Should Grade Retention Decisions Be
Indicators-based or Model-driven?

CSE Technical Report 716

Jia Wang and Haiwen Wang
CRESST/University of California, Los Angeles

May 2007

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
GSEIS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences (IES), U.S. Department of Education.

SHOULD GRADE RETENTION DECISIONS BE INDICATORS-BASED OR MODEL-DRIVEN?

Jia Wang and Haiwen Wang

CRESST/UCLA

Abstract

This study evaluates a large urban district's standards-based promotion policy decisions against a model-driven classification. Hierarchical logistic regression was used to explore factors related to grade retention at both the student and school level. Statistical results indicate that using students' next year achievement test scores as criteria, the model-driven classification of retention is better than the district's standards-based promotion policy decisions. Students who were predicted by the model to be retained, but were not retained, had lower test scores than students who were actually retained. School districts may incorporate this model-based approach into their current promotion and retention decision-making as an additional measure to improve the precision of their decisions.

Grade retention is sometimes considered to be a method to improve students' long-term achievement by allocating them more time to develop adequate academic skills (Holmes, 1989; Reynolds, 1992). Some educators view retention as a means of reducing skill variance in the classroom so that student needs can be better met (Owings & Magliaro, 1998). However, there are no uniform objective standards for making the retention decision and researchers even disagree as to the efficacy of grade retention (Holmes & Mattews, 1984; Jackson, 1975).

The effects of grade retention have been at the center of heated debates for many years. This is due to the consistent application of grade retention in public schools despite the large quantity of research evidences against the effectiveness of the practice (Byrnes & Yamamoto, 1986). In the current era of accountability and raising educational standards, retention is becoming a more popular practice across the nation

(National Association of School Psychologists, 2003). Lacking a practical alternative, some researchers point out the need to examine and improve the methods and criteria being used to retain students for better academic achievement (Office of University Communication, University of Pennsylvania, 2005). The purpose of this paper is to evaluate the efficacy of the standards-based promotion policy decisions of a large urban district against a model-driven classification.

Other researchers claim that there are long-term damaging effects for grade retention in students' later development (McCoy & Reynolds, 1999). For example, Meisels and Liaw (1993) used a sub-sample of black, white, and Hispanic children from the National Education Longitudinal Study of 1988 (NELS88) to estimate the effects of retention from kindergarten to eighth grade. They concluded that retained children had lower standardized test scores and academic grades than non-retained children, and retained children also had a higher incidence of special education placement. McCoy and Reynolds (1999) report similar results; they found a significant association between grade retention and significantly lower achievement in reading.

Negative effects of retention on both school achievement and social-psychological development are also reported in various research studies (e.g., Holmes, 1989; Peterson, DeGracie, & Ayabe, 1987). Holmes did a meta-analysis on 63 studies that compared the long-term performance of retained and non-retained students. In 54 studies, Holmes noted worse performance on standardized tests for the retained students when being promoted to the next grade, compared to their non-retained peers; Only nine studies showed positive retention effects, especially when individualized help to improve study skills was available.

In a longitudinal study conducted by Jimerson and Schuder (1996), a group of students who were retained during kindergarten to third grade were tracked until they were in high school. Jimerson and Schuder found that in addition to having a lower grade point average (GPA) and academic standing in the long run, these previously retained students were significantly disadvantaged in emotional well-being and disciplinary actions, compared to their fellow students who were never retained. Peterson et al. (1987) reached similar results with a twist. They found that the retained students improved in their relative classroom standings in the first three years after being retained, while compared to the promoted students at the same age. However, they failed to maintain the advantage after the first three years.

Past grade retention was also found to increase students' chances of dropping out, even after adjusting students' background variables and academic achievement (Rumberger & Larson, 1998). Roderick (1994) reported the risk of dropping out of school was two times greater for children who had been retained between kindergarten and the sixth grade than for children who were never retained. Analyzing the NELS88 data, Rumberger and Larson concluded that the odds of dropping out during high school were 2.56 times greater for students who were retained in grade eight and more than four times greater for students who were retained before grade eight.

Other longitudinal studies extending to high schools strongly suggest that children who were previously retained had significantly higher high school dropout rates (Holmes, 1989; Jimerson & Schuder, 1996; Roderick, 1994; Rumberger, 1995). Temple, Reynolds, and Miedel (1998) found that earlier grade retention between kindergarten and the eighth grade increased students' dropout rate before high school graduation by 13%, and it was a significant predictor of high school dropout even after controlling for students' background variables, etc. In the study conducted by Grissom and Shepard (1989), they tested whether grade retention was related to dropping out of school between grade seven and grade twelve. They found that retention increased the possibility of students' dropping out even after controlling for the effects of gender, ethnicity, socio-economic status (SES), and achievements levels in all the subjects.

While the effectiveness of grade retention is doubtful and gloomy, the cost of grade retention is very high. According to the Center for Policy Research in Education (1990), the annual cost of grade retention is about $4,050 per pupil. The estimated annual cost of grade retention for about 2.4 million students amounts to around $10 billion. If retention does not work or does not work efficiently, it might be a better idea to use the $10 billion on alternative remedial programs that are more effective.

Previous Research on Retention Risk Factors

Previous research studies have identified a list of student, teacher, and family factors that are associated with students' retention chances at schools. The main list consists of gender, race/ethnicity, SES, enrollment in special education programs, and students' reading and mathematics test scores (McCoy & Reynolds, 1999). Specifically, it has been found that boys are more likely to be retained than girls (Dauber, Alexander, & Entwisle, 1993; Gottfredson, Fink, & Gram 1994; Owings & Magliaro, 1998). Students from ethnic minority groups and from lower socio-economic backgrounds (including lower parent educational attainment and income) are more likely to be retained

(Bianchi, 1984; Meisels & Liaw, 1993; Owings & Magliaro, 1998; Thomas, Sammons, & Smees, 1997). Hauser, Pager, & Simmons (2000) identified white girls as the most successful group progressing through school and African American boys as the most held-back group. Students enrolled in special education programs are also found more likely to be retained (Yuan, Pernici, & Franklin, 2001). In addition to the above factors, Grissom and Shepard (1989) also indicated that students who are younger than their classmates or students with an immature appearance are more likely to be retained.

