

Individual, School, and District Effects on Reading Achievement Of 8th Grade Illinois Students, 2002, 2003

Prof. Charles Cappell
Prof. Harvey Smith
Illinois Interactive Report Card Project
Northern Illinois University
Version: April 9, 2005

Abstract

This paper reports the results of an initial 3-level hierarchical regression analysis of 8th grade reading scores of 145,336 students in 1,409 schools from 787 Illinois school districts during the 2002-03 school year. Eighty-six percent of the variation resides at the individual student level, 10.5% at the school level, and 3.1% at the district level. Organizational school level variables, class size, minutes on English instruction, and school size have opposite effects predicted by resource theory. However, district level variables point to the importance of expenditures, especially teacher salary and more highly educated staffs on producing higher scores, net of the individual and school level contributions. The results point to classroom resources as the key policy variables.

1. Introduction

Even as federal and state legislation mandates greater school accountability, and links school level student performance to incentives and penalties, research on school effectiveness produces ambiguous results regarding key variables like class size, teacher experience and education levels, teaching salaries, school size and composition. Most states are now generating comprehensive test scores and school effectiveness measures for public consumption, the state Report Cards. This paper reports an initial analysis of data recently released by the Illinois State Board of Education (ISBE) to the Illinois Interactive Report Card Project, directed by Harvey Smith at Northern Illinois University.ⁱ The data make it possible to examine several hypotheses about how the variance of student performance is distributed across individuals within the same school, across schools, and across districts. The bureaucratically generated measures do not include all of the useful measures at the individual level contained in comprehensive national surveys, such as the National Educational Longitudinal Survey (NELS), but basic demographic information about students is included. Furthermore, the state level data contains dozens of indicators of school and district level effects not available in national surveys. The comprehensive coverage of all schools and districts makes these data an interesting source to examine the effects of several important organizational variables.

Resource Theory

Most analysis of school effectiveness partition the list of organizational variables into those that are organizationally under the control of school authorities and those that are compositional, more due to the location of the school and its constituencies. We review previous studies by focusing on the variables related to school resources, those under administrative control. Compositional effects are usually found to be more influential than the effects from resource variables, but school related policies can't directly manipulate these variables. For this reason, our initial analysis is based on a small set of organizational resource variables. We will shortly extend this analysis to include compositional variables.

Perhaps the greatest controversy over the effect of resource variables involves spending levels. A recent review concludes: "Statistical evidence and recent historical experience suggest to me that school performance is unlikely to be improved solely by investing extra money in the nation's school. Increased spending on school inputs without any change in the current arrangements for managing schools offers little promise of improving either student performance or adult earnings." (Burtless, 1996, P. 41). Findings on this matter are ambiguous, as Burtless admits. Hanushek is generally unimpressed by the studies that assess achievement as a function

of expenditures (Hanushek, 1996); while others, employing meta-analysis on the same studies, conclude that many of the resource variables can be assumed to have positive effects (Hedges, Lane and Greenwald, 1993; Hedges, and Greenwald, 1996). Most ambiguous are the effects due to pupil-teacher ratios or class-size. In a school level analysis of Illinois schools using ordinary regression techniques, Sutton and Soderstrain (1999) find positive effect for pupil-teacher ratio, teacher salaries, among other resource variables, but these effects wash out when compositional variables are added to the equation. In an anecdotal review of the results from increasing spending on 15 Austin Texas schools by \$300,000 in one year, Murnane (1996) concludes that no change student performance resulted in 13 of the schools because only class-size was manipulated. When more dramatic changes to the pedagogical delivery of courses were incorporated by the two other schools, improvement occurred.

More specifically, various teacher characteristics have been examined. Level of education as measured by degree earned may not be as important as training in the content area being taught (Raudenbush, Rowan, and Cheong, 1993). Not much evidence has been found supporting the argument that matching a student and teacher by race, gender and ethnicity improves student performance (Ehrenber, Goldhaber, and Brewer, 1995). In an analysis that tries to control for student innate ability, school, and cohort effects, Rivkin, Hanushek, and Kain (2002) find evidence of specific teacher effects in performance, although they are unable with the data used to identify which characteristics of the teachers contributes to the results.

More rigorous tests of the nature of the school's organization and culture likewise have produced ambiguous results. Communal organizations are hypothesized by resource theorists to produce higher achievement because more of the resources of the school personnel are made collectively available. Academic press variables, the standards and norms for academic high achievement consciously produced by the school or district, are also hypothesized to affect achievement levels. In two large scale investigations of the effect of communalism, little evidence for improved achievement was found (Phillips, 1997; Yang, 2000). Yet the effects of academic press hold up (Phillips, 1997).

Most teacher salaries are set at the district level, and that is how Illinois reports this characteristic. Variation at the district level has been noted to affect how other organizational variables affect student performance in addition to the context itself (Friedkin and Necochea (1988); Hanaway and Talber, 1993; Witte and Walsh, 1990). These studies highlight the dramatic differences between urban, suburban, and rural school districts in the overall educational milieu, with many variables interacting with and confounding each other. Yet in Witte and Walsh's study of over 200 Milwaukee area schools, no consistent effects of school effectiveness or teacher control variables were found; the most consistent predictor was the poverty level of the school, as measured by percent in the free lunch program (Witte and Walsh, 1990, p. 206).

