Research in Higher Education, Vol. 43, No. 2, April 2002 (© 2002)

DOES HIGH SCHOOL MATTER? An Analysis of Three Methods of Predicting First-Year Grades

Gary R. Pike and Joseph L. Saupe

This research evaluated the usefulness of 3 approaches for predicting college grades: (a) traditional regression models, (b) high-school-effects models, and (c) hierarchical linear models. Results of an analysis of the records of 8,764 freshmen at a major research university revealed that both the high-school-effects model and the hierarchical linear model were more accurate predictors of freshman GPA than was the traditional model, particularly for lower ability students. Counter to expectations, the hierarchical linear model was not more accurate than the high school effects model.

KEY WORDS: enrollment management; college grades; high schools.

The growth of the "New Accountability" in American higher education during the 1980s and 1990s has focused increased public attention on the academic success of students as an indicator of the quality and effectiveness of colleges and universities (Adelman, 1999; Ewell and Jones, 1991). Partly in response to increased public scrutiny, and partly to bolster enrollments, colleges and universities have redoubled their efforts to implement programs that improve students' grades, persistence, and graduation (Kellogg Commission on the Future of State and Land-Grant Universities, 1997). Students who are at risk of dropping out of college because of poor preparation have been the focus of many of these interventions (American Council on Education, 1996).

Programs designed to improve students' academic skills can have a substantial effect on success in college. Kulik, Kulik, and Shwalb (1983) examined published and unpublished reports from 60 different programs and found that the interventions, on average, improved students' grade point averages by 0.27 of a standard deviation—the equivalent of a one letter-grade improvement in a

Gary R. Pike, Assistant Vice Chancellor, Student Affairs, and Joseph L. Saupe, Emeritus Professor of Education, University of Missouri–Columbia.

Address correspondence to: Gary R. Pike, Assistant Vice Chancellor, Student Affairs, University of Missouri–Columbia, 1 Parker Hall, Columbia, MO 65211-2325; PikeG@missouri.edu.

course each semester. Pascarella and Terenzini (1991) reviewed more than a dozen studies published after 1983 and also concluded that intervention programs have a substantial positive effect on students' grades, persistence, and graduation, particularly during the first year of college.

The success of academic intervention programs depends, in large part, on accurately identifying students in need of the programs' services (Eno, McLaughlin, Sheldon, and Brozovsky, 1999). Identifying these at-risk students frequently involves calculating predicted first-year grade point averages, or predicted grades in specific courses. According to Pascarella and Terenzini, first-year grades "are probably the single most revealing indicator of . . . successful adjustment to the intellectual demands of a particular college's course of study" (1991, p. 388). Moreover, grades are strongly related to persistence and graduation from an institution, admission to graduate or professional programs, and entry into high-level occupations (Baird, 1985; Pascarella & Terenzini, 1991; Tinto, 1975).

For almost a century, efforts to predict college grades have primarily focused on the predictive power of high school performance (i.e., grades and/or class rank) and scores on standardized tests, such as the American College Testing Program's *ACT Assessment (ACT)* and the College Board's *Scholastic Aptitude Test (SAT).* Models that include measures of high school performance and test scores can be reasonably accurate in predicting first-year grade point average, explaining between one fourth and one third of the variance in first-year grades (Mathiasen, 1984; Mouw & Khanna, 1993). For almost 70 years, researchers have recognized that the quality and effectiveness of the sending high school also has a significant effect on students' performance during college (Lee, Bryk, and Smith, 1993). Surprisingly, relatively few of the prediction models that are designed to identify at-risk students have included measures of high school quality and effectiveness. Given this gap in the literature, the present research asked the question, "Does high school matter in predicting students' grades during their first year of college?"

PREDICTING FIRST-YEAR GRADES IN COLLEGE

Research on the prediction of college grades is almost 100 years old, and most of the early research focused on the use of ability measures and high school performance measures to predict college grades (Fishman, 1957; Odell, 1927; Segal, 1934; Travers, 1949). Based on a review of studies conducted prior to 1983, Mathiasen (1984) concluded that test scores and high school performance were the best predictors of success in college. Since Mathiasen's review, a substantial number of studies have been conducted to assess the effects of student characteristics on first-year grades. These studies have generally found that standardized test scores and high school performance were related to first-year grades and could be used to make accurate and appropriate admission and placement decisions (Cabrera, Nora, and Castañeda, 1993; Eimers and Pike, 1997; Mouw and Khanna, 1993; Noble and Sawyer, 1987, 1997; Pike, 1991; Willingham, 1985).

Recently, Adelman (1999) found that including measures of students' high school course taking significantly improved the accuracy of academic success predictions. Analyzing data from the High School and Beyond (HS&B) sophomore cohort of 1982, he found that that the high school curriculum exerted a more powerful influence on bachelor's degree attainment than did test scores, high school class rank, and high school grade point average. When measures of ability, achievement, and curriculum were combined, they provided the best predictor of graduation.

Several studies have found that students' noncognitive characteristics (e.g., educational aspirations, study habits, and willingness to seek out support), as well as their involvement in high school activities, are significantly related to first-year grades in college (Pascarella and Terenzini, 1991; Williford, 1996). In their review of research on the prediction of college grades, Mouw and Khanna (1993) found that although students' noncognitive characteristics and high school extracurricular activities were significantly related to their first-year grades, the inclusion of these variables did little to improve the explanatory power of predictive models. The inability of noncognitive characteristics and high school involvement to contribute substantially to the prediction of college grades may be due to the strong relationships among noncognitive variables, standardized test scores, and high school performance measures (Noble, Davenport, Schiel, and Pommerich, 1999).

