DOCUMENT RESUME

ED 412 230 TM 027 456

AUTHOR Hargrove, Linda L.; Mao, Michael X.
TITLE Three-Level HLM Modeling of Academic and Contextual Variables Related to SAT Scores in Texas.
PUB DATE 1997-03-25
NOTE 26p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, March 25-27, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *High School Students; High Schools; *Institutional Characteristics; Mathematical Models; Multivariate Analysis; School Districts; *Scores; *Test Results
IDENTIFIERS Hierarchical Linear Modeling; *Scholastic Assessment Tests; *Texas

ABSTRACT

By including district variables at a third level, this study extended previous two-level hierarchical linear modeling (HLM) of multivariate relationships among Texas school-level characteristics and within-school mean Scholastic Assessment Test (SAT) scores and score differentiating factors (L. L. Hargrove, M. X. Mao, and G. Barkanic, 1996; Hargrove and L. T. Mellor, 1994; Hargrove, Mellor, and Mao, 1995). Most consistently, school percentage of students passing all state high school exit examinations, Advanced Placement (AP) examinees with scores of three to five, and students taking AP examinations related positively to school average scores and helped explain substantial total variance in separate three-level HLM models for verbal, mathematics, and total scores, while the district percentage passing all state exit assessments was positively linked to district average score in math and total score models. Similar and other school and district factors described smaller amounts of school and district variance in the average score differentiating effects of student grade point average, rank, advanced coursework, and college English placement/credit plans, further clarifying some construct patterns previously noted. Most importantly, it is noted that variables in these models were predominantly educational rather than demographic in concept, and thus among the more readily alterable variables of the educational context, using the (1989) terminology of J. Oakes. A number of findings were consistent with the previous modeling of Texas scores and have implications for policies regarding the use and reporting of SAT summary score results. (Contains 6 tables and 42 references.) (Author/SLD)

Three-Level HLM Modeling of Academic and Contextual Variables Related to SAT Scores in Texas

Linda L. Hargrove
Michael X. Mao

Texas Education Agency
Research and Evaluation Division
1701 North Congress Avenue
Austin, Texas 78701-1494
512-463-9701
E-mail: lhargrov@tmail.tea.state.tx.us

Abstract

By including district variables at a third level, this study extends previous two-level hierarchical linear modeling of multivariate relationships among Texas school-level characteristics and within-school mean SAT scores and score differentiating factors (Hargrove, Mao, & Barkanic, 1996; Hargrove & Mellor, 1994; Hargrove, Mellor, & Mao, 1995). Most consistently, school percentage of students passing all state high school exit assessments, Advanced Placement (AP) examinees with 3-5 score(s), and students taking AP exam(s) related positively to school average scores and helped explain substantial total variance in separate three-level HLM models for Verbal, Math, and Total scores, while the district percentage passing all state exit assessments was positively linked to district average score in Math and Total score models. Similar and other school and district factors described smaller amounts of school and district variance in the average score differentiating effects of student GPA, rank, advanced coursework, and college English placement/credit plans, further clarifying some construct patterns previously noted. Most importantly, variables in these models were predominantly educational rather than demographic in concept and, hence, among those more readily "alterable" variables of the educational context, using Oakes' (1989) terminology. A number of findings were consistent with the previous modeling of Texas scores and have implications for policies regarding the use and reporting of SAT summary score results.

Presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, Illinois, March, 25, 1997.

The views expressed in this paper are those of the authors and, thus, are neither the same, necessarily, as views held by staff of the Texas Education Agency nor the same as those held by State Board of Education members.

Three-Level HLM Modeling of Academic and Contextual Variables Related to SAT Scores in Texas

Linda L. Hargrove and Michael X. Mao

Previous hierarchical linear modeling (HLM) studies conducted by the authors (Hargrove, Mao, & Barkanic, 1996; Hargrove & Mellor, 1994; Hargrove, Mellor, & Mao, 1995) identified multivariate relationships among a number of between- and within-school variables related to 1991, 1993, and 1994 Scholastic Aptitude Test (SAT) Verbal, Math, and Total scores of Texas graduating seniors--predominantly academic context variables. In last year's models, school percentage of students passing state high school exit assessments, Advanced Placement (AP) examinees scoring 3-5, AP exam participation, diversity in non-AP advanced courses completed, and percentage of white teachers related positively to school average scores, explaining substantial total variance in separate two-level models for Verbal, Math, and Total scores. Similar school factors described within-school score differentiating effects such as student college English placement/credit plans, grade-point average (GPA), rank in class, and advanced coursework and suggested some discernible construct patterns. HLM (e.g., Arnold, 1992; Bryk and Raudenbush, 1992) had clearly allowed for more precise two-level specification and modeling of these relationships than standard single-level multiple regression models (cf. Texas Education Agency [TEA], 1990), given the hierarchical nesting of student SAT scores and other student data within schools that varied dramatically in key demographic and academic characteristics. For this study, three-level HLM modeling extended to the district level (rather than two-levels only) was integral to the further exploration of Class of 1994 Texas SAT score data because score results and their interpretation continue to have serious impact on educational experiences and achievement at the student, school, and district levels. This is especially so as new indicators and academic context variables are considered for public school and district performance indicator accountability system uses in Texas, as well as in other states, and as publicly available school- and district-level demographic and academic context data become more widely used in college admissions decisions in Texas and elsewhere in lieu of individual ethnicity variables.

Score Use and Indicator Context

Along with the American College Testing Program's ACT Assessment, SAT participation and performance results continue to be used as relatively low-stakes (sanction-free) indicators in determining non-monetary recognitions of districts and schools and for reporting, along with other indicators, in the statutorily-mandated accountability reports of district and school performance results to local communities, the state legislature, and the general public (TEA, 1995a). Periodically, new indicators are considered for these reports. Thus, in fall 1996, a statutory indicator of recommended high school graduation program completion by students, which incorporates completion of advanced coursework and credit by examination, was included, and the State Board of Education adopted College Board AP Exam participation and performance indicators for accountability reporting, as defined in TEA (1995c). With the recent and dramatic growth in the Texas AP program (e.g., College Entrance Examination Board [CEEB], 1994c; TEA, 1995c), interest in the extent to which taking AP courses, AP exams, and other advanced coursework (e.g., Herr, 1993) relates to SAT scores has intensified. Previously, Hargrove (1992) and Hargrove and Mellor (1993) examined other school characteristics, besides indicator performance, associated with recognition under Texas' 1992 and 1993 school award system; Cornett & Gaines (1993, 1994) summarized characteristics of and criteria used in these systems in Texas and in other states. In addition, a number of others (e.g., Bryk & Hermanson, 1993; Darling-Hammond, 1992; Oakes, 1989; Smith, 1988; Willms & Kerckhoff, 1995) have researched and/or critiqued the structure, interpretation, and uses of educational performance indicator systems in general. Others (e.g., L. Burt, The University of Texas at Austin, personal communication, February 11, 1997) are proposing uses of Texas' school performance indicator data along with other information as part of an adversity index in determining higher education scholarship awards for students and as part of the admissions process.

Apart from SAT score use in performance indicator systems, educators and the public are updated annually with the release of SAT summary score results reported by state of residence, ethnicity, gender, family income level, parent education level, amount and type of academic coursework, grade point average (GPA), urban/rural versus suburban locality, and so on (e.g., College Entrance Examination Board [CEEB], 1994a,b). A number of articles point to fallacies inherent in state-versus-state and other similar simple comparisons of aggregated SAT data, given the extent to which testing participation rates and other variables also correspond to state and other aggregate mean scores (Holland & Wainer, 1990; Page & Feifs, 1985; Taube & Linden, 1989; Wachter, 1989; Wainer, 1986, 1989a, 1989b, 1990; Wainer, Holland, Swinton, & Wang, 1985). Summaries more specific to district- or school-level results (e.g., Fetler, 1991; TEA, 1990,1996) note differences in SAT scores related to school percentages of economically disadvantaged, minority, and limited English-proficient (LEP) students, as well as district wealth, enrollment, average school achievement, teacher characteristics, and test-taking participation. TEA (e.g., 1995a) has acknowledged the relationship between district/school performance indicators, including the SAT (and ACT), and demographic characteristics through performance comparisons of districts/schools grouped by demographics.

More specifically, TEA (e.g., 1990, 1996) reports and others (e.g., Everson, Millsap, & Diones, 1995) have made distinctions between the non-academic (i.e., demographic or background) and academic (i.e., achievement and experience) types of variables related to SAT test score results, where the academic or educational variables are those most subject to control by educators. Oakes (1989) termed such academic variables as the "*alterable characteristics* of interest to educators and policy makers" (p. 186) and argued for the consideration of such educational context variables along with the analysis of district/school performance indicators. She also suggested three general constructs for describing school context indicators--student access to knowledge, school press for achievement, and professional teaching conditions within schools. To a limited extent, Texas' performance indicator system includes reporting of some of these academic variables (some mentioned earlier) along with the reporting of SAT (and ACT) test score data and other district/school performance. However, little guidance has been provided for the interpretation of performance results given observed relationships with both academic and non-academic context variables. One exception to this was a TEA (1990) report of multiple regression analysis results, which showed academic variables describing more of the total variance in students' SAT Total scores than student demographic variables. Clearly, need continues for analysis that can more explicitly inform the appropriate interpretation and relative importance of student, school, and district assessment and other performance data.

Study Objectives

Objectives for this study's three-level HLM modeling of 1994 SAT scores included: (a) building a third level into the modeling by including district-level variables along with student- and school-level variables previously included in two-level models developed with 1994 Texas SAT scores; and (b) evaluating the viability and explanatory power of the new, separate three-level HLM models for each set of SAT Verbal, Math, and Total scores.

Method

Given previous experience with the Bryk, Raudenbush, and Congdon (1994) HLM™2_3computer program, a pragmatic plan for data preparation, selection of variables, and analysis sequence was continued with the recent HLM™ program described in Bryk, Raudenbush, and Congdon (1996) to execute three-level HLM. Computational improvements made in the previous Bryk, Raudenbush, & Congdon (1994) HLM™2_3 PC computer program had removed previous limitations (e.g., Bryk, Raudenbush, Seltzer, and Congdon, 1989) on numbers of variables and Level-2 units, but physical constraints of individual PC system memory capacity and/or configuration still imposed processing and analysis limitations (J. Shao, personal communication, March 8, 1995). Because this study was extending two-level modeling executed with the previous HLM™2_3 program to three-level modeling on the same data set with the most recent HLM™ program version, our previous study's data preparation and some analysis procedures were extended to the district level in this study.

Data Sources

Three data sources were used in the analyses: (a) most recent SAT test score and Student Descriptive Questionnaire (SDQ) data from 1993/94 graduating seniors in all Texas public schools provided directly to TEA, with school permission, by the College Board from its Admissions Testing Program (cf. TEA, 1996); (b) 1994 AP exam data also released to TEA by the College Board (cf. TEA, 1995c); and (c) data on school- and district-level variables from TEA's Academic Excellence Indicator System (AEIS) (cf. TEA, 1995a,e) and Public Education Information Management System (PEIMS) (cf. TEA, 1995b). The Glossary in this paper's Appendix includes a grouping of study variables by Oakes' (1989) three general constructs for school context indicators, along with descriptions and value codes for all student, school, and district variables explored and ultimately included in this study's three-level HLM models; those explored in our previous two-level HLM models of 1991, 1993, and 1994 scores appear the Hargrove et al. (1996) Appendix. Descriptive statistics for variables in the Glossary are shown in Table 1.