Regarding school achievement, students who have poor reading and mathematics test scores have been found more likely to be retained, as expected (Dauber et al., 1993; Meisels & Liaw, 1993; McCoy & Reynolds, 1999; Reynolds, 1992). Subject-matter grades (Cadigan, Entwisle, Alexander, & Pallas, 1988; Meisels & Liaw, 1993) and classroom conduct (Dauber et al.) are also predictive of students' later grade retention.

Besides the above risk factors, children who are considered as having poor peer relationships by their teachers (Cadigan et al., 1988; Jimerson & Schuder, 1996), who have poor emotional well-being (Jimerson & Schuder), who have adjustment problems (Reynolds, 1992), and who are viewed by their parents as less capable (Cadigan et al.; Dauber et al., 1993) are more likely to be retained. McCoy and Reynolds (1999) concluded in their study that lack of parental participation in school activities and the number of times a student changes schools increase the possibility of retention.

Current Promotion Policy at the District

Since February 1999, the district from which we obtained the data has been using a standards-based promotion policy, as documented in its website in September 2002. According to the district policy, multiple measures of achievement and teacher judgment are used to determine whether a student can be promoted. The multiple measures include the Stanford Achievement Test, Version 9 (SAT9) reading, mathematics, and language scores and the English language arts performance assignment (PA) scores. For the purpose of this paper, we call the district's way of promotion and retention decisions "indicators-based" decisions, with achievement measures and teacher judgment as indicators.

Purposes

The main purpose of this study is to find out whether a statistical model provides a more accurate classification of students and a better reflection of how well they would do in the later school achievement tests than the current promotion and retention policy at the school district. The statistical prediction of students' promotion and retention status was based on the combination of student background variables and their previous school achievement scores. We were also interested in finding out whether there is any school variation in the decision of student promotion and retention decisions. This study further explored whether students' scores on a district language arts performance assessment contributed information to a retention decision beyond the effects of background variables, grade point average, and standardized achievement scores in reading and mathematics.

Research Questions

The following research questions were addressed in this study:

1. What is the agreement between students' actual retention status and the prediction based on a statistical model that took into consideration students' background variables and their previous school achievement measures?
2. Do students' PA scores contribute to their retention status beyond the effects of their background variables, GPA, and standardized achievement scores in reading and mathematics?
3. If we base the promotion and retention policy on a statistical model, will it classify students more accurately, judged by how well they would do in their later school achievement tests, compared with the current promotion policy?
4. Do schools vary in their student promotion/retention decisions? If they do, how do school characteristics affect their promotion/retention variations?

Data

We received data from a large urban school district on the West Coast. The district's database contains a wide range of information on students, including demographic variables, family background information, school program information, and previous achievement scores. The demographic variables include students' gender and ethnicity. The family variables include students' lunch program status, Title 1 status, language proficiency, immigrant status, special education status, and home language classification. The school variable is students' gifted status. The previous achievement measures are students' 2002-2001 GPA and the SAT9 reading and mathematics scores measured in normal curve equivalent scores (NCES).

The data came from 18,811 students who were enrolled at the district as 9th graders in 2000-2001, and were still enrolled in the district the next year. All of the students in the dataset had taken a district English language arts performance assessment in the 2000-2001 school year. English language learners (ELL) below the advanced level did not take the district test and are not included in the dataset.

Table 1 shows distribution information for the demographic, school, and family variables used in the study. The data reveals that 51% of the students are female; 63% are Title 1 students; 11%, gifted students; 1%, special education students; 23%, immigrants; and 61%, students receiving free or reduced-fee lunch at school. In terms of ethnicity, 74% of the students are Hispanic, 9% are black, 10% are white, 4% are Asian, and the remaining 3% belong to other ethnic groups. The population consists of 19% ELL students; 47% former ELL students, and 35% English Proficient (EP) students. Based on a home language survey conducted by the district, 67% of the students speak Spanish at home, 26% speak English, and 8% speak languages other than Spanish and English at home.

Table 1 shows the cross-tabulation of students' retention status by their background variables. Overall, 21% of the students were retained. The retention rate for boys (24.4%) is 7% larger than that for girls (17.2%). Hispanic students were found to have the highest retention rate (23.7%), followed by the black (19.2%), white (9.4%), all other (8.5%), and Asian students (6.9%). Students who paid for lunch had a slightly higher retention rate (21.3%) than students in the free or reduced fee lunch program (20.4%).

A larger percentage of Title 1 students (23.5%) were retained than non-Title 1 students (16.2%). Similar proportions of immigrant and non-immigrant students were retained (20.4% and 20.8% respectively). Not surprisingly, a much larger percentage of non-gifted students were retained (22.5%) than gifted students (6.3%; that any gifted student was retained might be considered surprising, however).