Much of the K–12 research on effective high schools indicates that the characteristics of high schools influence students' high school performance, test scores, and subsequent educational attainment (Lee et al., 1993). Recognizing that differences in high schools can affect students' college grades, several early studies attempted to improve predictions by adjusting measures of high school performance for differences in high schools. For example, Toops (1933) and Reitz (1934) adjusted high school grades by regressing high school grade point averages on aptitude test scores for individual high schools. In both cases, the correlations between high school and college grades increased as a result of the adjustments. Creaser (1965) converted high school class ranks to normalized standard scores and regressed the converted measures on college grade averages for each of 12 high schools. He then substituted the adjusted class ranks in a prediction model for all students. In this case, the adjusted measures predicted college grades better than the original measures. Bloom and Peters (1961) and Tucker (1963) developed regression-based models that adjusted predicted grades based on differences among high schools and differences among colleges. They found that these models significantly improved the prediction of first-year grades in college.

The improvement in predicting first-year grades that is achieved by modeling the effects of both students and high schools simultaneously comes at a cost. Specifically, the use of multiple regression techniques to estimate student and high school effects ignores a fundamental characteristic of the data—that students are nested within high schools. Failure to take into account the nesting of students within high schools violates the assumption of independence of observations in multiple regression (Ethington, 1997). Violating this assumption leads to standard errors for effect parameters that are too small and significance tests that are too liberal, increasing the probability of a Type I error (Ethington, 1997; Raudenbush and Bryk, 1988). The end result may be the inclusion of variables in the prediction model that are not significantly related to first-year grades. A model that includes both student and high school characteristics may also provide a distorted view of the direction of effects for a given high school (Burstein, 1980a, 1980b). The net effect may be a model that accurately predicts first-year grades for students in general, but inaccurately predicts first-year grades for students from a specific high school or set of high schools.

Developments in hierarchical linear modeling (HLM) have substantially improved the ability of researchers to represent accurately the effects of both student and high school characteristics on learning outcomes (Bryk and Raudenbush, 1992; Ethington, 1997). HLM can be viewed as a two-step process.[1] First, a student-level prediction (i.e., regression) model is specified and estimated for each high school. This model includes only student-level variables such as test scores, high school performance measures, and students' first-year grades. The second step in the process involves assessing the variability of the regression parameters across high schools and identifying high school characteristics that are related to the variability in regression parameters. In essence, the second step of the analysis involves regressing the student-level regression parameters on the high school variables (Ethington, 1997). As Raudenbush and Bryk (1988) have shown, the use of hierarchical linear models can produce results that differ substantially from the results produced by traditional regression models.

Based on the results of previous research, it is possible to form three general expectations concerning the relationships among student characteristics, high school characteristics, and first-year grade point averages in college. First, it is reasonable to expect that students' test scores, high school performance, and high school coursework will be significantly related to their first-year grade point averages. Second, the characteristics of sending high schools should also play a role in students' first-year grades. Specifically, high school effectiveness measures should be related to college grades. Third, given the fact that hierarchical linear models more accurately account for the nesting of students within high schools, these models should provide the most accurate predictions of students' first-year grade point averages. These expectations were formally tested in the present research.

RESEARCH METHODS

Conceptual Models

In order to examine the relationships among student characteristics, high school characteristics, and first-year college grades, three models were specified and tested. The first model included three predictors of first-year grades: standardized test scores, high school performance measures, and high school coursework. Analysis of this model provided a direct test of the first expectation. The results also served as a baseline against which the two remaining models could be evaluated. The second model included the three student-level predictors, plus a series of dummy variables representing students' sending high schools. This approach was similar to the procedures used in earlier studies that adjusted for difference among high schools in predicting college grades. The third model was developed using hierarchical linear modeling and contained the three-predictor baseline model at the student level, as well as a school-level model that included measures of high school effectiveness.

Research has identified at least four high school characteristics that may affect students' first-year grades in college. First, size of the sending high school may influence students' college performance. Although it is frequently presumed that students from large high schools will do better in college because they have access to more advanced courses and are better able to cope with the size and complexity of a college campus, research indicates that this is not the case (Lee et al., 1993). Although larger high schools do have greater student demand for varied courses, many of these courses are not academically oriented (Lee, Smerdon, Alfeld-Liro, and Brown, 2000). Moreover, large schools tend to have low levels of social and academic support (Lee et al., 1993, 2000). As a result, students from large schools tend to be less well prepared and have lower levels of achievement than students from smaller schools (Lee and Bryk, 1989; Lee et al., 1993).

A second high school characteristic that may influence success in college is the average ability level of the students in a school. Conventional wisdom suggests that students from high-ability high schools would perform better in college, but the evidence suggests that this is not the case. School-average ability has been found to be negatively related to a student's academic self-concept, high school performance, and educational and occupational aspirations in college (Alwin and Otto, 1977; Marsh, 1987, 1991). School-average ability can also have a substantial negative effect on students' high school class rankings (Marsh, 1991).

A third high school characteristic that may affect first-year college grades is the number or proportion of students from a high school attending a given college. There are at least two ways that attendance patterns can influence college grades. First, the fact that a substantial number of students from a high school attend a given college may encourage the high school to develop courses that better prepare students for specific college courses (Lee et al., 1993). In addition, having several friends and acquaintances from the same high school attend college together provides a peer support group that encourages high levels of involvement and academic success during college (Pascarella and Terenzini, 1991).

A fourth characteristic that can influence success in college is public versus private control of the high school. Several studies have found that, in comparison to student in public schools, students in private (i.e., Catholic) high schools have higher grade point averages, scores on standardized tests, and levels of educational attainment (Evans and Schwab, 1995; Sander, 2000; Sander and Krautmann, 1995). The positive effects of private high schools are most pronounced for inner-city, minority students (Neal, 1997; Sander, 2000). The evidence suggests that the success of students from private high schools is due to better preparation through a strong academic curriculum, an ethos of caring in private schools, and more time spent on homework (Lee et al., 1993; Sander, 2000).