Data Preparation

Additional preparation using the mainframe computer was necessary to include district-level data with our previous student- and school-level 1994 data before they could be submitted to the latest HLM™ PC program. Decisions about recoding variables, handling missing data, and defining the sampling populations and samples were required for data at the district level, just as had been the case previously for data at the student and school levels.

Recoding variables. Predictor variable values at each level were either implicitly numeric (i.e., ordinal, quantitative, or continuous) or were strictly non-numeric (non-ordinal) or categorical. Original coding was retained for the quantitatively valued variables and the dichotomously coded categorical variables; however, variables that reflected greater than two category values were recoded dichotomously through combining or collapsing two or more categories into one. Dichotomous coding was employed rather than the dummy coding of each category into one variable to keep the number of potentially relevant within-school variables small because of the practical limits of any given HLM analysis to handle a large number of within-school variables.

Standardizing between-district variables. To facilitate the comparable interpretation of between-school Gamma coefficients relative to any other Gammas in the two-level equations used to predict the HLM estimated within-school intercept and slope parameters, Bryk and Raudenbush (1992) and Arnold (1992) recommend unit standardization (i.e., mean=0; standard devation=1) of the values for between-school variables. Unit standardization was also applied for to district variables for the same reason for three-level modeling in this study.

Handling missing data. Nonmissing values are required on all school- and district-level variables included in three-level analysis (Bryk et al., 1996). With rare exception, most regular high schools in Texas with SAT data and schools with at least 20 SAT-tested students reflected nonmissing data on the school- and district-level variables. Thus, schools and districts with missing data on any of the main variables of interest were excluded from analysis rather than including them with mean or other values substituted for the missing data. In contrast, Arnold (1992, citing personal communication with S. Raudenbush, 1991) noted that HLM within-school parameter estimation remains relatively unaffected by missing values in the student data as long as adequate numbers of students with nonmissing values (i.e., so that there are adequate degrees of freedom) remain to support the number of within-school variables in the analysis.

Treatment of problematic and non-varying school and district variables. Schools with student-level variables exhibiting no variance (e.g., all White students for ethnicity) are excluded from the computation of HLM reliability and Chi-square statistics but are used in the estimation of the between-school Gamma coefficients (between-school intercepts and slopes) and residual parameter variances (Taus) associated with each within-school predictor variable (Bryk and Raudenbush, 1992; Bryk et al., 1989, 1994). Similar treatment is given to student data within a school in the HLM computations if they present a problem for matrix inversion (J. Shao, personal communication, March 28, 1995). When such exclusions occur, generalizing from the school to district level, interpretative comparisons of reliabilities and Chi-square statistics are no longer appropriate across slightly varying between-school and between-district models predicting the same within-school parameters. Because of the exploratory nature of this study, the decision was made to exclude districts with no variance on school variables, in addition to the previous exclusion of schools showing no variance and those with matrix inversion problems on student variables of fundamental interest. This worked in tandem with the other defining feature of the within-school sampling population described above--that is, including only those schools with 20 or more SAT-tested students.

Student, School, and District Samples

Sample selection for this year's three-level HLM modeling started with last year's two-level modeling sample.

Two-level sample. Analysis indicated one or more SAT-tested graduates (74,979 or 49.9% of graduates) from 1,111 Texas public high schools during school year 1993/94. Of those schools, 612 (55.1%) with 71,272 (95.1%) of the state's SAT-tested students showed 20 or more SAT-tested graduates; from those schools, 319 (28.7%) with 56,150 (74.9%) of the SAT-tested students showed greater than zero variance on key student variables considered for analysis. When those 319 schools in the student file were matched with schools reflecting nonmissing values on school-level variables of interest, a total of 305 schools remained in the sampling population. From those 305 schools, student- and school-level data for another 13 schools with matrix inversion problems were deleted. With the 292-school (26.3%) sample, data on 52,707 students (70.3% of the state total) remained in the student file.

Three-level sample. Only 169 (15.2%) high school campuses were members of districts with two or more high school campuses. Including only the 37,460 (49.9%) examinees within these 169 schools and 48 districts with multiple high school campuses helped ensure variance among schools within districts. After these data were submitted to the Bryk et al. (1996) HLM™ program, listwise deletion of student records with missing data on at least one of the four student predictor variables processed for the HLM sufficient statistics file yielded 32,877 (43.8%) student records, along with the 169 campus and 48 district records in the three-level data modeled.

Analysis

Key analysis features included the "intercept and slopes as outcomes" modeling at Level 2 and "intercepts-only as outcomes" at Level 3, modeling approach, statistics for interpreting results, and HLM™ program options executed.

Three-level HLM models. Although one equation can be used to formulate an HLM model, it is generally helpful for readers familiar with multiple regression analysis to see the model split into its three-level analog. Bryk and Raudenbush (1992) provide a number of three-level examples of formal HLM model formulations. For the within-school component of these models, predictor variable models for each of the three student-level dependent variables--SAT Verbal, Math, and Total scores--were specified. In the classical sense of single dependent variable multiple regression based on Arnold's (1992) two-level and Bryk and Raudenbush's three-level formulation, the general conditional HLM model ***within each school and district*** was:

$$Y_{mn} = P_{0mn} + P_{1mn}A_{1mn} + \ldots + P_{emn}A_{emn} + E_{mn}, \qquad (1)$$

where: Y_{mn} is the predicted SAT Verbal, Math, or Total score for students in school *m* and district *n*,
P_{0mn} is the intercept, or mean SAT score for school *m* in district *n*,
P_{1mn} is the Pi or slope coefficient for the first predictor variable for school *m* in district *n*,
P_{emn} is the Pi or slope coefficient for the e^{th} predictor variable for school *m* in district *n*,
A_{1mn} is the first predictor's value for school *m* in district *n*,
A_{emn} is the e^{th} predictor's value for school *m* in district *n*, and
E_{mn} is the random error of prediction for school *m* in district *n*.

For the between-school component of the overall three-level model, each of the within-school Pi parameters served as dependent variables for regression on a set of between-school predictors. Thus, the conditional model formulation of the set of ***school within district*** equations was:

$$P_0 = B_{00} + B_{01}X_{01} + \ldots + B_{0t}X_{0t} + R_0 \qquad (2)$$
$$P_1 = B_{10} + B_{11}X_{11} + \ldots + B_{1t}X_{1t} + R_1 \qquad (3)$$
$$P_e = B_{e0} + B_{e1}X_{e1} + \ldots + B_{et}X_{et} + R_e \qquad (4)$$

where: P_0 is the predicted intercept or average within-school SAT score across *mn* schools,
P_1 is the predicted Pi or average within-school slope coefficient for first predictor across *mn* schools (e.g., mean score "gap" or differentiating effect of the first within-school predictor on SAT scores),
P_e is the predicted Pi or average within-school slope coefficient for the e^{th} predictor across *mn* schools (e.g., mean score "gap" or differentiating effect of the e^{th} within-school predictor on SAT scores),
e is the number of prediction equations for the within-school parameters (i.e., one equation per parameter),
B_{e0} is the Beta intercept coefficient in the e^{th} equation,
B_{e1} is the Beta slope coefficient for the first school-level predictor in the e^{th} equation,
B_{et} is the Beta slope coefficient for the t^{th} school-level predictor in the e^{th} equation,
X_{e1} is the first school-level predictor variable's value in the e^{th} equation,
X_{et} is the t^{th} school-level predictor variable's value in the e^{th} equation, and
R_e is the between-school *mn* random error for the e^{th} equation.

In the "intercepts only" modeling for the between-district component of this study's three-level HLM models, only the within-district Beta intercept parameters served as dependent variables for regression on a set of between-district predictors, while Beta slope parameters were fixed (no Level-3 predictors and $R_e = 0$). The conditional model formulation for the ***between-district*** equation set follows:

$$B_{00} = G_{000} + G_{001}S_{001} + \ldots + G_{00t}S_{00t} + U_0 \qquad (5)$$
$$B_{01} = G_{010} \qquad (6)$$
$$B_{0t} = G_{0t0} \qquad (7)$$
$$B_{10} = G_{100} + G_{101}S_{101} + \ldots + G_{10t}S_{10t} + U_1 \qquad (8)$$
$$B_{11} = G_{110} \qquad (9)$$
$$B_{1t} = G_{1t0} \qquad (10)$$
$$B_{e0} = G_{e00} + G_{e01}S_{e01} + \ldots + G_{e0t}S_{e0t} + U_e \qquad (11)$$
$$B_{e1} = G_{e10} \qquad (12)$$
$$B_{et} = G_{et0} \qquad (13)$$

where: B_{00} is the predicted intercept or average within-district SAT score across *n* districts,

B_{0t} is the predicted Beta or average within-district slope coefficient across *n* districts for the t^{th} school predictor in the predicted *mn* school intercept equation (e.g., mean score "gap" or differentiating effect of the school predictor on scores),

B_{e0} is the predicted Beta or average within-district slope coefficient across *n* districts for the e^{th} student predictor (e.g., district mean "gap" or differentiating effect of the e^{th} student predictor on scores),

e is the number of prediction equations for the within-school parameters (i.e., one equation per parameter),

t is the number of prediction equations for the within-district parameters (i.e., one equation per parameter),

G_{e00} is the Gamma intercept coefficient in the B_{e0} equation,

G_{e0t} is the Gamma slope coefficient in the B_{e0} equation for the S_{e0t} predictor term,

G_{et0} is the Gamma intercept coefficient for the equation predicting the t^{th} school predictor in the e^{th} within-school equation,

S_{e0t} is the t^{th} district-level predictor variable's value in the B_{e0} prediction equation, and

U_e is the between-district random error for the e^{th} equation predicting within-district intercept coefficients.

When equations (2) through (4) are stated in the unconditional form--that is, with only the between-school Beta intercept (B_{e0}) and error (R_e) terms predicting P_0--B_{e0} represents the average within district intercept coefficient for equation *e*. Similarly, if equations (5), (8), and (11) are unconditionally specified, then G_{e00} represents the grand average intercept coefficient in these equations. These models were useful in gauging the extent to which school and district variation existed for specific within-school and within-district model parameters and in evaluating the extent to which variance was explained by the school and district predictors that were entered into subsequent models.

<u>Modeling approach.</u> The basic modeling approach was to build up from within-school models identified with last year's two-level modeling to identify school predictors explaining substantial amounts of within-school variation, followed by identifying district predictors explaining significant amounts of within-district variation and using parallel specification of predictors for all outcomes modeled within each level (e.g., Bryk and Raudenbush, 1992). In last year's two-level modeling, predictors in within-school models exhibiting little between-school variation were candidates for elimination in subsequent models explored, which helped to keep the number of within-school variables small, as Bryk et al. (1989) recommend for exploratory analysis. This also served to maximize the potential that predictors could be found to explain between-school variance and thereby explain larger amounts of the total variance for a given within-school parameter.