Table 1
Descriptive Information on the Student Background Variables (N=18,811)

Variable	Definition	Value Label	Percentage of Students in the Data	Percentage of Students Retained
Female	Gender			
		Female	50.5	17.2
		Male	49.5	24.4
Ethnicity	Ethnicity			
		Asian	3.9	6.9
		Black, Not Hispanic	9.0	19.2
		Hispanic	73.9	23.7
		White, Not Hispanic	10.3	9.4
		All Other	2.9	8.5
Lunch	Lunch Program			
		Free & reduced price	60.9	20.4
		Others	39.1	21.3
Title1	Title1 Student			
		Title1	62.5	23.5
		Non-Title1	37.5	16.2
Immigt	Immigrant Status			
		Immigrant	22.9	20.4
		Non-immigrant	77.1	20.8
H_lang	Home language			
		English	25.5	17.1
		Spanish	66.5	23.7
		Others	8.0	7.9
Langprof	Language Classification			
		EP	34.5	16.9
		ELL	18.4	35.1
		Former ELL	47.1	18.0
Gifted	Gifted student			
		Gifted	10.7	6.3
		Non-gifted	89.3	22.5
Sped	Special Education			
		Special education	0.6	23.5
		Non-special education	99.4	20.7

Students in special education had a higher retention rate (23.5%) than other students (20.7%). Students who speak Spanish at home had a higher retention rate (23.7%) than students who speak English at home (17.1%), who in turn had a higher retention rate than those who speak other languages at home (7.9). ELL students had a higher retention rate (35.1%) than the former ELL students (18.0%), who in turn had a higher retention rate than EP students (16.9%).

To get an idea about how students did in their school achievement tests, we calculated their mean GPA, reading, mathematics and performance assessment scores in 2000-2001, and reading and mathematics scores in 2001-2002, by their retention status. Looking at Table 2, the retained students had significantly lower mean scores in every measure, as expected. Please note that the achievement gaps in reading and mathematics scores were a little narrower for 2001-2002, which may indicate that retention was somewhat effective in improving students' performance.

Methodology

The hierarchical nature of the data—students nested within schools—made it necessary to choose a methodology that could account for this nesting of the data in its estimations. A brief methodological discussion is given below to facilitate the interpretation of the formal models employed for the estimations. Interested readers should refer to Kreft and Leeuw (1998), Raudenbush and Bryk (2002), and to selected references cited in these books for a more full explanation of the conceptual and methodological details of hierarchical modeling.

Table 2
Mean Scores and their Standard Deviations by Retention Status (N=18,811)

Retention Status	GPA 01*	Reading 01*	Math 01*	PA01*	Reading 02*	Math02*
Not retained	2.5	37.7	45.6	2.2	38.9	48.9
	(0.79)	(17.59)	(17.09)	(0.84)	(16.04)	(16.80)
Retained	1.3	32.4	39.8	1.7	28.6	37.3
	(0.65)	(14.92)	(13.52)	(0.71)	(13.41)	(12.38)
Total	2.3	36.7	44.4	2.1	36.7	46.5
	(0.92)	(17.23)	(16.61)	(0.84)	(16.08)	(16.66)

*Analysis of Variance shows significant difference between the mean scores of retained and not retained students at level .05.

In a multilevel model the random variability in the variables observed is decomposed between the basic smallest unit of analysis, conventionally known as Level 1, and the higher level grouping, Level 2. Emphasis is placed on defining and exploring variations at each level and how such variances are related to explanatory variables. In hierarchal models, the possibility of explicitly modeling cross-level interactions or variances associated with the errors at each level of the hierarchy allows for more interesting questions to be asked of the data.

In the present analysis, the nested nature of the data can be viewed in a hierarchical framework with student observations (Level 1) nested in higher level (Level 2) groups (i.e., schools). A multi-level model allows the impact of student variables to differ according to the school context (technically, variation in slopes). We chose the hierarchical generalized linear model with two sub-models, namely the Level-1 model and Level-2 model. Specifically, the Level-1 model represents the relations among the student level variables and the Level-2 model represents the influence of school-level factors on the outcome variable through student-level variables (Raudenbush & Bryk, 2002).

Since our outcome variable is a binary variable—student grade promotion/retention status—we adopted the most common link function (logit link) for our binomial Level-1 sampling model. The logit link function makes use of logistic regression (Raudenbush & Bryk, 2002). Logistic regression is a mathematical modeling approach that is typically used to describe the relationship of several independent variables to a dichotomous dependent variable (Long, 1997). It is designed to describe the probability (which will always be some value between 0 and 1) of analyzed subjects having the value of 0 or 1 for the dependent variable. Logistic regression is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables (Long).

The logit link function for our model is:

$$\eta_{ij} = \log (\varphi_{ij}/ (1-\varphi_{ij}))$$

where η_{ij} is the log of the odds of being retained and φ_{ij} is the probability of being retained: Prob (Retention = 1) = φ_{ij}. From the above equation we obtain that $\varphi_{ij} = \exp (\eta_{ij}) / (1 + \exp (\eta_{ij}))$, where $\exp (\eta_{ij})$ is the odds of being retained.

Model Specifications

We compiled four school variables from the district's public websites: (a) percentage of minority students, (b) percentage of students receiving free/reduced fee lunch, (c) number of students enrolled (in 1,000s), and (d) percentage of teachers with credentials. Descriptive information for these variables is summarized in Table 3. The inclusion of school variables is important since students with similar backgrounds tend to go to the same schools, and different schools tend to attract different teachers with different teaching philosophies, knowledge, and experience, which may result in different promotion and retention decisions for students in similar conditions. Excluding school variables and ignoring the student-school nesting would cause estimation biases and fail to capture the true relations among variables.

Table 3
Descriptive Information on School Level Variables

Variable	N	Mean	Std. Deviation
Percent minority students	52	0.88	0.14
Average free meal status	52	0.57	0.22
Percent English learners	52	0.24	0.11
Enrollment	52	3204	926
Percent full teacher credentials	52	0.74	0.08

The final model specification takes the following form for the Level 1 equation:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}\,\text{FEMALE} + \beta_{2j}\,\text{ASIAN} + \beta_{3j}\,\text{BLACK} + \beta_{4j}\,\text{HISPANIC}$$

$$+ \beta_{5j}\,\text{ETH\_OTHER} + \beta_{6j}\,\text{IMMIGRANT} + \beta_{7j}\,\text{SPANISH}$$

$$+ \beta_{8j}\,\text{HL\_OTHER} + \beta_{9j}\,\text{FREE-LUNCH} + \beta_{10j}\,\text{ELL} + \beta_{11j}\,\text{F\_ELL}$$

$$+ \beta_{12j}\,\text{GIFTED} + \beta_{13j}\,\text{TITLE 1} + \beta_{14j}\,\text{SPED} + \beta_{15j}\,\text{READING}$$

$$+ \beta_{16j}\,\text{MATHEMATICS} + \beta_{17j}\,\text{GPA} + \beta_{18j}\,\text{PA}$$