Comparing the results for the baseline model with the results for the high-school-effects model and the hierarchical model provided a test of the expectation that including high school variables would improve the prediction of first-year grades. By comparing the results for the high-school-effects model and the hierarchical model with each other, it was possible to test the expectation that the hierarchical model would be superior to the high-school-effects model. An analysis of results for the hierarchical linear model also provided a test of whether size of high school, school-average ability, attendance patterns, and public or private control would be related to students' first-year grade point average in college.

Participants

The setting for this research was a major research university in the Midwest. During the time period of the study, approximately 17,000 undergraduates were enrolled full time on campus, and slightly more than 80% of all first-year students lived on campus. Admission to the university is considered to be "selective" by the state's coordinating board and "moderately selective" by most college rating services. The university's admission standards are widely publicized within the state and include sliding-scale criteria for *ACT* score and high school percentile class rank.[2] In addition, students must have a high school degree or equivalent and have completed 17 high school units consisting of 4 units of English, 4 units of mathematics, 3 units of natural science, 3 units of social studies, 2 units of foreign language, and 1 unit of fine arts. According to policies of the university system and the state's coordinating board, 10% of the enrolled students in a cohort may be exceptions to admissions policy.

The participants in this study were 8,674 students who began matriculating at the subject university during the Fall semesters from 1996 to 1999. All of the participants were from 1 of 124 in-state high schools. High schools were included in the analyses if at least 20 of their students had entered the university between 1996 and 1998. Approximately 54.1% of the participants were female, 87.0% were white, 6.1% were African American, 2.6% were Asian American, 1.3% were Hispanic, 0.5% were Native American, and 2.5% were from some other ethnic group or did not identify their ethnicity. The average ACT Assessment composite score for these students was 25.6, and their average high school class percentile rank was 75.5. Approximately 83.6% of the students met the high school curriculum requirements that were implemented in Fall 1997. The mean first-year grade point average for the students was 2.75.

Because the purpose of this research was to assess the predictive power of models that included student and high school variables, the participants were divided into two groups. Students in the Fall 1996, 1997, and 1998 cohorts were assigned to a "model-development" group, and students from the Fall 1999 cohort were assigned to a "model-evaluation" group. The three predictive models were developed using the data from the first three cohorts, and the accuracy of the models was tested using data from the Fall 1999 cohort.

Although it would have been advantageous to control for cohort effects across the four entering classes by randomly assigning half of the students to the model-development group and half of the students to the model-evaluation group, this was not possible. All students from the first three cohorts needed to be included in the model-development group to provide sufficient numbers for stable estimates of high school effects. Moreover, a research design with random assignment would not have reflected the reality of institutional research practice where future behavior (i.e., first-year grades) must be predicted using past performance.

Table 1 contrasts the background characteristics, ability measures, and first-year grade point averages of the four entering cohorts. Students in these cohorts did not differ significantly in terms of their gender or ethnicity. Students in the Fall 1996 cohort were significantly less likely to have met core course requirements because these requirements did not go into effect until Fall 1997. There were also statistically significant differences in students' ACT composite scores, high school class percentile ranks, and first-year grade point averages across the four cohorts. However, these differences were trivial, accounting for less than 1% of the variance in ability measures and college grades.

TABLE 1. Comparisons Between the Model-Development and Model-Evaluation Groups

Measure	Fall 1996	Fall 1997	Fall 1998	Fall 1999	R^{2a}
Gender					
Female	55.3%	54.0%	54.1%	53.0%	
Male	44.7%	46.0%	45.9%	47.0%	
Ethnicity					
African American	5.9%	7.3%	5.7%	5.8%	
Asian American	2.7%	3.1%	2.5%	2.2%	
Hispanic	1.4%	1.6%	1.3%	1.1%	
Native American	0.5%	0.3%	0.3%	0.7%	
White	87.5%	85.0%	87.9%	87.5%	
Other/Missing	2.0%	2.7%	2.3%	2.7%	
*Core Course Requirements***					
Met	61.7%	90.5%	91.9%	90.7%	
Not Met	38.3%	9.5%	8.1%	9.3%	
ACT Composite Score**	25.4	25.7	25.8	25.4	0.002
High School Class Percentile Rank*	76.0	75.7	75.9	74.4	0.001
First-Year Grade Point Average**	2.69	2.77	2.74	2.79	0.002

[a]Percentage of variance attributable to group differences.
*$p < 0.01$; **$p < 0.001$.

Measures

All of the measures used in this research were taken directly from student records. Three measures were used to represent student-level variables. Test scores were represented by students' composite scores on the *ACT Assessment*, whereas high school performance was represented by class percentile rank and a dichotomous variable indicating whether students had (1), or had not (0), met the university's core course requirements. Five measures were used to represent school-level variables in this study. The first school-level measure consisted of 123 dummy-coded variables representing the 124 sending high schools. This measure was used to represent differences among high schools in the second model. The remaining four school-level variables were included in the second level of the hierarchical linear model. School size was represented by the mean of the number of students graduating from that high school each year from 1996 to 1998. School-average ability was represented by the mean ACT score of enrolled students from that high school. Attendance patterns were represented by the mean proportion of students in a high school graduating class that attended the university from 1996 to 1998. Control was represented by a dichotomous item indicating whether the high school was private (1) or public (0). First-year cumulative grade point average at the university was used as the criterion variable in this study.