<u>HLM statistics.</u> HLM statistics (e.g., Arnold, 1992; Bryk and Raudenbush, 1992; Bryk et al., 1989, 1994, 1996) used in evaluating results of three-level model development included:

- ***Sigma-square*** - proportion of total dependent variable variance potentially explainable in the within-school model;
- ***reliability*** - proportion of total variance for a within-school or within-district model parameter that is unexplained parameter variance (i.e., as school or district predictors are added to the HLM model, the reliability for the model parameter at the corresponding level should decrease);
- ***t***-statistic - value for testing the univariate significance of the Beta and Gamma coefficients (i.e., suggestive rather than definitive of HLM multivariate predictor associations) (Bryk, 1991, cited by Arnold, 1992);
- ***Chi-square*** - value for testing the significance of the null hypothesis of zero residual parameter variance (Tau) for either Beta/Gamma intercepts or Beta/Gamma slope coefficients;
- ***R-square**** - proportion of parameter variance explained for each within-school and within-district parameter (i.e., distinct from multiple regression R^2);
- ***deviance*** - a multiparameter measure of model fit (i.e., tests the significance of variances and covariances estimated from HLM based on the likelihood ratio test), with lower values indicating better model fit; and
- ***H***-statistic - value (with a Chi-square distribution and *m* degrees of freedom) for testing the significance of the difference between deviances from any pair of nested three-level models (i.e., a more complex model specification versus a reduced specification of the same model can be tested in three-level HLM, per Bryk et al., 1996).

Each within-school (Level-1) SAT Verbal, Math, or Total score model (identified from last year's study) with the smallest number of predictors best explaining large amounts of within-school variance (indicated by the difference in Sigma-square from the no-predictor model) and reliabilities indicating highest amounts of between-school variance relative to other models was selected for further modeling with between-school predictors. Then, R-square*, the difference between residual variances for a within-school parameter in the null model versus a model with between-school predictors, was used to determine the best Level-2 model in conjunction with the Chi-square tests. The same process was applied in evaluating null (unconditional at Level-3 only) vs. more complex Level-3 models of within-district parameters. At Level-2 only, the proportion of within-school parameter variance explained relative to its corresponding total variance was obtained by multiplying its reliability in the null (unconditional) model, by its R-square* in the conditional, or between-school predictor model. Because Chi-square tests of Level-2 slope parameter variances at Level-3 indicated nonsignificant variation, models of these parameters were fixed at Level-3 (i.e., Level-3 intercepts only in models). Of course, the objective was to maximize R-square* for each within-school and within-district parameter predicted in the three-level model.

HLM™ PC program options. HLM™ PC program options executed included: (a) centering within-school/within-district variables around their respective school/district means; (b) using 400 iterations for the HLM solution estimation routine (although fewer than 50 and generally fewer than 20 iterations were sufficient); (c) printing results of empirical Bayes (EB) residuals regressed on between-school/between-district variables not used in the model; and (d) employing the ***H***-statistic for testing a model's significance vs. a reduced, nested model. The group (school/district) mean centering for within-school/within-district variables facilitated interpretation such that the intercept in the within-school/within-district equation represented the average SAT score when other predictors in the equation were zero and, hence, at their average values (Arnold, 1992; Bryk & Raudenbush, 1992). Bryk et al. (1992) speak of the ***t***-statistics--the highest ***t***-value above 1.6 or so--printed from the EB analysis of between-school or between-district predictors not included in the HLM model as providing some information about the most promising predictor(s) for inclusion in the next model. However, these are to be used for general guidance only because the ***t***-values are univariate rather than multivariate statistics and apply only to inclusion of the next variable into the model conditional upon the variables already in the model.

Results

Descriptive Statistics

Descriptive statistics exhibited the full range of possible variation in SAT Verbal and Math scores (200 to 800) and SAT Total scores (400 to 1600) for the Texas student data (N=32,877) included in the HLM analysis. For this dataset, the SAT Verbal mean was 424--13 points higher than the mean for all Texas public school students. The SAT Math mean at 491 and SAT Total score mean at 915 for data used in this analysis were 17 and 30 points higher than the Texas public school averages from the complete dataset. Table 1 lists the ***n***-sizes, means, and standard deviations for the unstandardized values for all student, school, and district variables used in this study's three-level HLM modeling.

Insert Table 1 about here

Means for the student predictor variables in Table 1 described students in this analysis generally as: "B^{+}" students; ranking in the second 5th of their class; having no plans to "advance place" out of English courses (67%); and completing about 1.5 total AP and non-AP advanced courses. Gender and ethnic group representation mirrored that of all SAT examinees statewide. Examination of the unweighted means for the school-level variables shows the 169-school average percentage of economically disadvantaged students at 26 percent vs. 33 percent across secondary schools statewide, while the mean minority percentage was 53 percent vs. 39 percent for the state. Means for the remaining school-level variables closely approximated those same means statewide, except for means above those statewide for relevant secondary school enrollment (1992 vs. 816) and school percentage of graduates planning to go to college, taking either the SAT or ACT, graduating with an advanced seal on the diploma, as well as for the percentage of AP examinees with 3-5 score(s) and the number of types (diversity) of advanced (AP and non-AP) courses completed. None of the 48 districts or associated school campuses were located in non-metropolitan or rural locales according to TEA school district size/location categories. Rather, the 48 districts were located in areas categorized as either major urban, major suburban, suburban, or independent town.

Three-level Verbal, Math, and Total Models

Starting bases for the three-level HLM modeling of 1994 Texas SAT scores in this study included: (a) selecting and/or eliminating variables from the two-level HLM models of 1991, 1993, and primarily 1994 Texas Verbal, Math, and Total scores from our prior studies; (b) partial use and extension of conceptual frameworks (e.g., Oakes, 1989; TEA, 1990, 1996) for classifying potential student, school, and district predictor (or context) variables related to educational outcomes, such as SAT scores; and (c) use of ***R-square**** (percentage of variance explained) and ***H*** (difference between two deviance measures of model fit) statistics. As a result, a unique model was separately identified for each of the three dependent variables. At both the school and district levels modeled, both academic and demographic types of predictors were generally included, but they displayed some differences in their statistical significances and patterns of association with the SAT scores. All within-school predictors of Verbal, Math, and Total scores were classified as educational. Only statistically significant relationships are summarized below. *(See Table 2 for an overview of the three HLM model predictors and their directions of association and significances; see Table 3 for the proportions of parameter variance and total variance explained by these models.)*

Insert Tables 2 and 3 about here

Verbal Score Model. HLM modeling of between-school and between-district effects associated with within-school predictors of Class of 1994 SAT Verbal scores is summarized in Table 4. Because the between-school variable values were analyzed in their standardized form, the Beta intercept coefficients in equations predicting the within-school intercept and slope coefficients represent the HLM-weighted means of these coefficients for the 292 schools across 48 districts. That is, HLM weights each school's estimated value on a variable proportional to its precision, which is the reciprocal of its error variance and directly related to student sample size for the school on that variable (Arnold, 1992; Bryk & Raudenbush, 1992). Thus, estimates from schools with larger student samples receive relatively larger weights than smaller schools. This also extends to the district level with three-level HLM modeling, when the Gamma intercept coefficients represent the HLM-weighted means of within-district predictor variable intercept and slope coefficients.

Insert Table 4 about here

Given the treatment above, the HLM-weighted average SAT Verbal score was 419 across districts included in the analysis (see Table 4). The statistically significant Gamma intercept coefficients for the three within-school predictors in the model included -16.37 for GPA, 47.75 for college English placement plans, and 15.20 for number of advanced courses completed. Holding other within-school predictors in the model constant, -16 points was the HLM-weighted average differentiating effect of GPA (i.e., the lower the coded values, the higher the actual GPA). Otherwise, 48 and 15 points, respectively, were the average differentiating effects of college English placement plans and advanced courses completed, with students planning college English placement and completing more advanced courses scoring higher.

Controlling for all other variables in the model, Table 4 also shows between-school effects associated with the average within-school SAT Verbal score (intercept). For instance, for every s.d. unit increase above average in the school number of types of AP subject exams taken, the within-school SAT Verbal score averaged 12 points higher; likewise, the within-district average score also rose 7 points higher per s.d. increase in the district version of the same variable. The HLM model accorded similar average score advantages per s.d. unit increase above average in the school percentage of students passing all TAAS Exit exams taken (+9 points), AP examinees scoring 3-5 (+9 points), and students taking AP exam(s) (+10 points). In contrast, for every s.d. unit increase above average in school percentages of limited-English proficient students and number of types of AP subject course(s) taken, the SAT Verbal score averaged 9 points and 11 points less, respectively. Within-district average scores decreased 28 points for every s.d. unit increase in district minority student percentages.

One school variable, number of types of AP exams taken, was positively linked with wider than average score gaps associated with English placement plans (48-point gap) by about 8 points per s.d. unit increase above average. The percentage of students in advanced courses, per s.d. unit increase above average, corresponded with a 4-point less advanced courses completed average score advantage of 15 points, as did increases in the district percentage of minority students (-2 points) and the district percentage completing non-AP advanced course(s) (-4 points). Otherwise, this score advantage widened by 2 points per s.d. unit increase in the district number of types of AP subject exams taken, as did the GPA score advantage (+1 point) per s.d. increase in the same district variable.

Overall, this Verbal model explained 76 percent of the parameter variance, or 73 percent of total variance (.7631 multiplied by the unconditional model reliability of .954), for within-school mean SAT Verbal scores (see Table 3).

Between-school predictors accounted for 29 percent of the parameter variance in college English placement plans; 35 percent, in number of advanced courses completed; and 41 percent, in GPA. At the district level (Level-3), the model explained 75 percent of parameter variance for within-district mean SAT Verbal scores; 46 percent, for GPA; 55 percent, for college English placement plans; 51 percent, for number of advanced courses completed. Proportions of total variance estimation by level in the fully unconditional model were 86 percent for the student level; 8 percent, school level; and 6 percent, district level.

Math Score Model. Table 5 presents a model with an HLM-weighted average SAT Math score of 486 across districts and weighted average score gaps of 41 and 21 points, respectively, for student high school rank and number of advanced courses completed, with students ranking higher and completing more advanced courses scoring higher. Between-school predictors in the model described 78 percent of parameter variance and 74 percent of total variance in average SAT math scores at the school level and 79 percent of parameter variance at the district level (see Table 3). For high school rank and advanced coursework, school-level predictors described 48 and 39 percent of Level-2 parameter variance, respectively, and a respective 93 and 73 percent of Level-3 parameter variance. Overall examination of variance estimates from the completely unconditional three-level HLM analysis showed a total variance partitioning of about 87 percent at the student level, 7 percent at the campus level, and 6 percent at the district level.

Insert Table 5 about here

Of the between-school predictors of average SAT Math score, the school number of types of total (AP and non-AP) advanced courses completed had the largest impact, adding 11 points for every s.d. unit increase above the school average number, while the district percentage passing all TAAS Exit Exams taken had the largest effect (28 points per s.d. unit increase) of district predictors. The school percentage of students in advanced courses had not only the least, but the only nonsignificant impact (univariate *t* value = 1.97).