The Level 2 model specifications are as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,\text{MINORITY\%} + \gamma_{02}\,\text{FREE-LUNCH\%} + \gamma_{03}\,\text{ENROLLMENT}$$

$$+ \gamma_{04}\,\text{TEACHER\_CRED \%} + u0j$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50}$$

$$\beta_{6j} = \gamma_{60}$$

$$\beta_{7j} = \gamma_{70}$$

$\beta_{8j} = \gamma_{80}$

$\beta_{9j} = \gamma_{90} + u_{9j}$

$\beta_{10j} = \gamma_{100}$

$\beta_{11j} = \gamma_{110}$

$\beta_{12j} = \gamma_{120}$

$\beta_{13j} = \gamma_{130}$

$\beta_{14j} = \gamma_{140}$

$\beta_{15j} = \gamma_{150}$

$\beta_{16j} = \gamma_{160}$

$\beta_{17j} = \gamma_{170} + u_{17j}$

$\beta_{18j} = \gamma_{180}$

where γ_{01} to γ_{04} are the estimated coefficients for school level variables and u_{0j} is the unique effect of school j on the Level-1 intercept. The values of γ_{10} to γ_{170} are the estimated coefficients for student level variables, u_{9j} is the unique effect of school j on the coefficient of student lunch program status, and u_{17j} is the unique effect of school j on the coefficient of GPA score.

In the previous model specification, besides having Level-2 predictors explain the intercept β_{0j}, we relaxed the assumption that the impact of student lunch program status and performance assessment scores on retention status is the same across schools. We chose to let student lunch program status and GPA scores have a school-specific effect, based on previous literature (Raudenbush & Bryk, 1988). Therefore there are the error terms of u_{9j} and u_{18j} in the corresponding models to see whether student lunch program status and GPA score impact students' retention status differently across schools.

We also applied the same models to a larger dataset of 24,648 Grade 9 students from the district by relaxing the requirement of having taken the performance assessment in 2000-2001. The description and results are summarized in Appendix A.

Statistical Results

This section reports the results based on the last model specification (Please see Appendix B for the other model specifications we explored). The last model specification included both student variables and school level variables (to model school intercept), while allowing the effects of both student lunch program and GPA scores to vary across schools.

The HLM results are reported in Table 4 and Table 5. Based on the coefficient results summarized in Table 4, we can see that the effects of (a) immigrant status, (b) home language, (c) language proficiency, (d) special education status, (e) Title 1, and (f) gifted student status on whether a student was retained were not statistically significant, after controlling for all other variables in the estimation model. Holding other variables constant and comparing to their corresponding peers, female students were less likely to be retained, students in free and reduced lunch program were less likely to be retained, and Hispanic and Black students were more likely to be retained than white students. The odds of being retained for Asian and other ethnic students were not significantly different from white students.

Table 4
Estimation of Fixed Effects of Student and School Level Variables (N =18,811)

Variable		Coefficients		Standard Error	Adjusted Odds Ratio
Level-1: Student characteristics					
Female	γ_{10}	-0.236	*	(0.04)	0.79
Ethnicity-Asian	γ_{20}	0.384	*	(0.15)	1.47
Ethnicity-Black	γ_{30}	0.250	*	(0.01)	1.28
Ethnicity-Hispanic	γ_{40}	0.496	*	(0.10)	1.64
Ethnicity-Other	γ_{50}	0.016		(0.14)	1.02
Immigrant Status	γ_{60}	-0.050		(0.04)	0.95
Home Language - Spanish	γ_{70}	-0.094		(0.09)	0.91
Home Language - Other	γ_{80}	-0.134		(0.11)	0.87
Free/Reduced Fee Lunch	γ_{90}	-0.146	*	(0.06)	0.86
ELL	γ_{100}	-0.054		(0.09)	0.95
F_ELL	γ_{110}	-0.046		(0.08)	0.96
Gifted	γ_{120}	-0.096		(0.08)	0.91
Title1	γ_{130}	-0.003		(0.07)	1.00
Special Education	γ_{140}	-0.119		(0.29)	0.89
Reading 2001	γ_{150}	-0.005	*	(0.00)	1.00
Math 2001	γ_{160}	-0.013	*	(0.03)	0.99
GPA 2001	γ_{170}	-1.944	*	(0.05)	0.14
PA 2001	γ_{180}	-0.144	*	(0.03)	0.87
Level-2: School characteristics					
Intercept	γ_{00}	6.639	*	(1.11)	764.33
Percent of minority students	γ_{01}	-3.22	*	(0.67)	0.04
Percent of free meal	γ_{02}	1.535	*	(0.44)	4.64
Enrollment	γ_{03}	0.255	*	(0.00)	1.29
Percent of teacher credential	γ_{04}	-3.189	*	(1.23)	0.04

*Variable coefficients denoted with * indicate the variables are significant at .05 significant level. The standard deviations of the coefficients are shown in parenthesis.

Both reading and mathematics scores were negatively related to students' possibility of being retained or not. The higher the reading and mathematics scores were, the less likely that the student was retained. GPA also had a negative correlation with students being retained. The higher the GPA score, the less likely a student was to be retained. The same relation was found for the writing performance assessment scores—the higher the performance assessment score, the less likely that the student was retained.

At the school level, the percent of minority students and percent of teachers with credentials are negatively related to students being retained or not, while school enrollment size and the percent of students receiving free/reduced lunch have a positive relationship with being retained. These findings indicate that in schools with larger proportions of minority students and the larger percentages of teachers with credential, students were less likely to be retained. Students in schools with larger percentages of students receiving free/reduced fee lunch and with larger enrollment are more likely to be retained.