Data Analyses

Prior to specifying and testing the three prediction models, the independent variables were all centered about their respective grand means. That is, the grand mean for an independent variable was subtracted from each student's observed value for that variable. Centering the data allowed the intercepts for the prediction models to be interpreted as the expected first-year grade point average of an average student at an average high school. The effect parameters in the regression model (i.e., *b*s) represented the change in the average student's grades resulting from a 1-unit change in an independent variable (e.g., ACT composite score).

Formal data analyses were carried out in two phases corresponding to model development and model testing. In the first phase of the analyses, data from the Fall 1996–1998 cohorts were analyzed using multiple regression and hierarchical linear modeling. To develop a baseline model, students' first-year grade point averages were regressed on students' ACT scores, class ranks, and core-course indicators. For the high-school-effects model, first-year grade point averages were regressed on the three predictor variables used in the baseline model plus 123 dummy-coded variables representing the sending high schools. An important property of this model was that the intercept represented the expected grade point average of a student who was typical of all students in terms of ACT scores, high school ranks, and meeting the core course requirements, from the high school identified by zeros in all of the dummy codes.

The steps used to develop predictions based on HLM followed the procedures outlined by Bryk and Raudenbush (1992) and Ethington (1997). These procedures utilized the effect parameters (i.e., regression coefficients) to determine the statistical significance of relationships between independent and dependent variables and used the variances in effect parameters across high schools to assess the explanatory power of the relationships.

The first step in the hierarchical linear modeling process involved determining whether there was sufficient variance in first-year grade point averages across high schools to warrant the use of HLM procedures. To answer this question, a model was specified and tested that included an intercept in the student-level model and no other variables. Intercepts represented the mean college grade point averages for each high school. This model was equivalent to a one-way analysis of variance (ANOVA) in which high school was the independent variable and first-year cumulative grade point average was the dependent variable.

Dividing the variance of the intercepts (i.e., high school means) by itself plus the pooled variance within high schools (i.e., the total variance in grades) provided an estimate of the proportion of the variance in grade point averages that was attributable to high schools. This estimate of explained variance was equivalent to a traditional eta-squared coefficient produced by an ANOVA.

The second step in the HLM process involved the within-school regression of first-year grade point averages on the student-level variables (i.e., ACT score, high school class percentile rank, and course requirements being met). As with traditional OLS regression, tests of the effect parameters provided an indication of whether the student-level variables were significantly related to first-year grades. In addition, variances in the effect parameters provided an indication of whether there was sufficient variability in the parameters across high schools to warrant developing a school-level model. Two tests of the variances were utilized. First, chi-square significance tests of the variances were calculated to determine if group differences existed. Second, reliability coefficients were examined to determine if the observed differences among high schools were meaningful.[3] Reliability coefficients of 0.70 or greater were considered an indication of meaningful differences. Examining changes in the pooled within-school variances (i.e., the residuals) for the first and second models provided an indication of the explanatory power of the student-level variables. Dividing the decrease in the pooled within-school variance component from the first to the second model by the within-school variance for the first model identified the proportion of the student-level variance in first-year grades that could be attributed to ACT score, high school class percentile rank, and meeting course requirements.

The final step in the HLM analyses involved specifying and testing a two-level model that included the model used in the second step and a high-school-effects model. High school size, average ability, attendance, and control were included as independent variables in the school-level model. Only those student-level parameters that showed significant and meaningful variability were associated with the school-level measures. Significance tests for the effect parameters identified those school characteristics that were associated with differences in the student-level effects. The reduction in the variance of effect parameters from the second to the third models, when divided by the variance component for the second model, provided an indication of the proportion of the variance in effect parameters that was accounted for by characteristics of the high schools.

Once all three models had been developed using the Fall 1996, 1997, and 1998 cohorts, the intercepts and effect parameters from the models were used to calculate three predicted grade point averages for the Fall 1999 cohort. The mean (1996–1998) values for the high school effectiveness measures were used in these calculations. Two sets of tests were used to evaluate the accuracy of the predicted grades. First, differences (i.e., residuals) and intraclass correlations between actual and predicted grades were calculated to assess the accuracy of the predictions overall. Intra-class correlations were used instead of traditional Pearson product–moment correlations because intraclass correlations are sensitive to differences in both the patterns and magnitudes of scores, whereas product–moment correlations are only sensitive to differences in the patterns of scores (Rummel, 1970). Second, actual and predicted grade point averages were categorized as "at risk" (FYGPA < 2.00), in "good standing" (FYGPA = 2.00–3.24), and "scholarship eligible" (FYGPA ≥ 3.25) based on university policies. A comparison of the proportions of accurate predictions within the three categories provided an evaluation of the models for use in identifying at-risk and high-achieving students.

RESULTS

Model Development

Regression of students' first-year grade point averages on their ACT composite scores, high school class percentile ranks, and measures of whether they had met high school course requirements for admission to the university explained 34.1% ($R = 0.584$) of the variance in first-year grades. The results of this analysis are displayed in Table 2. The expected grade point average of a typical student (i.e., the intercept in the multiple regression model) was 2.731, and the effect parameters for ACT score (0.038), high school class percentile rank (0.021), and meeting course requirements (0.179) were all statistically significant. Results for the high-school-effects model are also presented in Table 2. Including variables representing sending high schools significantly improved the

TABLE 2. Multiple Regression Results for the Baseline and High School Effects Models

Measure	Effect Parameter
Baseline Model ($R^2 = 0.341$)	
Intercept	2.731*
ACT Score	0.038*
High School Class Rank Percentile	0.021*
Course Requirements Met	0.179*
High-School-Effects Model ($R^2 = 0.401$)	
Intercept	2.590*
ACT Score	0.027*
High School Class Rank Percentile	0.026*
Course Requirements Met	0.101*
High School Effects	–0.493 to 1.114