Although none of the univariate *t*-values were statistically significant for between-school predictors of rank as a within-school predictor of SAT Math, these predictors taken together contributed to the significance of the conditional model versus a saturated model according to the *H*-statistic. Of school predictors with significant *t*-values, the average 21-point SAT Math score advantage associated with student number of advanced (AP and non-AP) courses completed shrank by 4 points and 2 points, respectively, per standard deviation unit increase above average in the school percentage of students enrolled in advanced courses and in the school number of types of advanced courses completed. Of district predictors with significant *t*-values, the average score advantage linked with number of advanced courses completed lessened by 5 and 6 points, respectively, per s.d. unit increase in the district percentage of students completing AP course(s) and in the district percentage completing non-AP course(s). Conversely, this same score advantage grew as the district percentage passing all TAAS Exit Exams and district number of types of AP subject exams taken went up. Score advantages associated with rank also grew larger as the district number of types of AP subject exams taken rose but narrowed as the district percentage of students completing non-AP advanced course(s) went up.

Total Score Model. Table 6 shows the model with three academic predictors of SAT Total scores within schools over districts. The HLM-weighted average SAT Total score was 905, and weighted average score gaps were 39, 66, and 38 points, respectively, for student GPA, college English placement plans, and number of advanced courses completed.

Insert Table 6 about here

Much consistency appeared in the significant between-school predictors of average scores across the SAT Verbal, Math, and Total models. In the Total score model, increments in the school percentage of students passing all TAAS Exit Exams taken, per s.d. increase above average passing rate, corresponded with a 26-point higher average SAT Total score, whereas a similar increment in the TAAS passing rate at the district level was linked with a 50-point higher average score. Similarly, school mean SAT Total scores were boosted about 16 points for every s.d. unit increase above average school percentage of AP examinees scoring 3-5, 24 points per s.d. unit increase in the number of types of AP and non-AP advanced courses completed by students on campus, and about 22 points per s.d.

unit increase in the school percentage of enrollees taking AP exam(s). Conversely, mean SAT Total scores were reduced by nearly 22 points for every s.d. unit increase above average in the school number of types of AP subject courses completed.

One between-district predictor associated with the within-school Total score differentiating effects (Level-1 slope coefficients) was also linked with the same Verbal score differentiating effects. That is, Total score advantages associated with higher GPA and greater number of advanced courses completed decreased by 3 points and 8 points, respectively, per s.d. unit increase above average in the school percentage of students completing non-AP advanced courses. The average Total score advantage for students with plans vs. those without plans to "place out" of college English widened by 16 points per s.d. unit increase above average for the same between-district predictor and also grew by 11 points per s.d. unit as the school percentage taking AP exam(s) went up. Alternatively, an s.d. unit addition above average in the school percentage passing all TAAS Exit exams taken was linked with shrinkage in the GPA advantage for Total score by 4 points. The Total score advantage associated with the quantity of advanced courses completed was further enhanced as the district percentage passing TAAS Exit exams went up (6 points per s.d. unit above average) but was reduced as the school percentage of students enrolled in advanced courses and school number of types of AP subject courses completed rose (-8 and -6 points, respectively, per s.d. unit above average).

In sum, between-school effects described 76 percent of parameter variance and 73 percent of total variance in within-school average SAT Total scores and 34 percent, 18 percent, and 41 percent of the respective parameter variances for the GPA, college English placement planning status, and advanced courses quantity slope coefficients (see Table 3). Between-district effects explained parameter variances of 77 percent, 30 percent, 72 percent, and 46 percent for within-district average scores, GPA, college English placement planning status, and advanced courses quantity slopes. Total variance apportionment in the fully unconditional model was 85 percent for Level 1; 8 percent, Level 2; and 7 percent, Level 3.

Discussion

Commonalties and distinctions among the three-level HLM models developed for 1994 Texas SAT Verbal, Math, and Total scores are discussed, as are predictors with statitistically significant intercept and slope coefficients. In addition, these models are compared to the two-level HLM models from our previous studies. Conclusions, caveats, and potential implications of the results are discussed, and possible directions for future work are noted.

Predictors of School and District Average SAT Scores

A substantial amount of total variation in SAT average scores at both school and district levels in all three null models was unexplained parameter variance. That is, intercept reliabilities for the three models (unconditional at Levels 2 and 3), which contained within-school predictor variables only, ranged from .95 to .96 (see Table 3), and are impressive amounts of parameter (or between-school) variance relative to total variance for this parameter at Level 2. Reliabilities for the within-school predictors in the unconditional versions of this year's three HLM models ranged from .33 to .56. Lower reliabilities for the slope parameters than for the intercept has been a typical observation in other studies applying HLM in modeling school average achievement (cf. Arnold, 1993).

Between-school predictors in each of the three HLM conditional models described a substantial amount of parameter variance in school average SAT scores compared to unconditional Level 2 models--78 percent for Math and 76 percent for both Verbal and Total conditional models at Level 2 (see Table 3). Similarly, parameter variance explained in district average SAT scores, comparing to unconditional Level 3 models, was 79 percent for Math, 77 percent for Total, and 75 percent for Verbal conditional models at Level 3. Three school characteristics that were positively and significantly related to school average score in all three models included the percentage of: (a) students passing all TAAS Exit Exams, (b) AP examinees with 3-5 score(s), and (c) students taking AP exam(s). In addition, three school characteristics were positively and significantly related to school average score in one or two of the models including: (a) percentage of white (non-minority) teachers, (b) number of types of advanced (AP and non-AP) courses completed, and (c) number of types of AP subject exams taken. Two additional school characteristics, percentage of limited-English proficient students (Verbal model) and number of types of AP subject courses completed (Verbal and Total models), were negatively and significantly related to school average scores. Although the last negative effect (for types of AP courses completed) was at odds with the positive relationships between other academic predictors and school average scores, this may be a reflection of the recent rapid inclusion or expansion of AP course offerings in high schools as a result of new graduation program offerings and the AP incentive program in Texas (see TEA, 1995c). Thus, these may be schools with relatively less experience with AP courses or a wider variety of AP course offerings, which is perhaps reflected in somewhat lower school average Verbal and Total scores. For significant district predictors of district average scores, both the percentage passing all TAAS Exit exams (Math and Total models) and the number of types of AP subject exams taken (Verbal model) had positive intercept coefficients, while the percentage of minority students had a negative intercept coefficient.

Using Oakes' (1989) framework, the eight school and three district variables above were considered in this study either to be part of the school press for academic achievement, as cultivated within the school, peer, family, and community culture, or linked with access to educational opportunity. Within certain parameters, these are subject to some degree of manipulation in the educational context. Which of these factors is easiest to alter from an educational or policy perspective is debatable; the likelihood is that all are important along with additional factors in effecting increases in overall SAT scores.

Perhaps of more constructive interest for front-line educators are all academically-related variables discussed above rather than the more demographically-related school percentage of limited-English proficient students (Verbal model), school percentage of white teachers (Math model), and district percentage of minority students (Verbal model), especially given the generally predominant size of the TAAS passing rate effect in all three models. All three demographic variables rather than the 11 academic variables are less readily alterable in the educational context. However, all three demographic variables can be considered indicative of educational access opportunities. For instance, higher percentages of minority students usually attend higher teacher minority (lower white teacher percentage) schools in Texas (TEA, 1994), which contaminates interpretations about the quality of instruction provided or overall expectations in these schools. Short of additional information about the professional working conditions for teachers in these schools, this result supports, perhaps, the observation that disadvantaged student groups tend to receive instruction historically from the least rewarded teachers and least supported schools (e.g., Darling-Hammond, 1990, 1992). As such, existence of these effects may suggest areas for targeting broader-based educational policies and interventions.

Predictors of Average Score Differentiation

Verbal and Total score conditional three-level models included the same three student predictors--GPA, college English placement plans, and number of advanced courses completed--with the score differentiation effect of college English placement plans the largest of the three predictors in both models. Number of advanced courses was included in all three models--Verbal, Math, and Total. High school rank appeared instead of GPA in the Math model and the associated score differentiating effect was about twice that of the differentiating effect associated with advanced courses. Notably, all school predictors were educational rather than demographic, conceptually speaking.

Verbal/Math/Total score differentiation related to advanced coursework. Over the three models, the score advantage positively related to number of advanced courses completed, an indicator of individual press for achievement, lessened as significant predictors representing educational access opportunities went up including the school percentage of students in advanced courses, school number of types of advanced (AP and non-AP) courses completed (Math model only), and types of AP courses completed (Total model only). Some similar significant predictors were noted for this same differentiating effect at the district level, with the effect diminishing as district percentages increased for: (a) non-AP advanced course(s) completed (all models), (b) AP course(s) completed (Math and Total models), and (c) minority students (Verbal model only). One potential explanation for such attenuation with these predictors is that the individual achievement advantage associated with advanced coursework becomes less discernible in SAT scores within an atmosphere of commonly available educational opportunities or with commonly unavailable opportunities, as may be the case historically as minority student percentages increase.

Otherwise, significant district press for achievement predictors--percentages passing all TAAS Exit exams (Math and Total models) and the number of types of AP subject exams taken (Verbal and Math models)--were positively linked with enhancement of the advanced courses score advantage. Both predictors are indicative of the extent to which there are general expectations within districts for students to excel on and attempt other academic exams.

Math score differentiation related to rank. The average math score advantage associated with higher rank in class, an indicator of individual press for achievement, lessened somewhat as the district percentage of students completing non-AP advanced course(s) rose and grew slightly as the district number of types of AP exams taken went up. If the first significant predictor represents educational access, its corresponding "leveling" effect on rank is consistent with interpretations made for its similar effect on the average advanced courses score advantage, also an individual press for achievement variable. Perhaps under such conditions rank becomes less reliable in differentiating among students on average and/or a student's own self-reported rank in class is harder to estimate. That the SAT Math score advantage linked with rank increases as a significant district press for achievement predictor, types of AP exams taken, goes up is line with the interpretation of relatively higher expectations.

<u>Verbal/Total score differentiation related to GPA.</u> In the Verbal and Total models, the average score advantage linked with GPA, similar to rank as an indicator of individual press for achievement, also shrank as the district percentage of students completing non-AP advanced course(s) increased (both models) and as the school percentage passing all TAAS Exit exams went up (Total model only) but enlarged as the district number of types of AP exams taken rose (Verbal model only). Similar to the effect of the first and third predictors on rank above, similar interpretations of the relationships apply and also apply to TAAS passing percentage as a school press for achievement predictor. Furthermore, GPA may become less reliable on average as a predictor of Verbal and Total scores because in association with the one school and two district predictors, grading practices may also vary--either grade inflation or deflation--perhaps also making GPA a less reliable indicator of students' college academic potential. Thus, if district/school X's "B" is district/school Y's "A," or vice versa, any effects on GPA warrant further study, given the extent to which grading practices can greatly influence students' perceptions about their abilities and, consequently, decisions they make about their academic and work futures. Of course, the same applies when any substantial unreliability in SAT score is apparent.

<u>Verbal/Total score differentiation related to college English advanced placement.</u> Plans for college English advanced placement may indicate the extent to which students at the individual level are able or willing to bring resources to bear to achieve this educational advantage (i.e., planning for educational access opportunity). Academic press for achievement variables--school number of types of AP subject exams taken (Verbal model) and school percentage taking AP exam(s) (Total model)--were positively associated with an even greater score advantage for students indicating plans for college English advanced placement vs. those with no plans. Perhaps this occurs because the relatively greater expectations within schools, as indicated by the percentage of students taking challenging exams or in the diversity of challenging exams taken, help support students in planning for access to further educational opportunity. Otherwise, one district educational access predictor, percentage of non-AP advanced courses completed (Total model), was also positively associated with even further score advantage for students with college English placement plans, perhaps because in a context of greater educational access, students receive relatively more information about how they can realistically access further educational opportunities.