Table 5
Estimation of Variance Components

	Variance component	P-value
Intercept	1.531 *	0.000
Free Lunch	0.083 *	0.034
GPA	0.140 *	0.000

*Variance components denoted with * indicate they are significant at .05 significant level.

Table 5 reports the random effects in the model. The estimations presented allow us to describe the extent of the unexplained variance (after controlling for student and school factors) in student retention status and its division between the two levels of analysis. The random effects for school means and the coefficients of free lunch and GPA score are significant at the p level of 0.05. These indicate that after controlling for the school characteristics, there are still significant variations in the retention profile among schools, and lunch program and GPA score have varying effects on retention across schools. At the same time, the group (school) effects are significant, pointing to the influence of school specific (unobserved) factors.

We calculated the school specific retention values for each school in our data (see Table 6) using the coefficients reported in Table 4. Looking at Table 6, we could see that there was a large variation in mean school retention probability—the probability values ranged from 1.6% to 37.1% for this group of 52 schools. These numbers represented the school-specific typical retention probability for its students. For example, the probability of being retained for School 34 was estimated to be 1.6% and it was 37.1%

for School 24. Table 7 also lists out the number of students included in the analysis, school-specific intercept, school-specific lunch program status effect, and school-specific GPA effect.

School characteristic variables such as percentage of minority students, percentage of students receiving free/reduced fee lunch, school enrollment size, and percentage of teachers with a credential were all found to have significant effects on schools' overall retention rate. Student lunch program status and GPA score variables were also found to have varying effects on whether a student was retained across schools. We found that students were not retained according to a uniform standard across schools and there were a wide range of school variations.

Achievement Comparison Based on Classifications

Using the coefficients estimated in the last model specification, as reported in Table 4, we calculated the predicted retention status for students in the data to see how well the final estimation model did in classifying students. Based on whether the students were predicted to be retained and whether they were actually retained, we grouped students into the following four groups:

Group A - students not predicted to be retained and were not retained;
Group B - students not predicted to be retained but were actually retained;
Group C - students predicted to be retained but were not retained;
Group D - students predicted to be retained and were retained.

Table 6
Retention Probability, Intercept and Lunch and GPA Slopes for Each School

School Name	Number of students	Retention probability	Intercept	Lunch Slope	GPA slope
Mean	355	20.1%	0.20	-0.17	-2.18
Minimum	46	1.6%	-7.72	-0.48	-2.69
Maximum	857	37.1%	4.77	0.26	-1.40
School 1	104	24.0%	4.14	0.00	-2.69
School 2	150	24.0%	3.77	-0.05	-2.48
School 3	397	20.9%	0.56	-0.18	-2.21
School 4	857	23.0%	1.29	-0.41	-2.36
School 5	634	27.3%	-1.50	0.04	-1.82
School 6	425	26.4%	2.36	-0.33	-2.32
School 7	254	30.7%	2.68	0.02	-2.50
School 8	516	17.2%	0.72	-0.45	-2.04
School 9	521	20.0%	0.67	0.08	-2.40
School 10	284	18.3%	-0.30	-0.05	-2.18
School 11	216	26.9%	1.73	-0.28	-2.36
School 12	118	17.8%	-1.05	-0.10	-2.08
School 13	341	15.5%	0.67	0.12	-2.41
School 14	591	10.8%	-0.70	-0.02	-2.30
School 15	310	16.8%	-1.07	-0.22	-2.02
School 16	234	1.8%	-6.21	-0.33	-1.57
School 17	110	10.1%	-1.27	-0.28	-2.07
School 18	516	32.6%	-1.31	-0.29	-1.73
School 19	303	12.2%	-0.90	-0.14	-2.14
School 20	567	16.9%	-1.70	0.11	-1.93
School 21	617	16.9%	2.85	-0.06	-2.67
School 22	90	28.9%	-1.50	-0.33	-1.85
School 23	224	11.2%	-3.10	-0.08	-1.96
School 24	369	37.1%	-0.78	-0.03	-2.03
School 25	700	29.0%	2.41	-0.26	-2.27
School 26	406	35.2%	2.03	-0.36	-2.06
School 27	211	8.6%	-3.32	-0.33	-1.80
School 28	364	10.4%	-0.96	-0.13	-2.11
School 29	211	6.2%	3.91	-0.12	-2.59
School 30	396	21.5%	0.29	0.10	-2.35
School 31	46	26.1%	0.03	-0.07	-2.25
School 32	384	15.4%	-2.80	-0.48	-1.69
School 33	91	26.4%	4.34	-0.15	-2.44
School 34	315	1.6%	-7.72	-0.44	-1.40
School 35	515	14.6%	-1.04	-0.28	-2.09

	Mean	250	19.8%	0.87	-0.22	-2.21
	Minimum	70	8.8%	-3.60	-0.39	-2.56
	Maximum	378	34.8%	4.77	-0.09	-1.84
School 36		493	28.0%	0.66	-0.07	-2.33
School 37		362	13.5%	0.29	-0.34	-2.25
School 38		540	21.7%	1.69	-0.19	-2.57
School 39		314	16.9%	3.22	-0.13	-2.58
School 40		239	28.5%	1.93	-0.11	-2.42
School 41		641	21.1%	-2.86	-0.39	-1.62
School 42		570	34.6%	-0.51	0.26	-2.15
School 43		321	8.4%	-1.98	-0.08	-1.97
School 44		731	20.5%	0.44	-0.17	-2.20
School 45		462	25.3%	2.77	-0.13	-2.56
School 46		378	19.6%	-0.90	-0.31	-1.84
School 47		230	16.5%	0.19	-0.09	-2.07
School 48		333	34.8%	2.96	-0.27	-2.56
School 49		328	26.2%	4.77	-0.39	-2.53
School 50		115	20.0%	0.23	-0.13	-2.23
School 51		297	8.8%	-3.60	-0.14	-1.89
School 52		70	12.9%	2.45	-0.19	-2.35

As reported in Table 7, there were 3,901 students who were actually retained in the student data, and 2,890 students were predicted to be retained based on the final model. Out of the total 18,811 students, our model predicted 854 students to be retained while they were not (Group C), and predicted 1,865 students to not be retained while they were actually retained (Group B). The classification agreement is 85.5%t—these two methods gave the same classification for 16,092 students (Groups A and D).