*$p < 0.001$.

power of the predictive model. The percent of variance in students' first-year grade point averages accounted for by the model increased to 40.1 ($R = 0.633$). The expected grade point average of a typical student from the uncoded high school was 2.590, and the effect parameters for ACT score (0.027), high school class percentile rank (0.026), and meeting course requirements (0.101) were all statistically significant. Effect parameters for the dummy-coded high school variables ranged from –0.493 to 1.114.[4]

Table 3 presents the results for the three hierarchical linear models that were specified and tested. Results for the first hierarchical model, essentially a one-way ANOVA in which high schools were the independent variable, produced a statistically significant effect for high schools (2.740). Dividing the variance component for high schools (0.017) by the total variance in the model (0.017 + 0.669) produced an estimate of the proportion of explained variance in unadjusted grade point averages of 0.025. Including students' ACT scores, high school class percentile ranks, and course requirement variables in the second hierarchical model significantly improved the predictive power of the model

TABLE 3. Parameter Estimates, Variance Components, and Reliabilities for the Hierarchical Linear Models

Variable	Parameter Estimate	Variance Component	Reliability
Group Differences			
Intercept	2.740***	0.017***	0.510
Residual		0.669	
Student Effects			
Intercept	2.665***	0.061***	0.759
ACT	0.028***	0.000**	0.104
H.S. Class Rank	0.027***	0.000***	0.349
Course Requirements	0.099***	0.027***	0.267
Residual		0.400	
Final Model			
Intercept		0.039***	0.680
Intercept	2.665***		
Mean ACT	0.062***		
Private	0.125*		
Prop. Attending	1.561***		
ACT	0.026***	0.000*	0.092
H. S. Class Rank	0.027***	0.000***	0.363
Course Requirements	0.100***	0.028***	0.271
Residual		0.400	

*p < 0.05; **p < 0.01; ***p < 0.001.

($R^2 = 0.402$). As was the case with the baseline and high-school-effects models, all student-level parameters in the hierarchical model were statistically significant. An examination of the variance components for the second hierarchical model revealed that there was statistically significant variance in each of the level-1 effect parameters across high schools. However, the reliability coefficients for the level-1 effects indicated that only the variability in the intercepts for individual high schools (i.e., the direct effects of high schools on college grades) was meaningful. Differences among high schools did not substantively alter the relationships between student characteristics and college grades.

In the third hierarchical model, the level-1 intercepts were regressed on the four high school effectiveness measures. Inspection of the results for this model indicated that size of the sending high school was not significantly related to the variance in level-1 intercepts. Consequently, the high school size variable was dropped from the analysis, and the model was respecified and tested. Results for the final model indicated that including the three high school measures explained approximately 36.1% of the variance in the level-1 intercepts. Although meaningful, the magnitude of the remaining variance component and the reliability estimate for the level-1 intercept indicated that a significant amount of the variance in the effects of sending high schools remained unexplained.

Model Evaluation

In order to evaluate the accuracy of the three models, predicted grade point averages were calculated for students in the Fall 1999 entering cohort and then compared with students' actual grades. The results of these comparisons are presented in Table 4. An examination of the results in Table 4 revealed that all three models, on average, underpredicted students' first-year grade point averages. The differences between actual and predicted grade point averages were relatively small, with the high school effects model producing the smallest average residual (0.071) and the hierarchical model producing the largest average

TABLE 4. Accuracy of the Three Prediction Models

	All Students		Proportion of Correct Classifications by GPA Categories		
Model	Mean Residual	Intra-Class Correlation	0.00–1.99	2.00–3.24	3.25–4.00
Traditional	0.076	0.478	0.231	0.864	0.319
H. S. Effects	0.071	0.523	0.333	0.847	0.386
Hierarchical	0.107	0.508	0.342	0.827	0.377

residual (0.107). In addition, the predicted grade point averages produced by the high school effects model had the largest intraclass correlation with actual grades. Although the intraclass correlation between predicted and actual grades was not as great for the hierarchical model, it was larger than the intraclass correlation between actual grades and the predicted grades derived from the traditional model. Thus, for the Fall 1999 entering cohort as a whole, predicted grade point averages calculated using the high-school-effects model were more accurate than grade point averages calculated using either the traditional model or the hierarchical linear model.

Because the effectiveness of an intervention program depends on accurately identifying students who are at risk, the second set of evaluations focused on the classification accuracy of the three prediction models. An examination of these data in Table 4 revealed that none of the prediction models were particularly accurate at identifying at-risk students (i.e., students with actual grade point averages below 2.00). The traditional model accurately classified 23.1% of the students with actual grade point averages below 2.00, whereas the high-school-effects model and the hierarchical model were more accurate (33.3% and 34.2%, respectively). Prediction of students who were not at risk was more accurate. All three models correctly classified more than 80% of the students in good standing and correctly classified between 30% and 40% of the students who were scholarship eligible. Overall, the tests of classification accuracy suggested that both the high-school-effects model and the hierarchical model were more accurate than the traditional model in classifying at-risk students. No substantive differences in predictive accuracy were found between the high-school-effects model and the hierarchical linear model.

DISCUSSION

The results of the present research can be summarized as follows:

1. Consistent with previous research, test scores, high school performance, and courses taken during high school were significantly related to first-year grade point averages. These precollege characteristics explained approximately one third of the variance in students' first-year grades.
2. Including measures of the sending high schools measurably improved the accuracy of predicted grade point averages. Both the high-school-effects model and the hierarchical linear model were able to explain an additional 6% to 7% of the variance in first-year grade point averages. In addition, the models that included measures of the sending high schools more accurately identified students who were at risk of poor grade performance.
3. Counter to expectations, the hierarchical linear model was not measurably more accurate than the high-school-effects model at predicting first-year grades. For the entire Fall 1999 cohort, the hierarchical model was less accu-

rate than the high-school-effects model, and for at-risk students the hierarchical model was only slightly more accurate than the high-school-effects model.