<u>1994 Two-Level vs. Three-Level Models</u>

Compared to previous two-level HLM models of Texas SAT scores, three-level HLM allowed a more precise specification of predictors attributable not only to the school, but also to the district level. For educators trying to understand relationships among educational and demographic factors associated with various levels of the organization, three-level HLM appears to provide a more powerful elucidation. Not only were three-level HLM models generally more parsimonious than previous two-level models for 1994 scores (cf. Hargrove et al., 1996), but the explanatory power of these models was also generally improved and in a few instances equivalent or only somewhat worse than previous models for corresponding scores and within-school predictors. Some between-school predictors in the previous two-level models for 1994 were either added as district variables or replaced with an increased number of AP exam diversity, participation, and performance variables, as well as advanced coursework diversity and participation variables, which probably added to the explanatory power of this year's HLM models. For example, the proportion of parameter variance explained by three-level modeling for each 1994 SAT Verbal, Math, and Total score Level-2 model intercept was somewhat higher (ranging from 74% in Verbal, 78% in Math, and 76% in Total models) than for the same intercept in the previous two-level models of the 1994 scores (ranging from 72% in the Verbal model to 75% in Math and Total models).

Limitations to the above three-level vs. two-level model comparisons are potentially critical differences between this study and the previous studies. For instance, data for a lesser number of schools (N=169) and students (N=32,877)--those in non-rural or metropolitan districts with two or more high schools--were included in this study's analyses vs. those in last year's study (292 schools and 52,707 students). In this year's study, the Bryk et al. (1996) HLM™ computer program was used, whereas last year's study used the Bryk et al. (1994) HLM™$^{2}_{3}$ computer program. In addition, this year's modeling employed parallel specification of predictors for all dependent variables within a given level, as is generally recommended in Bryk and Raudenbush (1992), whereas last year's modeling did not.

Conclusions

Overall, findings such as the preponderance of educational access and school/community press for achievement variables related to school (and now also district) average SAT scores were consistent with those in the previous Hargrove et al. (1995, 1996) and Hargrove and Mellor (1994) studies. Whether these were the relatively more alterable educational variables or the less alterable demographic ones, school- and district-level variables typically associated positively with educational advantage (e.g., higher school TAAS passing percentages, school AP examinee percentages scoring 3-5, school percentages taking AP exams, and district percentags completing non-AP advanced courses) were also linked with higher average SAT scores, while variables indicating educational disadvantage (e.g., higher school limited-English proficient student percentages and district minority student percentages) were linked with lower average scores.

Even more discernible in this year's than in last year's study were patterns among the district, school, and student variables considered representative of the educational access and press for achievement constructs. For example, SAT score advantages linked with greater individual educational access (e.g., having college English placement/credit plans) tended to further increase in districts with greater access to educational opportunities (e.g., higher percentages completing non-AP advanced courses in the SAT Total model only) and in schools with a relatively greater press for achievement via AP exam-taking diversity (Verbal model) and AP exam participation (Total model). In contrast, average score advantages associated with individual press for achievement (e.g., higher GPA, rank, and advanced coursework) tended to diminish somewhat in districts with generally greater access to educational opportunity (e.g., higher percentages completing non-AP advanced courses). In other instances, average score advantages accompanying individual achievement press (e.g., GPA and more advanced coursework in Verbal and Math models) were strengthened when the district achievement press via exam diversity (e.g., types of AP exams taken) went up.

Given that most of the variables above are alterable characteristics of the educational context, perhaps attention to these can help channel policy discussions and actions in a more productive direction regarding the reporting and use of SAT score results in the state's accountability system. For instance, one suggestion is the disaggregated reporting of district/school average scores by groups of districts/schools with contextually similar educational opportunities and prior/current press for educational achievement rather than by groups of districts/schools similar on relatively "unalterable" demographic variables alone. If school- or district-level SAT performance is exceptionally out of range in either a positive or negative direction, focus on some of the predominant district, school, and student characteristics from this study's analyses may be useful in diagnosing reasons for performance success or failure. Such focus is in accordance with Goldstein's (1991) recommendation that HLM, per se, not be used to rank institutions on performance but rather be put to diagnostic use in the preliminary phase of a more complete school (and district) performance evaluation.

Because the purpose of this study was to examine the viability and explanatory power of three-level models for SAT scores, results and interpretations are conditional with respect to the smaller number of students and schools and, hence, districts analyzed compared to last year's two-level modeling of a larger set of these data. Most interpretations raise questions needing more in-depth study and confirmation. In addition, the relatively large portions of unexplained Level-1 variance in the models (see Table 3) suggest perhaps the need to include other student predictors that may increase the amount of variance explained at Level-1, even if these are predictors without sufficient amounts of variance for modeling at Levels 2 and 3, thereby requiring associated error terms to be set to zero at these levels. Thus, conclusions regarding the extent to which variables in the models may accurately "predict" test scores for any individual, school, or district are stronger than in previous work but still somewhat tentative. Further work remains to be done in mapping out the most important of the complex multivariate relationships existing among district, school, and student variables/constructs related to SAT scores. This includes continuing to study the consistency of the Verbal, Math, and Total score models from one year to the next, especially with the advent of the SAT-I beginning with the Class of 1995 and the recentering of Verbal, Math, and Total scores initiated with the Class of 1996. In sum, continued multilevel (HLM) conceptualization of the Texas SAT data can help enrich understanding about how student, school, and district characteristics relate to SAT score diversity. This, in turn, can help improve the quality and fairness of comparisons made among schools and districts and, ultimately, similar judgments made about students within schools and districts.

References

Arnold, C. L. (1992). Methods, plainly speaking: An introduction to hierarchical linear models. *Measurement and Evaluation in Counseling and Development, 25*, 58-90.

Arnold, C. L. (1993, April). *Results from the 1990 NAEP Data Reporting Program*. Paper prepared for the American Educational Research Association Annual Meeting, Atlanta, Georgia.

Bryk, A. S., & Hermanson, K. L. (1993). Educational indicator systems: Observations on their structure, interpretation, and use. In L. Darling-Hammond (Ed.), *Review of Research in Education, Vol. 19* (pp. 451-484). Washington, DC: American Educational Research Association.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Bryk, A. S., & Raudenbush, S. W., & Congdon, R. T., Jr. (1994). HLM™2_3. *Hierarchical Linear Modeling with the HLM/2L and HLM/3L Programs.* Chicago: Scientific Software, Inc.

Bryk, A. S., & Raudenbush, S. W., & Congdon, R. T., Jr. (1996). *HLM™: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software International.

Bryk, A. S., & Raudenbush, S. W., Seltzer, M., & Congdon, R. T., Jr. (1989). *An introduction to HLM™: Computer program and user's guide*. Chicago: Scientific Software, Inc.

College Entrance Examination Board. (1994a). *College Bound Seniors: 1994 Profile of SAT and Achievement Test Takers. National Report.* New York: Author.

College Entrance Examination Board. (1994b). *College Bound Seniors: 1994 Profile of SAT and Achievement Test Takers. Texas Report.* New York: Author.

College Entrance Examination Board. (1994c). *1994 AP Texas and national summary reports.* New York: Author.

Cornett, L. M., & Gaines, G. F. (1993, February). Incentive programs: A focus on program evaluation. *Career Ladder Clearinghouse*. Atlanta: Southern Regional Education Board.

Cornett, L. M., & Gaines, G. F. (1994, April). Reflecting on ten years of incentive programs. The 1993 SREB Career Ladder Clearinghouse Survey. *Career Ladder Clearinghouse*. Atlanta: Southern Regional Education Board.

Darling-Hammond, L. (1990). Teacher quality and equality. In J. Goodlad & P. Keating (Eds.), *Access to knowledge: An agenda for our nation's schools.* New York: College Entrance Examination Board.

Darling-Hammond, L. (1992). Educational indicators and enlightened policy. *Educational Policy*, (3), 235-265.

Goldstein, H. (1991). Commentary. Better ways to compare schools? *Journal of Educational Statistics, 16*, 89-91.

Everson, H., Millsap, R. E., & Diones, R. (1995, April). *Exploring group differences in SAT performance using structural equation modeling with latent means.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Fetler, M. E. (1991). Pitfalls of using SAT results to compare schools. *American Educational Research Journal, 28*, 481-491.

Hargrove, L. L. (1992, November). *A retrospective examination of characteristics of Texas schools considered for state monetary awards based on aggregate academic performance gain.* Paper presented at the Annual Meeting of the American Evaluation Association, Seattle.

Hargrove, L. L., & Mellor, L. T. (1993, November). *Three years of school incentive monetary awards in Texas: Characteristics of eligible schools.* Paper presented at the Annual Meeting of the American Evaluation Association, Dallas.

Hargrove, L. L., & Mellor, L. T. (1994, April). *An HLM exploration of between-school effects related to within-school SAT score differences in Texas: Accountability implications.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, Louisiana.

Hargrove, L. L., Mellor, L. T., & Mao, M. X. (1995, April). *Further HLM exploration of between- and within-school effects related to SAT scores in Texas: Policy and reporting implications.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.

Hargrove, L. L., Mao, M. X., & Barkanic, G. (1996, April). *HLM Modeling of Coursework, AP and Other Academic Context Variables Related to SAT scores in Texas.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.

Herr, N. E. (1993). The relationship between Advanced Placement and Honors science courses. *School and Science Mathematics, 93,* 183-187.

Holland, P. W., & Wainer, H. (1990). Sources of uncertainty often ignored in adjusting state mean SAT scores for differential participation rates: The rules of the game. *Applied Measurement in Education, 3,* 167-184.

Oakes, J. (1989). What educational indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis, 11,* 181-199.

Page, E. B., & Feifs, H. (1985). SAT scores and American states: Seeking for useful meaning. *Journal of Educational Measurement, 22,* 305-312.

Smith, M. S. (1988). Educational indicators. *Phi Delta Kappan, 69,* 487-491.

Taube, K. T., & Linden, K. W. (1989). State mean SAT score as a function of participation rate and other educational and demographic variables. *Applied Measurement in Education, 2,* 143-159.

Texas Education Agency. (1990). *Report of results of college admissions testing in Texas for 1988/89 graduating seniors* (TEA Publication No. GE0 543 04). Austin, TX: Author.

Texas Education Agency. (1994). Teacher diversity and recruitment. *Policy Research Report, Report Number 4.* Austin, TX: Author.

Texas Education Agency. (1995a). *Accountability manual: The 1995 accountability rating system for Texas public schools and school districts and blueprint for the 1996-2000 accountability systems.* Austin, TX: Author.

Texas Education Agency. (1995b). *Public Education Information Management System (PEIMS) Data Standards 1994-95* (Publication No. T12 082 01). Austin, TX: Author.

Texas Education Agency. (1995c). *Reporting Texas Advanced Placement Examination performance: Promoting a head start on the transition to college* (TEA Publication No. RE6 601 03). Austin, TX: Author.

Texas Education Agency. (1995d). *Snapshot '94: 1993-94 school district profiles* (TEA Publication No. GE5 602 01). Austin, TX: Author.

Texas Education Agency (1996). *Results of college admissions testing in Texas for 1993-94 graduating seniors* (TEA Publication No. GE6 601 05). Austin, TX: Author.

