across the subject areas, and for some subject areas, explain differences in rates of change in MSPAP performance over time. In addition, the perceived impact of MSPAP on instruction/assessment practices was also found to significantly explain either differences in MSPAP performance levels or rates of change over time across the subject areas. Based on the same set of questionnaires, other analyses have also found that classroom instruction and assessment practices appeared to change over time with the educational reform movement in Maryland (Lane, Stone, Parke, Hansen, & Cerrillo, 2000). However, other studies indicate that there are still gains to be made in the congruence between classroom instruction/assessment practices and the state-defined learning outcomes and the MSPAP assessment program (Cerrillo, Hansen, Parke, Lane, & Scott, 2000).

There is also correlation evidence that, in lower performing schools, some of this change may be in the form of the use of tasks that resemble the "assessment format" rather than instruction/assessment that focuses on a variety of process learning outcomes and a broader array of reform-oriented problem types. Other analyses examining the classroom artifacts which teachers provided also found that specific test preparation materials more closely resembled MSPAP and the Maryland Learning Outcomes, as compared to classroom instruction and assessment materials. Finally, a school characteristic, percent free lunch which was used as a proxy for SES, was found consistently to be related to MSPAP performance levels but generally not related to rates of change in performance over time.

Although the results should be interpreted cautiously since the samples were relatively small, there was a high degree of similarity in the pattern of findings across the five subject areas. Thus, this cross-validation of the results provides some degree of generalizability in the findings. Further, it should be emphasized that the results for the Math and Language subject areas in comparison with the results for Science and Social Studies involved a different set of schools and were from different instructional years.

As noted above, the design of such a validity or impact study could be improved by measuring the outcomes in the present study concurrently with assessment performance over time. Thus, changes in classroom instruction and assessment practices, student learning and motivation, professional development, students' and teachers' beliefs about and attitude toward the assessment program could be examined in connection with changes in assessment performance over time. Although school characteristics may or may not change appreciably over time, these could be measured at each time-point.

One of the advantages of estimating growth curve models in a SEM framework is that more general analyses can be conducted, such as models with multiple outcome variables with different growth processes.

Finally, the present study may be improved by examining the growth processes separately for elementary and middle schools as well as incorporating a three level model. The present study combined elementary and middle schools in order to increase the sample size for the analyses. However, other studies found differences on the teacher questionnaire dimensions between elementary and middle school teachers (e.g., Lane et. al., 2000). It would be interesting to explore any differences in predicting factors related to the performance and changes in performance over time for elementary versus middle schools. The present study also involved a two level model where the unit of analysis involved measurements at the school level and variability in the schools was examined. In a three level model, measurements at the class level provide the repeated measurements at Level 1, variation in classes within schools is modeled at Level 2, and finally, variation among schools is modeled in Level 3. It would be expected that teachers would vary within schools and variables could be introduced to explain differences between teachers as well as variables introduced to explain variation in schools.

References

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

Bentler, P.M. & Bonnet, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: Applications and data analysis. Beverly Hills, CA: Sage.

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.

Cerrillo, T.C., Hansen, M.A., Parke, C.S., Lane, S., & Scott, K. (2000). The relationship between MSPAP and science classroom instruction and assessment materials. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

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.

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.

Laird, N.M. & Ware, J.H. (1982). Random-effects models for longitudinal data. Biometrics, 38, 963-974.

Lane, S., Stone, C.A., Parke, C.S., Hansen, M.A., & Cerrillo, T.L. (2000). Consequential evidence for MSPAP from teacher, principal, and student perspectives (Mathematics, Reading, Writing, Science, and Social Studies). Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Lane, S., Parke, C.S., & Stone, C.A. (1998). Consequences of the Maryland school performance assessment program. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Linn, R.L. (2000). Assessments and accountability. Educational Researcher, 29, 4-16.

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. (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.

Resnick, L.B. & Resnick, D.P. (1992). Assessing the thinking curriculum; New tools for educational reform. In B.G. Gifford & M.C. O'Connor (Eds.), Changing assessments: Alternative views of aptitude, achievement and instruction (pp. 37-75). Boston: Kluwer Academic Publishers.

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.

NOTICE

Reproduction Basis