Care should be taken not to overgeneralize these results. The results are specific to a single research university and may not apply to other universities—particularly other universities with different missions and student populations. Moreover, the results of the present research may not be totally representative of the institution in which the study was conducted. Only in-state students and students from high schools that sent at least 20 students to the institution between 1996 and 1998 were included in the research. Including all students and/or a broader range of high schools might have produced different results. The generalizability of the findings were also limited by the predictors included in the models. This is particularly true for the high school characteristics used in the study. Whereas some measures (e.g., public versus private control) accurately represented the high schools, other measures may not have accurately reflected the quality and effectiveness of the sending high schools. The use of mean ACT scores of enrolled students as a measure of school average ability is a case in point. It is doubtful that the mean ACT score of students attending a research university is a good indicator of school-average ability. Students attending the state's elite public institution would, most likely, be among the very best students at some of the high schools. Another basic limitation of this research concerns the methods used to evaluate the three prediction models. Establishing the accuracy of one model over another requires a controlled experiment in which all students received the same academic support services (i.e., no educational intervention affected students' grades). In this study, a controlled experiment was not possible. Consequently, this study assumes that the effects of students' educational experiences represent a constant bias across all three models. This assumption was not tested and represents a limitation of the present research.

Two other interrelated factors may have confounded the results of this research. Combining the Fall 1996, Fall 1997, and Fall 1998 entering cohorts into the model-development group, and using the Fall 1999 entering cohort as the model-evaluation group, leaves open the possibility that cohort effects could have influenced the predictive accuracy of the models. Although cohort effects were relatively minor, the possibility remains that differences in the cohorts could have influenced the findings of this study. A second possible confounding factor in the study is grade inflation, a specific type of cohort effect. Although first-year grade point averages were higher for the Fall 1999 cohort than for the Fall 1996 cohort, the differences were relatively minor and not linear. Moreover, grade inflation should have equally affected the predictive accuracy of all three models.

Despite these limitations, the results of the present research do have important implications for institutional research and practice. First and foremost, this study demonstrates that it is possible to use measures of student aptitude, high school performance, and high school coursework to accurately predict students' first-year grade point averages. This finding is consistent with a substantial body of empirical research (Mathiasen, 1984; Mouw and Khanna, 1993; Noble and Sawyer, 1987, 1997; Pascarella and Terenzini, 1991; Willingham, 1985). Also consistent with previous research is the finding that models based on aptitude, performance, and coursework are not accurate predictors of poor academic performance (Ramist, Lewis, and McCamley, 1990; Ramist, Lewis, and McCamley-Jenkins, 1993). This finding should not come as a surprise given the evidence that success in college has less to do with students' precollege characteristics than with the nature and quality of their college experiences (Pascarella and Terenzini, 1991).

The results of this study also show that including measures of high school quality and effectiveness in models of first-year grades substantially improves the predictive accuracy of the models. In addition, the results of the present study suggest that including measures of high school quality has the greatest impact on the identification of at-risk students. The fact that some of the relationships between high school effectiveness characteristics and first-year grade point averages were not consistent with previous research suggests that additional research is needed to clarify these relationships. In fact, it is possible that the effects of sending high schools on grades are unique to each college and university.

Initially, the finding that the hierarchical model was not a better predictor of college grades than the high-school-effects model seemed surprising. Because the hierarchical model better represents the nesting of students within high schools, it is reasonable to expect that the hierarchical model would more accurately account for the effects of high schools on first-year college grades. Careful reflection suggests two reasons hierarchical models may not be superior to a high-school-effects model in this instance. The first reason grows out of the two distinct uses of multiple regression—prediction and explanation. When the ultimate use of multiple regression is for prediction, the focus is on the accuracy of the numerical value that is produced by the weighted linear combination of variables in the model (i.e., $\hat{Y}$). When multiple regression is used for explanation, interest turns to the contributions made by specific variables (i.e., *b*s) and the statistical significance of those contributions. Violating the assumption of independence of observations in the high school effects model may invalidate significance tests for the effect parameters, but it does not threaten the validity of the overall prediction (Ethington, Thomas, and Pike, in press).

The second reason a hierarchical model may not be superior to a high-school-effects model grows out of the relationship between the two models. Porter and Umbach (2001) noted that a multiple regression model with dummy variables representing level-2 units (e.g., high schools) is equivalent to a hierarchical model in which the variance in intercepts across level-2 units is perfectly explained. Given that the hierarchical model was not able to account for all of the variance among high schools using three measures of effective high schools, it is unrealistic to expect that predictions based on the hierarchical model would be more accurate than predictions based on the high-school-effects model.

Does this mean that a high-school-effects model will always be superior to a hierarchical model? It does not. A hierarchical model may prove to be superior to a high-school-effects model in at least two situations. First, a high-school-effects model is only useful when students are from high schools that have been included in the model previously. When a student is from a high school that heretofore has not been included in the model, it is not possible to calculate a predicted grade point average for that student. Because the hierarchical model makes use of general high school characteristics to represent the effects of individual high schools, it may be possible to calculate grade point averages for students from new feeder high schools. Hierarchical models may also be more useful than high-school-effects models when the relationships between criterion and predictor variables differ by high school (i.e., there is an interaction between high school and an independent variable, such as high school class percentile rank). Although it is theoretically possible to represent these contingent effects in a high school effects model using dummy-coded interaction terms, the procedure can produce inaccurate and difficult to interpret results and significantly reduces degress of freedom in the tests of the model (Stapleton, and Lissitz, 1999).