Wachter, K. (1989). Statistical adjustment: Comment on Howard Wainer's "Eelworms, Bulletholes, & Geraldine Ferraro." *Journal of Educational Statistics, 14,* 183-186.

Wainer, H. (1986). Five pitfalls encountered while trying to compare states on their SAT scores. *Journal of Educational Measurement, 23,* 69-81.

Wainer, H. (1989a). Eelworms, bulletholes and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics, 14,* 121-140.

Wainer, H. (1989b). Responsum. *Journal Educational Statistics, 14,* 187-200.

Wainer, H. (1990). Adjusting NAEP for self-selection: A useful place for "Wall Chart" technology. *Journal of Educational Statistics, 15,* 1-7.

Wainer, H., Holland, P. W., Swinton, S., & Wang, M. (1985). On "State Education Statistics." *Journal of Educational Statistics, 10,* 293-325.

Willms, J. D., & Kerckhoff, A. C. (1995). The challenge of developing new educational indicators. *Educational Evaluation and Policy Analysis, 17,* 113-131.

Table 1: Unweighted Means and Standard Deviations for Unstandardized Student-, School-, and District-Level Variables Explored in Three-Level HLM Modeling of 1994 Texas SAT Verbal, Math, and Total Scores

Variable	N	Mean	Std. Dev.	Min.	Max.
Within-Schools Within Districts (Student-Level):					
1994 SAT Total Score*	32,877	915.37	209.56	400	1600
1994 SAT Math Score*	32,877	491.26	118.10	200	800
1994 SAT Verbal Score*	32,877	424.12	109.13	200	800
High School GPA*	32,877	4.02	1.80	1	12
High School Rank*	32,877	2.73	1.19	1	6
English Placement Plans*	32,877	0.33	0.47	0	1
# Non-AP Advanced Courses Completed*	32,877	1.46	1.75	0	10
Between-Schools Within Districts (School-Level):					
Student Enrollment Number	169	1992.18	647.74	331	3997
% Economically Disadvantaged Students	169	25.95	20.87	0	95.7
% Minority Students	169	52.87	27.54	6.3	98.7
% Limited English-Proficient Students*	169	6.68	8.81	0	56.9
% Passing TAAS Math Exit-Level Exam	169	59.18	15.10	23.2	93.5
% Passing TAAS Reading Exit-Level Exam	169	78.28	11.99	47.3	98.2
% Passing All TAAS Exit-Level ExamsTaken*	169	54.52	15.88	19.2	91.8
% Expenditures for Vocational Instruction	169	11.49	5.66	0	34.5
% 1993 Dropouts	169	3.43	2.56	0	13
% Graduates Planning College	169	75.40	19.09	1.3	100
% Graduates Taking >= 1 SAT/ACT	169	68.98	15.48	18.4	99.8
% Graduates with Advanced Seals on Diploma	169	41.72	17.69	0	100
% Students in Advanced Courses*	169	14.39	7.39	2.2	48.7
# Students per Teacher (Student/Teacher Ratio)	169	17.17	1.82	11.7	21.1
% Teachers White (Non-minority)*	169	79.55	17.68	28.3	99.2
% Gr. 11-12 Enrollees Taking >= 1 AP Exam*	169	6.81	5.32	0.2	27.9
# Types AP Subject Exams Taken*	169	9.06	5.32	1.0	26
% Gr. 11-12 AP Exam Scores 3-5	169	45.65	16.98	0	100
% Gr. 11-12 AP Exam Takers with >= 1 Score 3-5*	169	67.45	21.33	0	100
% Students Completing >= 1 AP Advanced Course	169	4.17	5.11	0	28.5
% Students Completing >= 1 non-AP Adv. Course	169	12.72	6.78	0	51.5
# Types AP Subject Advanced Courses Completed*	169	3.43	4.44	0	21
# Types Non-AP Advanced Courses Completed	169	9.74	4.48	0	29
# Types AP/Non-AP Advanced Courses Completed*	169	13.17	5.69	0	32
Between Districts (District-Level):					
Student Enrollment Number	48	37380.44	32816.76	6454	200445
% Economically Disadvantaged Students	48	39.05	20.86	8.7	93.5
% Minority Students*	48	49.92	25.28	11.7	97.2
% Limited English-Proficient Students	48	10.66	10.54	0.8	44
% Passing TAAS Math Exit-Level Exam	48	60.60	11.32	34.1	83.5
% Passing TAAS Reading Exit-Level Exam	48	79.21	8.85	58.1	94.4
Operating Cost per Pupil	48	4257.29	385.52	3443	5283
Property Value per Student	48	169933.38	74502.25	44428.6	392915.9
% Expenditures for Vocational Instruction	48	3.50	0.91	1.6	5.7
(Non-)Central/North Regional Location	48	0.60	0.49	0	1
% Passing All TAAS Exit-Level ExamsTaken*	48	55.92	11.92	30.2	80.8
% Students Mobile	48	22.02	5.05	13.8	36.5
% 1993 Dropouts	48	2.58	1.44	0.3	6.7
% Graduates Planning College	48	75.33	15.18	40.9	99.0
% Graduates Taking >= 1 SAT/ACT	48	66.94	12.93	31	90.5
% Graduates with Advanced Seals on Diploma	48	40.86	12.09	3.6	71.1
# Math Classes	48	460.42	325.85	0	1744
# Science Classes	48	464.94	317.02	0	1692
% Students in Advanced Courses	48	12.93	4.32	6.6	25.1
# Students per Teacher (Student/Teacher Ratio)	48	16.54	1.07	13.6	19
Average Teacher Salary	48	29649.46	1576.21	26333	34805
Average Teacher Experience	48	11.51	1.12	8.8	14.5
% Teachers with Advanced Degrees	48	30.99	8.79	14.6	57.7
% Teachers Minority (Non-white)	48	19.86	19.14	1.6	73
% Teachers White (Non-minority)	48	80.14	19.14	27	98.5
% Gr. 11-12 Enrollees Taking >= 1 AP Exam	48	6.42	4.54	0.9	22.1
# Types AP Subject Exams Taken*	48	14.48	6.16	3	26
% Gr. 11-12 AP Exam Scores 3-5	48	46.44	10.99	23.5	80.5
% Gr. 11-12 AP Exam Takers with >= 1 Score 3-5	48	70.10	14.51	29.2	91.7
% Students Completing >= 1 AP Advanced Course*	48	3.97	4.18	0	14.7
% Students Completing >= 1 non-AP Adv. Course*	48	11.64	3.63	5.1	24.2
# Types AP Subject Advanced Courses Completed	48	5.35	5.65	0	21
# Types Non-AP Advanced Courses Completed	48	14.63	6.16	6	38
# Types AP/Non-AP Advanced Courses Completed	48	19.98	7.42	6	51

**Variable used in at least one of three HLM models described in this paper.*

Table 2: Overview of Three-Level HLM Exploratory Modeling of 1994 Texas SAT Scores

Within-School Model Parameters	Between-School Effects	Between-District Effects	Models: Verbal	Math	Total
Intercept 1 (mean score)	*On Intercept 2*	*On Intercept 3*	*X***	*X***	*X***
		% Minority Students	X(-)**		
		% Passing All TAAS Exit Exams Taken		X**	X**
		# Types AP Subject Exams Taken	X*	X	
		% Completing >= 1 AP Adv. Course		X	X
		% Completing >= 1 Non-AP Adv. Course	X(-)	X	X
	% Limited-English Proficient Students		X(-)**		
	% Teachers White (Non-minority)			X**	
	% Passing All TAAS Exit Exams Taken		X**	X**	X**
	% Students in Advanced Courses		X	X	X
	# Types AP Subject Exams Taken		X**		
	% Gr. 11-12 AP Takers with >= 1 Score 3-5		X**	X**	X**
	# Types AP/Non-AP Adv. Courses Compl.			X**	X**
	# Types AP Subject Courses Completed		X(-)**		X(-)**
	% Gr. 11-12 Enrls. Taking>=1 AP Exam		X**	X**	X**
GPA [Rank] Slope Coef.	*On Intercept 2*	*On Intercept 3*	*X(-)***	*[X(-)**]*	*X(-)***
		% Minority Students	X(-)		
		% Passing All TAAS Exit Exams Taken		[X(-)]	X
		# Types AP Subject Exams Taken	X*	[X*]	
		% Completing >= 1 AP Adv. Course		[X(-)]	X
		% Completing >= 1 Non-AP Adv. Course	X(-)*	[X(-)**]	X(-)*
	% Limited-English Proficient Students		X**		
	% Teachers White (Non-minority)			[X(-)]	
	% Passing All TAAS Exit Exams Taken		X(-)	[X(-)]	X(-)*
	% Students in Advanced Courses		X(-)	[X(-)]	X(-)
	# Types AP Subject Exams Taken		X(-)		
	% Gr. 11-12 AP Takers with >= 1 Score 3-5		X(-)	[X(-)]	X(-)
	# Types AP/Non-AP Adv. Courses Compl.			[X(-)]	X(-)
	# Types AP Subject Courses Completed		X		X
	% Gr. 11-12 Enrls. Taking>=1 AP Exam		X(-)	[X(-)]	X(-)
Eng. Placemnt. Slope Coef.	*On Intercept 2*	*On Intercept 3*	*X***		*X***
		% Minority Students	X(-)		
		% Passing All TAAS Exit Exams Taken			X
		# Types AP Subject Exams Taken	X		
		% Completing >= 1 AP Adv. Course			X(-)
		% Completing >= 1 Non-AP Adv. Course	X		X**
	% Limited-English Proficient Students		X		
	% Teachers White (Non-minority)				
	% Passing All TAAS Exit Exams Taken		X(-)		X(-)
	% Students in Advanced Courses		X		X
	# Types AP Subject Exams Taken		X*		
	% Gr. 11-12 AP Takers with >= 1 Score 3-5		X		X
	# Types AP/Non-AP Adv. Courses Compl.				X
	# Types AP Subject Courses Completed		X(-)		X(-)
	% Gr. 11-12 Enrls. Taking>=1 AP Exam		X		X*
#Adv. Courses. Slope Coef.	*On Intercept 2*	*On Intercept 3*	*X***	*X***	*X***
		% Minority Students	X(-)*		
		% Passing All TAAS Exit Exams Taken		X**	X**
		# Types AP Subject Exams Taken	X*	X**	
		% Completing >= 1 AP Adv. Course		X(-)**	X(-)**
		% Completing >= 1 Non-AP Adv. Course	X(-)**	X(-)**	X(-)**
	% Limited-English Proficient Students		X(-)		
	% Teachers White (Non-minority)			X	
	% Passing All TAAS Exit Exams Taken		X(-)	X(-)	X
	% Students in Advanced Courses		X(-)**	X(-)**	X(-)**
	# Types AP Subject Exams Taken		X		
	% Gr. 11-12 AP Takers with >= 1 Score 3-5		X	X	X
	# Types AP/Non-AP Adv. Courses Compl.			X(-)*	X(-)
	# Types AP Subject Courses Completed		X(-)		X(-)*
	% Gr. 11-12 Enrls. Taking>=1 AP Exam		X	X	X

Note. Gamma coefficient is statistically significant using two-tailed univariate test probabilities for t when single asterisk (*$p \leq .05$) or double asterisk (**$p \leq .01$) is shown. Minus signs in parentheses indicate direction of effect given coded values for a variable--for example, GPA and high school rank are coded such that lower values mean higher GPA or rank. Brackets indicate effects are for the rank slope intercept 2 and intercept 3 coefficients.