Table 7 also presents students' mean 2000-2001 and 2001-2002 reading and mathematics scores by the groupings defined above. For the 2000-2001 test scores, Group C and D students were similar to each other; their test scores are identical or within one point difference. They had the lowest scores, as compared to the other two groups. Group A had the highest scores and Group B students scored above both Groups C and D students.

Table 7
Mean 2002 NCE Scores and Standard Deviations for Retention Subgroups (N=18,811)

		Not predicted to be retained		Predicted to be retained	
2002 NCE scores		Not retained (Group A) N=14,056	Actually Retained (Group B) N=1,865	Not retained (Group C) N=854	Actually retained (Group D) N=2,036
Reading 2001		39.7	31.6	25.0	25.8
		(15.9)	(13.7)	(12.0)	(12.5)
Math 2001		49.8	40.9	34.0	34.0
		(16.7)	(12.6)	(10.6)	(11.2)
Reading 2002	*	38.5	35.6	23.2	29.3
		(17.5)	(15.1)	(13.0)	(14.0)
Math 2002	*	46.3	42.7	32.9	36.9
		(17.1)	(14.2)	(11.2)	(12.2)

*Post Hoc tests show significant differences between any pair of mean scores among the four categories at a joint level .05.

For the 2001-2002 achievement scores, the similarities between Groups C and D disappeared. Group C students had the lowest scores in both reading and mathematics. The next higher group was Group D, followed by students in Group B. Students in Group A had the highest 2002 SAT9 test scores. Based on the post hoc tests in analysis of variance, we found that the four group differences for any pair were statistically significant at a joint statistical significance level of 0.05.

In summary, the statistical model predicted 2,890 students to be retained while 2,036 of them were actually retained. Compared to these 2,036 students, the other 854 students who were not retained but were predicted to be retained had almost identical reading and mathematics scores in 2000-2001 and significantly lower test scores in 2001-2002 than those 2,036 students. Besides the 2,036 students who were predicted to be retained and actually were retained, there were another 1,865 retained students who were not predicted to be retained by our model. This group of students scored about a half standard deviation higher than those 2,036 students in both 2000-2001 and 2001-2002. They also scored about a half standard deviation higher than those 854 students in 2000-2001 and almost one standard deviation higher than those 854 students in 2001-2002.

In other words, among students who were predicted to be retained, those who were actually retained (Group D) scored higher than those who were not (Group C). Among students who were not predicted to be retained, those who were not retained (Group A) scored higher scores than those who were actually retained (Group B). These findings indicate that the promotion decision based on our final estimation model is better in classifying students than the current policy that is in place at the school district, using students' next year achievement scores as the measure.

Summary and Conclusion

Similar to previous research (Dauber et. al., 1993; Gottfredson et al., 1994; Hauser et al., 2000; McCoy & Reynolds, 1999; Meisels & Liaw, 1993; Owings & Magliaro, 1998; Reynolds, 1992; Thomas et al., 1997), factors of being male, being Hispanic and black, and having lower academic performance scores increase a student's chance of being retained, controlling for other variables. A unique finding of the current study is that students with lower SES status (as represented by free lunch status) were less likely to be retained after controlling for other background and score variables. This could be due to sample characteristics, as we did not find any significant relation between student lunch program and student retention in the larger school district data (see Appendix A and its tables).

The main purpose of this study is to find out whether the students' retention classification based on a statistical model provided a more accurate prediction of how well the students would do in the later school achievement tests than the current promotion/retention policy classification used at the school district. The statistical prediction of students' promotion/retention status was derived from the estimation that consists of student background variables and their previous school achievement scores. We found an 85.5% overall agreement in student classification between what was predicted and what actually occurred in the school district.

Even though there is a large degree of agreement between the classifications based on these two different retention decision approaches, the agreement number alone did not provide solid evidence of whether one is better than the other in classifying students. To decide which one is more accurate, we followed up with comparing students' achievement scores in the next year, 2001-2002. We found that, compared to these 2,036 students who were actually retained and who were predicted to be retained, the 854 students who were not retained, but were predicted to be retained, had significantly lower test scores in 2001-2002. There was one finding, however, that did not aligned with our previous claim. The group of 1,865 students who were retained, but were predicted not to be retained by our model, scored about a half standard deviation higher than those 2,036 students in 2001-2002, and they scored almost one standard deviation higher than those 854 students in 2001-2002. These findings, based on students' later school achievement scores in reading and mathematics, point to the district's or schools' misclassification of students – not retaining students who should be retained and retaining students who should not be retained.

We were also interested in investigating whether there was any school variation in the decision of promotion or retention decisions. We found that students were not retained according to a uniform standard across schools and there was a wide range of school variations. The average school retention probability ranged from 1.6% to 37.1%. This is the one unique finding of this study. Our data indicate that schools with less credentialed teachers and schools with lower social economic status (as indicated by percentage of students receiving free/reduced fee lunch) tend to retain more students. This might be due to a lack of resources for alternatives to improve student achievement, such as reading intervention, direct instruction, cognitive behavioral modification, and systematic formative evaluation as suggested by Jimerson (2001). While studies in the field mainly focus on the contribution of student background and retention, it is important to investigate the role of the school in grade retention. Hauser et al. (2000) found that grade retention rate is different across geographic locations, after student background is under control. We found in this study that in a large urban school district, students are retained differently based on school characteristics. Further qualitative study investigating the cause to these differences will contribute to a better understanding of grade retention.