It is important to recognize that there are limits to the gains that can be made by including additional precollege characteristics in a prediction model. As Baird (1985) noted, as much as one half of the variance in students' college grades may be due to college characteristics and college experiences. College characteristics that have been found to influence students' grades include the selectivity of the institution, the homogeneity of the freshman cohort, and grading practices at the institution (Baird, 1984; Pascarella and Terenzini, 1991; Ramist et al., 1990, 1993). College experiences that may influence grades include academic major, quality of student effort, interaction with faculty and peers, and the supportiveness of the campus environment (Cabrera et al., 1993; Eimers and Pike, 1997; Pascarella and Terenzini, 1991; Pike, 1991). Many of these factors can confound effects to predict students' first-year grades based on their precollege characteristics.

The results of this research also have practical implications for institutional researchers and other university officials. For example, predictions of academic success are frequently used as control variables by institutional researchers interested in evaluating the effects of a particular program, net the effects of entering ability. The measure of entering ability will normally be based on an admission test score, measures of high school performance, and perhaps other variables. The inclusion of high school attended, or characteristics of that high school, in the prediction of entering ability can improve the accuracy of the control variable, thereby providing more accurate assessments of program effectiveness.

Academic advisers also use information about the expected academic performance of freshmen in planning programs of study that will maximize the likelihood of student success. This information may also be useful to instructors in tailoring classes to students' capabilities. It is important that this information be as accurate as possible, and this research demonstrates that taking into consideration a students' high schools improves the accuracy of information about their abilities. This is particularly true for those students who may be most in need of intrusive advising and classroom experiences that are tailored to their needs.

Knowledge about the effects of high schools may be useful at institutions with selective admission policies. This information can be used in making admission decisions aimed at enrolling a student body that will be successful. Of course, the manner in which data about high schools are used needs to be formulated with care because both the integrity if the admission process and positive relationships with sending high schools are at stake. In particular, the relationship between high school and college personnel could be undermined by careless communication. The potential sensitivity of the results of institutional research on the effects of high schools should serve as a reminder that the research be carried out with scrupulous care.

CONCLUSION

It may be, as Ewell and Jones (1991) claim, that success has replaced access as the primary criterion by which colleges and universities are judged in the era of the New Accountability. However, political and financial pressures continue to impel colleges and universities, particularly state land-grant institutions, to make higher education accessible to a growing number of Americans (Kellogg Commission on the Future of State and Land-Grant Universities, 1997). If colleges and universities are to make progress toward the twin goals of access and success, they must deliver effective support programs to students who are at risk of performing poorly during their first year of college. Effective programs, in turn, require delivering services to the students who need them. Because so many different factors can affect students' first-year grades, institutions must make use of available information to improve the accuracy of their efforts to target at-risk students. As this research demonstrates, high school does matter, and institutions would be wise to incorporate information about sending high schools in their targeting efforts.

Acknowledgments. An earlier version of this paper was presented at the annual meeting of the Association for Institutional Research, Long Beach, CA, June 2001.

NOTES

1. Although hierarchical linear modeling is described as a two-step process, calculations of the student-level and school-level effects are performed simultaneously (see Ethington, 1997).
2. To be admissible to a selective state institution, the state's coordinating board requires that a student's ACT score percentile and high school class rank percentile sum to 120.
3. The HLM reliability coefficient is defined as the ratio of the variance in a parameter estimate across level-2 units to itself plus error variance. Thus, the reliability coefficient represents the proportion of variance in a level-1 parameter that is attributable to differences among level-2 units.
4. A complete list of the effect coefficients for individual high schools is available from the first author.

REFERENCES

Adelman, C. (1999). *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment.* Washington, DC: U. S. Government Printing Office.

Alwin, D. F., and Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education* 50: 259–273.

American Council on Education (1996). *Remedial Education: An Undergraduate Student Profile.* Washington, DC: Author.

Baird, L. L. (1984). Predicting predictability: the influence of student and institutional characteristics on the prediction of grades. *Research in Higher Education* 21: 261–278.

Baird, L. L. (1985). Do grades and tests predict adult accomplishment? *Research in Higher Education* 23: 3–85.

Bloom, B. S., and Peters, F. R. (1961). *The Use of Academic Prediction Scales for Counseling and Selecting College Entrants.* New York: The Free Press of Glencoe.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.

Burstein, L. (1980a). The analysis of multilevel data in educational research and evaluation. In D. Berliner (ed.), *Review of Research in Education*, Vol. 8, pp. 158–233. Washington, DC: American Educational Research Association.

Burstein, L. (1980b). The role of levels of analysis in the specification of education effects. In R. Dreeben and J. A. Thomas (eds.), *The Analysis of Educational Productivity, Volume I: Issues in Microanalysis*, pp. 119–190. Cambridge, MA: Ballinger Publishing.

Cabrera, A. F., Nora, A., and Castañeda, M. B. (1993). College persistence: structural modeling of an integrated model of student retention. *Journal of Higher Education* 64: 123–139.

Creaser, J. W. (1965). Predicting college success from equated high school ranks: a cross-validation study. *College and University* 41: 96–100.

Eimers, M. T., and Pike, G. R. (1997). Minority and nonminority adjustment to college: differences or similarities? *Research in Higher Education* 38: 77–97.

Eno, D., McLaughlin, G. W., Sheldon, P., and Brozovsky, P. (1999). Predicting freshman success based on high school record and other measures. *AIR Professional File, No. 72*: 1–9.

Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 12, pp. 165–194. New York: Agathon.

Ethington, C. A., Thomas, S. L., and Pike, G. R. (in press). Back to the basics: regression as it should be. In J. Smart (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 17. New York: Agathon.

Evans, W. N., and Schwab, R. M. (1995). Finishing high school and starting college: do Catholic schools make a difference? *Quarterly Journal of Economics* 110: 941–974.

Ewell, P. T., and Jones, D. P. (1991, July). *Assessing and Reporting Student Progress: A Response to the "New Accountability."* Denver, CO: National Center for Higher Education Management Systems.

Fishman, J. A. (1957). *1957 Supplement to College Board Scores, No. 2.* New York: College Entrance Examination Board.

Kellogg Commission on the Future of State and Land-Grant Universities (1997, April). *Returning to Our Roots: The Student Experience.* Washington, DC: National Association of State Universities and Land-Grant Colleges.

Kulik, C., Kulik, J., and Shwalb, B. (1983). College programs for high-risk and disadvantaged students: a meta-analysis of findings. *Review of Educational Research* 53: 397–414.

Lee, V. E., and Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education* 62: 172–192.

Lee, V. E., Bryk, A. S., and Smith, J. B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (ed.), *Review of Research in Education*, Vol. 19, pp. 171–267. Washington, DC: American Educational Research Association.

Lee, V. E., Smerdon, B. A., Alfeld-Liro, C., and Brown, S. L. (2000). Inside large and small high schools: curriculum and social relations. *Educational Evaluation and Policy Analysis* 22: 147–172.

Marsh, H. W. (1987). The big-fish little-pond effect on academic self-concept. *Journal of Educational Psychology* 79: 280–295.

Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal* 28: 445–480.

Mathiasen, R. L. (1984). Predicting college academic achievement: a research review. *College Student Journal* 18: 380–386.

Mouw, J., and Khanna, R. (1993). Prediction of academic success: a review of the literature and some recommendations. *College Student Journal* 27: 328–336.

Neal, D. (1997). The effect of Catholic secondary schooling on educational attainment. *Journal of Labor Economics* 15: 98–123.

Noble, J., Davenport, M., Schiel, J., and Pommerich, M. (1999). *Relationships Between Noncognitive Characteristics, High School Coursework and Grades, and Test Scores of ACT-Tested Students.* ACT Research Report Series, 99-4. Iowa City, IA: American College Testing Program.

Noble, J., and Sawyer, R. (1987). *Predicting Grades in Specific College Freshman Courses from ACT Test Scores and Self-Reported High School Grades.* ACT Research Report Series, 87-20. Iowa City, IA: American College Testing Program.

Noble, J., and Sawyer, R. (1997). Alternative methods for validating admission and course placement criteria. *AIR Professional File, No. 63*: 1–9.

Odell, C. W. (1927). Attempt at predicting success in freshman year at college. *School and Society* 25: 702–706.

Pascarella, E. T., and Terenzini, P. T. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research.* San Francisco: Jossey-Bass.

Pike, G. R. (1991). The effect of background, coursework, and involvement on students' grades and satisfaction. *Research in Higher Education* 32: 15–30.

Porter, S. R., and Umbach, P. D. (2001). Analyzing faculty workload data using multilevel modeling. *Research in Higher Education* 42: 171–196.

Ramist, L., Lewis, C., and McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT. In W. Willingham, C. Lewis, R. Morgan, and L. Ramist (eds.), *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades*, pp. 253–288. Princeton, NJ: Educational Testing Service.

Ramist, L., Lewis, C., and McCamley-Jenkins, L. (1993). *Student Group Differences in Predicting College Grades: Sex, Language, and Ethnic Groups.* College Board Report 93–1. New York: The College Board.

Raudenbush, S. W., and Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Rothkopf (ed.), *Review of Research in Education*, Vol. 15, pp. 423–477. Washington, DC: American Educational Research Association.

Reitz, W. (1934). Prediction college achievement with marks and ranks adjusted for inter-high school variability. *Bulletin of the American Association of College Registrars* 10: 162–181.

Rummel, R. J. (1970). *Applied Factor Analysis.* Evanston, IL: Northwestern University Press.

Sander, W. (2000). Catholic high schools and homework. *Educational Evaluation and Policy Analysis* 22: 299–311.

Sander, W., and Krautmann, A. C. (1995). Catholic high schools, dropout rates, and educational attainment. *Economic Inquiry* 33: 217–233.

Segal, D. (1934). *Prediction of Success in College* (Bulletin no. 15, U. S. Office of Education). Washington, DC: U. S. Government Printing Office.

Stapleton, L. M., and Lissitz, R. W. (1999, March). Evaluating faculty salary equity using hierarchical linear modeling. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research* 45: 89–125.

Toops, H. A. (1933). The transmutation of marks. *Ohio College Association Bulletin, No. 88*: 1093–2000.

Travers, R. M. W. (1949). Significant research on the prediction of academic success. In W. T. Donahue, C. H. Coomb, and R. M. W. Travers (eds.), *The Measurement of Student Adjustment and Achievement*, pp. 147–190. Ann Arbor: University of Michigan Press.

Tucker, L. R. (1963). *Formal Models for a Central Prediction System* (Psychometric Monograph no. 10). Richmond, VA: William Byrd Press.

Williford, L. E. (1996, October). The freshman year: How do personal factors influence academic success and persistence? *1996 SAIR Annual Report: Charting the Course in a Changing Environment.* Mobile, AL: Southern Association for Institutional Research and Southern Region of the Society for College and University Planning.

Willingham, W. (1985). *Success in College: The Role of Personal Qualities and Academic Ability.* New York: The College Board.