Table 3: Proportion of Parameter Variance and Total Variance Explained for Three Conditional Three-Level HLM Models of 1994 Texas SAT Scores

Model & Parameters	Unconditional (Null) Model Reliability	Conditional Model Reliability	Residual Parameter Var. (Tau)	d.f.	Prob. of Tau=0	Prop. of Parameter Var. Exp.	Prop. of Total Var. Explained
Verbal Score Model							
Level-1:							
(Dev.=398483; 4 param. est.)							
Error	--	--	6695.50	--	--	.3671	--
Level-2:							
(Dev.=383907; 25 param. est.)							
Intercept (mean Total)	.954	.838	245.68	114	.000	.7631	.7280
GPA Slope	.459	.343	9.62	114	.000	.4145	.1903
English Placement Slope	.552	.476	213.72	114	.000	.2853	.1575
# Advanced Courses Slope	.331	.254	7.55	114	.000	.3491	.1156
Level-3:							
(Dev.=383703; 53 param. est)							
Intercept (mean Total)	1.000	1.000	270.87	44	.000	.7477	--
GPA Slope	1.000	1.000	7.24	44	.000	.4548	--
English Placement Slope	1.000	1.000	63.45	44	.004	.5504	--
# Advanced Courses Slope	1.000	1.000	10.81	44	.000	.5084	--
*(Dev.=383622; 65 param. est.)**							
Math Score Model							
Level-1:							
(Dev.=403506; 4 param. est.)							
Error	--	--	7059.25	--	--	.4276	--
Level-2:							
(Dev.=385550; 16 param. est.)							
Intercept (mean Math)	.948	.814	215.63	115	.000	.7789	.7384
Rank Slope	.556	.404	27.79	115	.000	.4822	.2681
# Advanced Courses Slope	.474	.374	14.56	115	.000	.3851	.1825
Level-3:							
(Dev.=385365; 34 param. est.)							
Intercept (mean Math)	1.000	1.000	280.09	43	.000	.7868	--
Rank Slope	1.000	.998	1.76	43	.148	.9251	--
# Advanced Courses Slope	1.000	1.000	11.66	43	.000	.7331	--
*(Dev.=385248; 46 param. est.)**							
Total Score Model							
Level-1:							
(Dev.=440711; 4 param. est.)							
Error	--	--	20759.95	--	--	.4565	--
Level-2:							
(Dev.=421230; 25 param. est.)							
Intercept (mean Total)	.961	.863	937.20	115	.000	.7608	.7311
GPA Slope	.503	.410	40.82	115	.000	.3410	.1715
English Placement Slope	.556	.511	774.92	115	.000	.1793	.0997
# Advanced Courses Slope	.473	.364	42.67	115	.000	.4138	.1957
Level-3:							
(Dev.=421035; 49 param. est.)							
Intercept (mean Total)	1.000	1.000	1019.12	44	.000	.7722	--
GPA Slope	1.000	1.000	30.02	44	.000	.2979	--
English Placement Slope	1.000	1.000	89.83	44	.121	.7164	--
# Advanced Courses Slope	1.000	1.000	59.36	44	.000	.4552	--
*(Dev.=420936; 61 param. est.)**							

Note. Deviances for the unconditional models are shown by level; asterisked deviances are for the fully conditional model across all levels of the corresponding model. Variance explained at the third level for each fully conditional model was compared to a model with unconditional variance at the third level only (assuming fully conditional models at Levels 1 and 2). Because Level 3 reliabilities calculated by the HLM program were generally 1.0, total variance explained was not calculated for Level 3.

Table 4: Effects of School and District Characteristics on Student-Level Predictors of Texas 1994 SAT Verbal Scores

School and District Effects	Gamma Coefficient	Standard Error	*t* Value
For Intercept 1 (average SAT Total)			
For Intercept 2 /Intercept 3	419.33	2.81	149.12**
District % Minority Students	-27.60	3.34	-8.26**
District # Types AP Subject Exams Taken	6.75	3.07	2.20*
District % Completing >= 1 Non-AP Adv. Course	-0.47	3.55	-0.13
For School % Limited-English Proficient Students /Intercept 3	-8.54	2.39	-3.57**
For School % Passing All TAAS Exit Exams Taken /Intercept 3	9.10	3.17	2.88**
For School % Students in Advanced Courses /Intercept 3	3.95	2.85	1.39
For School # Types AP Subject Exams Taken /Intercept 3	12.17	2.97	4.10**
For School % Gr. 11-12 AP Takers with >= 1 score 3-5 /Intercept 3	8.79	2.00	4.40**
For School # Types AP Subject Courses Completed /Intercept 3	-10.69	4.28	-2.50**
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	9.61	2.72	3.53**
For GPA Slope Coefficient			
For Intercept 2 /Intercept 3	-16.37	0.61	-26.67**
District % Minority Students	-0.78	0.72	-1.09
District # Types AP Subject Exams Taken	1.34	0.68	1.98*
District % Completing >= 1 Non-AP Adv. Course	-1.89	0.77	-2.45*
For School % Limited-English Proficient Students /Intercept 3	2.82	0.81	3.50**
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-0.80	1.01	-0.80
For School % Students in Advanced Courses /Intercept 3	-0.38	0.89	-0.43
For School # Types AP Subject Exams Taken /Intercept 3	-0.58	0.89	-0.65
For School % Gr. 11-12 AP Takers with >= 1 score 3-5 /Intercept 3	-0.13	0.66	-0.20
For School # Types AP Subject Courses Completed /Intercept 3	1.19	1.26	0.94
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	-0.66	0.82	-0.81
For English Placement Slope Coefficient			
For Intercept 2 /Intercept 3	47.75	2.16	22.07**
District % Minority Students	-3.48	2.52	-1.38
District # Types AP Subject Exams Taken	3.89	2.35	1.66
District % Completing >= 1 Non-AP Adv. Course	4.92	2.72	1.81
For School % Limited-English Proficient Students /Intercept 3	0.75	3.25	0.23
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-3.40	4.05	-0.84
For School % Students in Advanced Courses /Intercept 3	1.03	3.52	0.29
For School # Types AP Subject Exams Taken /Intercept 3	7.82	3.65	2.14*
For School % Gr. 11-12 AP Takers with >= 1 score 3-5 /Intercept 3	3.46	2.61	1.33
For School # Types AP Subject Courses Completed /Intercept 3	-8.18	4.94	-1.66
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	4.58	3.30	1.39
For # Advanced Courses Slope Coefficient			
For Intercept 2 /Intercept 3	15.20	0.69	22.08**
District % Minority Students	-1.82	0.82	-2.22*
District # Types AP Subject Exams Taken	1.79	0.76	2.35*
District % Completing >= 1 Non-AP Adv. Course	-3.65	0.83	-4.37**
For School % Limited-English Proficient Students /Intercept 3	-0.92	0.86	-1.08
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-0.99	1.13	-0.88
For School % Students in Advanced Courses /Intercept 3	-3.68	1.02	-3.60**
For School # Types AP Subject Exams Taken /Intercept 3	0.94	0.99	0.95
For School % Gr. 11-12 AP Takers with >= 1 score 3-5 /Intercept 3	1.29	0.77	1.68
For School # Types AP Subject Courses Completed /Intercept 3	-1.87	1.30	-1.44
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	1.20	0.85	1.42

Two-tailed univariate test probabilities:
* $p \leq .05$, ** $p \leq .01$

Table 5: Effects of School and District Characteristics on Student-Level Predictors of Texas 1994 SAT Math Scores

School and District Effects	Gamma Coefficient	Standard Error	*t* Value
For Intercept 1 (average SAT Math)			
For Intercept 2 /Intercept 3	485.80	2.81	173.19**
District % Passing All TAAS Exit Exams Taken	28.17	3.80	7.41**
District # Types AP Subject Exams Taken	3.16	3.17	1.00
District % Students Completing >=1 AP Advanced Course	3.95	3.50	1.13
District % Students Completing >=1 non-AP Advanced Course	2.04	3.68	0.55
For School % Teachers White (Non-minority) /Intercept 3	9.77	3.68	2.66**
For School % Passing All TAAS Exit Exams Taken /Intercept 3	9.33	2.71	3.44**
For School % Students in Advanced Courses /Intercept 3	5.36	2.72	1.97
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	8.02	1.93	4.16**
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	10.71	2.34	4.58**
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	8.73	2.34	3.73**
For Rank Slope Coefficient			
For Intercept 2 /Intercept 3	-40.84	0.74	-55.27**
District % Passing All TAAS Exit Exams Taken	-1.04	0.90	-1.15
District # Types AP Subject Exams Taken	1.87	0.78	2.40*
District % Students Completing >=1 AP Advanced Course	-0.99	0.83	-1.20
District % Students Completing >=1 non-AP Advanced Course	-4.98	0.96	-5.19**
For School % Teachers White (Non-minority) /Intercept 3	-2.48	1.94	-1.28
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-1.87	1.38	-1.35
For School % Students in Advanced Courses /Intercept 3	-1.08	1.37	-0.79
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	-0.99	1.02	-0.97
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	-1.89	1.17	-1.61
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	-0.97	1.12	-0.87
For # Advanced Courses Slope Coefficient			
For Intercept 2 /Intercept 3	21.10	0.75	28.15**
District % Passing All TAAS Exit Exams Taken	3.44	0.98	3.49**
District # Types AP Subject Exams Taken	3.77	0.83	4.54**
District % Students Completing >=1 AP Advanced Course	-5.40	0.88	-6.12**
District % Students Completing >=1 non-AP Advanced Course	-5.73	0.96	-5.97**
For School % Teachers White (Non-minority) /Intercept 3	2.89	1.54	1.88
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-0.22	1.15	-0.19
For School % Students in Advanced Courses /Intercept 3	-3.84	1.14	-3.37**
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	0.58	0.87	0.67
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	-2.48	0.97	-2.56*
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	0.56	0.89	0.63

Two-tailed univariate test probabilities:
* $p \leq .05$, ** $p \leq .01$

Table 6: Effects of School and District Characteristics on Student-Level Predictors of Texas 1994 SAT Total Scores