Overall, our results suggest that the model-driven decision on grade retention was more accurate in predicting how well students would do in the next year's school achievement tests compared with the current district student promotion policy. Of course, school achievement was not the only measure used in the real life decisions regarding students' retention and promotion. As we have also found, school characteristics played significant roles in student retention decisions and this finding increased the complication of looking at the local and individual decisions with a measuring tape. This study, however, definitely indicates the possibility of student misclassification and the wide school variation in student retention related to both explained and unexplained school characteristics. School districts may incorporate this model-based approach into their current promotion and retention decision-making as an additional measure to improve the precision of their decisions on which students to promote and which students to retain.

References

Bianchi, Suzanne. M. (1984). Children's progress through school: A research note. *Sociology of Education*, 57(3), 184-192.

Byrnes, & Yamamoto, (1986). Views on grade repetition. *Journal of Research and Development in Education, 20(1),* 14-20.

Cadigan, D., Entwisle, D.R., Alexander, K.L., & Pallas, A.M. (1988). First-grade retention among low-achieving students: A search for significant predictors. *Merrill-Palmer Quarterly*, 34(1), 71-88.

Center for Policy Research in Education (1990). *Repeating grades in school: current practice and research evidence.* CPRE Policy Briefs.

Dauber, S.L., Alexander, K.L., & Entwisle, D.R. (1993). Characteristics of retainees and early precursors of retention in grade: Who is held back? *Merrill-Palmer Quarterly*, 39(3), 326-343.

Gottfredson, D.C., Fink, C.M., & Graham, N. (1994). Grade retention and problem behavior. *American Educational Research Journal*, 31(4), 761-784.

Grissom, J.B., & Shepard, L.A. (1989). Structural Equation Modeling of Retention and Overage Effects on Dropping Out of School. Shepard & M.L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp.16-33). New York: Falmer.

Hauser, R.M., Pager, D. I., & Simmons, S.J. (August 2000). *Race-ethnicity, social background, and grade retention*. Paper presented at the Annual Meeting of the American Sociological Association: Washington, DC.

Holmes, C.T. (1989). Grade level retention effects: A meta-analysis of research studies. In L.A. Shepard & M.L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp.16-33). New York: Falmer.

Holmes, C.T., & Matthews, K.M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*, 225-236.

Jackson, G. (1975). The research evidence on the effects of grade retention. *Review of Educational Research,* 45, 613-635.

Jimerson, S. R. (2001). A synthesis of grade retention research: Looking backward and moving forward. The California School Psychologist, 6, 47-59.

Jimerson, S.R., & Schuder, M.R. (1996, June). *Is grade retention an appropriate academic intervention? Longitudinal data provide further insights.* Paper presented at Head Start's Third National Research Conference, Washington, D.C.

Kreft, I. & de Leeuw, J. (1998). *Introducing Multilevel Modeling.* London: Sage.

Long, J. Scott. *Regression models for categorical and limited dependent variables* (1997). Sage Publications: Thousand Oaks.

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology*, 37(3), 273-298.

Meisels, S. J., & Liaw, F.R. (1993). Failure in grade: Do retained students catch up? *Journal of Educational Research*, 87(2), 69-77.

National Association of School Psychologists (2003). Position Statement on Student Grade Retention and Social Promotion. Retrieved September 14, 2005 electronically from http://www.nasponline.org.

Office of University Communication, University of Pennsylvania (2005). Grade Retention. Retrieved September 14, 2005 electronically from http://www.upenn.edu/pennnews/sourcesheet.php?id=37.

Owings, W. A., & Magilaro, S. (1998). Grade retention: A history of failure. *Educational Leadership*, 56(1), 86-88.

Peterson, S. E., DeGracie, J. S., & Ayabe, C. R. (1987). A longitudinal study of the effects of retention/promotion on academic achievement. *American Educational Research Journal,* 24(1), 107-118.

Raudenbush, S.W. & Bryk, A.S. (1988). Methodological advances in studying effects of classrooms and schools on student learning. *Review of Research in Education,* 1988-1989 edition, 15, 423-476.

Raudenbush, S.W., & Bryk, A, S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage Publications: Thousand Oaks.

Reynolds, A.J. (1992). Grade retention and school adjustment: An explanatory analysis. *Educational Evaluation and Policy Analysis,* 14(2), 101-121.

Roderick, M. (1994). Grade retention and school dropout: Investigating the association. *American Educational Research Journal,* 32(4), 729-759.

Rumberger, R. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal*, 32(3), 583-625.

Rumberger, R., & Larson, K. (1998). Student Mobility and the Increased Risk of High School Dropout. *American Journal of Education*, 107 (1), 1-35.

Temple, J.A., Reynolds, A. J., & Miedel, W. T. (1998). Can Early Intervention Prevent High School Dropout? Evidence from the Chicago Child-Parent Centers. (Report No. DP-1180-98)

Thomas, S., Sammons, P., & Smees, R. (1997). Differential secondary school effectiveness: Comparing the performance of different pupil groups. *British Educational Research Journal*, 23, 451-487.

Yuan, X. S., Pernici, S.C., Franklin, B.J. (2001). Grade level retention rates in Louisiana public schools: 1997-98 to 2000-01. Retrieved April 29, 2003, from the ERIC database.

APPENDIX A

Testing the hierarchical model using a larger data

The data used in the main text may not be a representative sample for the district as it only included non-ELL students and ELL students at the 'advanced' stage of their English Language Development. To validate our findings, we repeated the analyses using the same model on an overall sample of 24,648 students from the district by relaxing the requirement of having taken PA. The purpose of using the overall sample is to see whether the statistical model would also provide a more accurate classification than the current promotion policy for the general population of students as well as the selected sample of students.