School and District Effects	Gamma Coefficient	Standard Error	*t* Value
For Intercept 1 (average SAT Total)			
For Intercept 2 /Intercept 3	905.00	5.43	166.72**
District % Passing All TAAS Exit Exams Taken	49.82	7.28	6.84**
District % Students Completing >=1 AP Advanced Course	8.18	6.47	1.27
District % Students Completing >=1 non-AP Advanced Course	7.77	6.68	1.16
For School % Passing All TAAS Exit Exams Taken /Intercept 3	25.70	5.26	4.88**
For School % Students in Advanced Courses /Intercept 3	6.21	5.43	1.14
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	16.05	3.77	4.26**
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	23.75	5.02	4.73**
For School # Types AP Subject Courses Completed /Intercept 3	-22.08	8.56	-2.58**
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	22.12	4.78	4.62**
For GPA Slope Coefficient			
For Intercept 2 /Intercept 3	-38.92	1.17	-33.32**
District % Passing All TAAS Exit Exams Taken	0.23	1.55	0.15
District % Students Completing >=1 AP Advanced Course	1.31	1.38	0.95
District % Students Completing >=1 non-AP Advanced Course	-2.90	1.46	-1.99*
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-3.68	1.61	-2.29**
For School % Students in Advanced Courses /Intercept 3	-0.33	1.66	-0.20
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	-0.60	1.22	-0.49
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	-0.91	1.48	-0.62
For School # Types AP Subject Courses Completed /Intercept 3	2.75	2.44	1.12
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	-1.78	1.48	-1.20
For English Placement Slope Coefficient			
For Intercept 2 /Intercept 3	66.13	3.41	19.38**
District % Passing All TAAS Exit Exams Taken	0.90	4.45	0.20
District % Students Completing >=1 AP Advanced Course	-2.13	4.06	-0.53
District % Students Completing >=1 non-AP Advanced Course	15.60	4.37	3.57**
For School % Passing All TAAS Exit Exams Taken /Intercept 3	-6.70	6.32	-1.06
For School % Students in Advanced Courses /Intercept 3	1.78	6.37	0.28
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	4.41	4.71	0.94
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	10.06	5.75	1.75
For School # Types AP Subject Courses Completed /Intercept 3	-15.24	8.89	-1.71
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	10.96	5.61	1.95*
For # Advanced Courses Slope Coefficient			
For Intercept 2 /Intercept 3	38.13	1.48	25.83**
District % Passing All TAAS Exit Exams Taken	6.11	1.97	3.10**
District % Students Completing >=1 AP Advanced Course	-5.54	1.73	-3.21**
District % Students Completing >=1 non-AP Advanced Course	-8.32	1.80	-4.62**
For School % Passing All TAAS Exit Exams Taken /Intercept 3	0.59	1.93	0.30
For School % Students in Advanced Courses /Intercept 3	-8.10	1.99	-4.08**
For School % Gr. 11-12 AP Takers with >= 1 Score 3-5 /Intercept 3	1.13	1.48	0.77
For School # Types AP & Non-AP Advnc. Courses Completed /Intercept 3	-1.40	1.74	-0.80
For School # Types AP Subject Courses Completed /Intercept 3	-5.93	2.65	-2.24*
For School % Gr. 11-12 Enrollees Taking >= 1 AP Exam /Intercept 3	2.27	1.62	1.40

Two-tailed univariate test probabilities:
* $p \leq .05$, ** $p \leq .01$

Glossary

Student-Level Variables:

VARIABLE NAME	DESCRIPTION/VALUE CODING
* SAT VERBAL SCORE	1994 graduate's most recent SAT Verbal score ***(200-800).***
* SAT MATH SCORE	1994 graduate's most recent SAT Math score ***(200-800).***
* SAT TOTAL SCORE	1994 graduate's most recent SAT Total score (Verbal+Math) ***(400-1600).***
* OVERALL GPA	Student's overall grade point average ***(1=A+, 2=A, 3=A-, 4=B+, 5=B, 6=B-, 7=C+, 8=C, 9=C-, 10=D+, 11=D, 12=E or F)***
* HIGH SCHOOL RANK	Student's high school rank ***(1=highest 10th, 2=second 10th, 3=second 5th, 4=third 5th, 5=fourth 5th, 6=lowest 5th).***
* COLLEGE ENGLISH PLACEMENT PLANS	A value of ***"1"*** if the student plans to "advance place" out of English, ***"0"*** otherwise.
*AP & NON-AP ADVANCED COURSES COMPLETED	Number of types of AP and non-AP advanced courses, as defined for Texas public schools, completed by SAT-tested graduates in Grades 11-12 ***(0-90 possible statewide).***

School- and/or District-Level Variables by construct area[1]:

Professional Teaching Conditions	
** AVERAGE YEARS OF TEACHER EXPERIENCE	District average years of teacher experience.
** AVERAGE TEACHER SALARY	District average teacher salary.
** AVERAGE STUDENT/TEACHER RATIO	School/District average student/teacher ratio.

**Variable included in at least one of three HLM models presented in this paper.*
***Variable analyzed in current study.*

[1]For purposes of this study, school- and district-level variables were organized under three broad constructs discussed by Oakes (1989): (a) access to knowledge - extent to which schools/districts provide students with opportunities to learn; (b) professional teaching conditions - extent to which certain conditions encourage or constrain teachers from implementing a quality instructional program; and (c) press for achievement - amount of institutional pressure exerted to get students to work hard and achieve.

School- and/or District-Level Variables by construct area[1] (continued):

VARIABLE NAME	DESCRIPTION/VALUE CODING
Access to Knowledge	
* % IN ADVANCED COURSES	School/District percentage of students enrolled in advanced high school courses ***(0%-100%)***.
** % ECONOMICALLY DISADVANTAGED	School/District percentage of students economically disadvantaged ***(0%-100%)***.
* % MINORITY	School/District percentage of minority students ***(0%-100%)***.
**ENROLLMENT	Total number of students enrolled in school/district during school year.
** OPERATING COSTS PER PUPIL	School district operating costs per pupil ***($3443- $5283 in data subset)***.
** PROPERTY VALUE PER STUDENT	School district wealth defined as total taxable property value per student ***($44,428.60- $392,915.90 in data subset)***.
** (NON-)CENTRAL/NORTH REGION	Coded ***"1"*** if district was located in the South, East, or Panhandle/West area of the state; or ***"0"*** if district was located in the Central or North area of state.
** % MINORITY TEACHERS	District percentage of total teachers that were minority ***(0%-100%)***.
* % WHITE TEACHERS	School/District percentage of total teachers that were white ***(0%-100%)***
* % LIMITED ENGLISH-PROFICIENT	School/District percentage of students limited English-proficient ***(0%-100%)***.
** STUDENT MOBILITY RATE	Percentage of students at school district < 2 years ***(0%-100%)***.
** # MATH CLASSES	Total number of math classes offered in district ***(0-1744 in data subset)***.
** # SCIENCE CLASSES	Total number of science classes offered in district ***(0-1692 in data subset)***.
* % COMPLETING AP ADVANCED COURSES	School/District percentage of students completing at least one AP advanced course in 1993-94 as defined for Texas public schools ***(0%-100%)***.
* % COMPLETING NON-AP ADVANCED COURSES	School/district percentage of students completing at least one 1993-94 non-AP advanced course as defined for Texas public schools ***(0%-100%)***.
** # TYPES NON-AP ADVANCED COURSES COMPLETED	Number of types of 1993-94 non-AP advanced courses, as defined for Texas public schools, completed by students within a school/district ***(0-62 statewide)***.
* # TYPES OF AP & NON-AP ADVANCED COURSES COMPLETED	Number of types of 1993-94 AP and non-AP advanced courses, as defined for Texas public schools, completed by students within a school/district ***(0-90 statewide)***.

**Variable included in at least one of three HLM models presented in this paper.*
***Variable analyzed in current study.*

[1]For purposes of this study, school- and district-level variables were organized under three broad constructs discussed by Oakes (1989): (a) access to knowledge - extent to which schools/districts provide students with opportunities to learn; (b) professional teaching conditions - extent to which certain conditions encourage or constrain teachers from implementing a quality instructional program; and (c) press for achievement - amount of institutional pressure exerted to get students to work hard and achieve.

School- and/or District-Level Variables by construct area[1] (continued):

Press for Achievement	
** % PASSING TAAS EXIT-LEVEL READING EXAM	School/District percentage of students passing the Texas Assessment of Academic Skills (TAAS) Exit-Level (grade 10) reading exam *(0%-100%)*.
** % PASSING TAAS EXIT-LEVEL MATH EXAM	School/District percentage of students passing the Texas Assessment of Academic Skills (TAAS) Exit-Level (grade 10) mathematics exam *(0%-100%)*.
* % PASSING ALL TAAS EXIT-LEVEL EXAMS TAKEN	School/District percentage of students passing all tests taken on the Texas Assessment of Academic Skills (TAAS) Exit-Level (grade 10) exam *(0%-100%)*.
** % DIPLOMAS WITH ADVANCED SEALS	School/District percentage expected to graduate with "advanced", or "advanced with honors" seals on their diplomas *(0%-100%)*.
** % SAT- AND/OR ACT-TESTED GRADUATES	School/District percentage of students taking at least one college admissions test (SAT and/or ACT) *(0%-100%)*.
* % ENROLLEES TAKING AT LEAST ONE AP EXAM	School/District percentage of Grade 11-12 enrollees taking at least one AP exam in 1994 *(0%-100%)*.
** % AP EXAM SCORES 3-5	School/District percentage of 1994 Grade 11-12 AP exam scores 3-5 *(0%-100%)*.
* % AP EXAMINEES WITH AT LEAST ONE SCORE 3-5	School/District percentage of 1994 Grade 11-12 AP examinees with at least one score 3-5 *(0%-100%)*.
* # TYPES AP EXAMS TAKEN	School/District number of types of 1994 AP subject exams taken *(0-29)*.
* # TYPES OF AP ADVANCED COURSES COMPLETED	Number of types of 1993-94 AP advanced courses, as defined for Texas public schools, completed by students in a school/district *(0-28)*.
** % EXPENDITURES FOR VOCATIONAL EDUCATION	District percentage of instructional operating expenditures for vocational education *(0%-100%)*.
** % TEACHERS WITH ADVANCED DEGREES	Percentage of teachers with advanced degrees (Master's or Doctoral) in the school district *(0%-100%)*.
** DROPOUT RATE	School/District percentage of students who dropped out of school in grades 7 through 12 *(0%-100%)*.
** % GRADUATES PLANNING COLLEGE	School/District percentage of high school graduates planning to attend college *(0%-100%)*.

**Variable included in at least one of three HLM models presented in this paper.*
***Variable analyzed in current study.*

[1]For purposes of this study, school- and district-level variables were organized under three broad constructs discussed by Oakes (1989): (a) access to knowledge - extent to which schools/districts provide students with opportunities to learn; (b) professional teaching conditions - extent to which certain conditions encourage or constrain teachers from implementing a quality instructional program; and (c) press for achievement - amount of institutional pressure exerted to get students to work hard and achieve.

NCME 1997

TM027456

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Three-Level HLM Modeling of Academic and Contextual Variables Related to SAT Scores in Texas

Author(s): Linda L. Hargrove and Michael X. Mao

Corporate Source: Texas Education Agency

Publication Date: March 25, 1997

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/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. 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 the identified document, please CHECK ONE of the following options and sign the release below.

[✓] Sample sticker to be affixed to document | Sample sticker to be affixed to document [ ]

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here
Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical 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."

Signature: Linda L. Hargrove	Position: Project Director
Printed Name: Linda L. Hargrove	Organization: Texas Education Agency Research and Evaluation Div.
Address: 1701 N. Congress Avenue Austin, TX 78701-1494	Telephone Number: (512) 463-9701
	Date: June 2, 1997

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 24, 1997

Dear NCME Presenter,

Congratulations on being a presenter at NCME[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a written copy of your presentation.

We are gathering all the papers from the NCME Conference. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our process of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with two copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the ERIC booth (523) or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: NCME 1997/ERIC Acquisitions
O'Boyle Hall, Room 210
The Catholic University of America
Washington, DC 20064

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an NCME chair or discussant, please save this form for future use.

ERIC Clearinghouse on Assessment and Evaluation