Table A1 has the student distribution information for the overall sample by the demographic, school, and family variables used in the study. Results indicate that by including ELL students at lower levels of English Language Development, the overall sample has slightly larger proportions of male students, Title 1 students, immigrants, ELL and former ELL students, non-gifted students, and students who received free or reduced lunch. It also has a slightly larger proportion of students whose home languages are not English and students who are Black or Hispanic.

Table A2 reports the cross-tabulation of students' retention status by their background variables. We will skip the results here since the findings are similar to what we found for the selected sample, as explained in the descriptive results section in the main text.

Table A3 presents students' mean 2000-2001 GPA, reading, mathematics and PA scores along with their 2001-2002 reading and mathematics scores by students' retention status. All the performance scores are a little lower in the overall sample than in the selected sample. As in the selected sample, retained students had significantly lower mean scores in every aspect studied than students who were not retained. The gap between retained and non-retained students in reading and mathematics total NCE scores is a little narrower in 2001-2002 than in 2000-2001, which suggests that retention, in general, improved students' performance.

The procedure of the analysis for the overall sample is exactly the same as that for the selected sample in the main body. Results are summarized in Table A4. Table A4 shows the estimation of fixed effects at both student and school levels. The student level results are similar to what we obtained with the selected sample except that the coefficient for free lunch is now not significant, and the coefficient for special education is significant in the analysis of overall sample. At the school level, percentage of students receiving free/reduced fee lunch and school size are no longer significant predictors of retention.

(Please contact the authors for tables mentioned here.)

APPENDIX B

Description and results of the other models explored

To measure the magnitude of variation between schools in grade retention, we first estimate a model with no predictors at either level.

The level-1 model is:

Log-odds of being retained $\eta_{ij} = \beta_{0j}$ (1a)

where β_{0j} is the mean value for the log odds of being retained for school j.

The level-2 model is:

$\beta_{0j} = \gamma_{00} + u_{0j}$ (1b)

where γ_{00} is the grand-mean outcome (log odds of retention) across the population, and u_{0j} is the random effect associated with school j and is assumed to have a mean of zero and variance of τ_{00}. Here τ_{00} is the variance between schools in their average log odds of retention.

The estimated γ_{00} is -1.39 (se = 0.075) and τ_{00} is 0.404 (se = 0.084). The corresponding odds is 0.25 and the probability is 20% ($\varphi_{ij} = \exp(\eta_{ij}) / (1 + \exp(\eta_{ij})) = 0.25/(1 + 0.25) = 0.20$). This indicates that for a school with a "typical" retention rate, that is, for a school with a random effect $u_{0j} = 0$, the probability of retention is 0.20. Since the school variance τ_{00} is significant, we conclude that schools have different effects on their students' retention status and proceed to run our hierarchical models.

In the next estimation, we added predictors into the level-1 model. The intercept is allowed to as we found in the previous analysis that schools differ in students' retention. The variables we added are student background and achievement variables, please refer to Table 1 for the detailed descriptive information on them. The level-1 model is as the following:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}\,\text{FEMALE} + \beta_{2j}\,\text{ASIAN} + \beta_{3j}\,\text{BLACK} + \beta_{4j}\,\text{HISPANIC}$$
$$+ \beta_{5j}\,\text{ETH\_O THER} + \beta_{6j}\,\text{IMMIGRANT} + \beta_{7j}\,\text{SPANISH}$$
$$+ \beta_{8j}\,\text{HL\_OTHER} + \beta_{9j}\,\text{FREE-LUNCH} + \beta_{10j}\,\text{ELL} + \beta_{11j}\,\text{F\_ELL}$$
$$+ \beta_{12j}\,\text{GIFTED} + \beta_{13j}\,\text{TITLE 1} + \beta_{14j}\,\text{SPED} + \beta_{15j}\,\text{READING}$$
$$+ \beta_{16j}\,\text{MATHEMATICS} + \beta_{17j}\,\text{GPA} + \beta_{18j}\,\text{PA} \quad (2a)$$

The Level-2 model is consisted of the followings:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2b)$$

$$\beta_{1j} = \gamma_{10} \quad (2b1)$$

$$\beta_{2j} = \gamma_{20} \quad (2b2)$$

$$\beta_{3j} = \gamma_{30} \quad (2b3)$$

$$\beta_{4j} = \gamma_{40} \quad (2b4)$$

$$\beta_{5j} = \gamma_{50} \quad (2b5)$$

$$\beta_{6j} = \gamma_{60} \quad (2b6)$$

$$\beta_{7j} = \gamma_{70} \quad (2b7)$$

$$\beta_{8j} = \gamma_{80} \quad (2b8)$$

$$\beta_{9j} = \gamma_{90} \quad (2b9)$$

$$\beta_{10j} = \gamma_{100} \quad (2b10)$$

$$\beta_{11j} = \gamma_{110} \quad (2b11)$$

$$\beta_{12j} = \gamma_{120} \quad (2b12)$$

$$\beta_{13j} = \gamma_{130} \quad (2b13)$$

$$\beta_{14j} = \gamma_{140} \quad (2b14)$$

$$\beta_{15j} = \gamma_{150} \quad (2b15)$$

$$\beta_{16j} = \gamma_{160} \quad (2b16)$$

$$\beta_{17j} = \gamma_{170} \quad (2b17)$$

$$\beta_{18j} = \gamma_{180} \quad (2b18)$$

where γ_{10} to γ_{180} are the estimated coefficients for the student level variables listed in model 2a, and u_{0j} is the unique effect of school j on the Level-1 intercept. The values of γ_{10} to γ_{180} give the estimates of change in the respective log-odds given one unit change in the predictors.

In the model specification above, we had all 18 student-level variables have the same effects on student retention status across all schools, while we let the intercept vary across schools so each school has school-specific base student retention possibility.

It is to assume that the impact of student level variables is the same independent of schools. However, research findings have increasingly found contradicting evidence. In the final mode, we relaxed that assumption and chose to let student lunch program status and GPA scores have school-specific effect, based on previous literature.