One of the major methodological complications in many of the previous research articles concerns the use of aggregated data to a single level, usually the school. There are serious problems with such approaches. This paper examines many of the same core variables considered in resource theory and uses a 3 level, multi-level model to decompose the variance of scores and estimate effects at each level. In general, resource theory predicts that all of the following organizational level variables we examine will have positive effects on student performance: at the district level – per pupil expenditures, average teacher salary, and percent teachers with advanced degrees; at the school level – minutes per day on English instruction, and perhaps school size. At the school level, the following variables should have a negative relationship with student achievement - class size, percent of teachers not highly qualified, and perhaps school size. School size may reflect greater resources, but it may also reflect greater bureaucratization accompanied by less attention to individual students.

2. Data and Variables

The data for this analysis were obtained from the Illinois State Board of Education. Each year the Board assesses all 2, 3, 5, 6, 8, and 11th graders in the core subjects of reading, math, science, English, social studies, and writingⁱⁱ. The data aggregated at the school and district level

is posted at the Board's website. We obtained in addition the individual level file, without student identification numbers, but with school and district identifiers which made it possible to estimate a 3 level model.

At the individual level, this paper analyzes the 8th grade reading score. The distribution of this dependent variable is presented in Figure 1. The mean score is 156 with a standard deviation of 12.31

The individual level file contains very limited information about the student, just their gender, race, whether they're enrolled in the free lunch program, and whether they're classified as a migrant. Free lunch status is a proxy for family poverty level. The file contains information whether the student is classified with Limited English Proficiency (LEP), but we consider this a diagnostic response variable to the student's previously assessed language ability and not a possible cause of reading achievement. In the raw student level data file, some information for 151,513 8th graders was included. Due to listwise deletion of missing values at the individual and higher levels, the analysis was based on 147,912 8th graders.

Descriptive statistics for district level variables for all school districts are presented in Table 1. In Illinois, in 2002-2003, there were a total of 891 school districts, 789 school districts that contained a school with an 8th grade (the remainder were high school districts). One school district was deleted due to missing values on at least one variable, so this analysis used 788 school districts. We include in the 3rd level model the set of administrative variables relevant to resource explanations of achievement that are not included at the school level: IEPP, the instructional expenditure per pupil, the average teacher salary for the district, and the percent of teachers with an M.A. or higher degree.

We note that the descriptive statistics for these variables show that there is variation of these characteristics across districts. The percent of teachers with M.A. or higher degrees has the most relative variation, with a coefficient of variation equal to 49.8, indicating a standard deviation about half the size of the mean. The relative variation of per pupil expenditure is 23.9; that of average teacher salary, 21.1. The inter-quartile range of teacher salaries shows that 25% of the districts have average teacher salaries below \$38, 207, while 25% exceed \$49,694. This difference of over \$11,000 between the top and bottom quarter districts is substantively appreciable and would seem to translate into quality of life differences for teachers in the respective districts, even when adjusted for local cost of living. A large difference like this also could generate quality differences, providing salaries are used as incentives to attract better teachers and not as compensation for poor working conditions. The variance in the percentage of teachers with M.A. or higher degrees is also appreciable: the lowest quarter of districts average 23.4% or lower, the top quarter of districts have at least half of their faculty with higher degrees. Similarly, the difference of over \$1,200 in per pupil expenditure between the 25th and 75th quartiles translates into a substantial difference. We also incorporated a few additional variables from census files aggregated at the level of school district, median household income, percent whites below the poverty level, percent unemployed, and percents in the professional and educational occupations.

(insert Table 1 about here)

School level variables are presented in Table 2. The I.S.B.E. file contained information on 3,919 schools, but only 1,446 schools have an 8th grade. When schools were deleted that had missing values on any of the variables, the number available for analysis decreased to 1,410. Again, in this paper, we limit the analysis to a handful of resource variables hypothesized to relate to academic performance from the full list in Table 2. Variables that reflect demographic school composition will be included in a later analysis. Overall school enrollment (SEN211A) shows an average school enrollment in Illinois of 161 students (for schools with 8th grade classes, average size is 182), with appreciable variation across schools. The average 8th grade class size (SACS8) was 22.9; the average number of minutes spent in 8th grade on English was 93.7. Both of these variables show variation across schools. The fourth and last school level variable included in this analysis was the percent of teachers not highly-qualified to teach (SCNHQT). The average across all schools was 2.0%; with a standard deviation of 5.6 producing a corresponding coefficient of variation equal to 285.2. This indicates a distribution with a subset of schools where the percentages are higher and a vast majority of schools with very few unqualified teachers.

(insert Table 2 about here)

3. Results

The analytical strategy began by estimating a baseline model, a fully unconditional, intercepts only model. This result breakdowns the total variance in individual 8th grade reading test scores into the portion due to individual variation of individuals within schools, the portion due to the variation among schools, and the portion due to variation among districts. Next, the individual level equation was specified by adding four individual characteristics sequentially, an indicator of whether the student was non-white (1) v. white (0), whether the student was female (1) v. male (0), whether the student participated in the free lunch program (1) or not (0), and whether the student was classified as a migrant (1) or not (0). After each variable was assessed, an additional model was estimated that allowed the coefficient to be random. An assessment was made based on the reliability coefficient and the variance component chi-square test whether to fix the coefficient or estimate it as a random coefficient at both the school and district levels.

Once a final individual level equation was established, school level variables were entered sequentially to estimate their effect on the expected intercept in the individual level equation, an estimate of the expected score of a white male, non-migrant, not on a lunch program at that school. This estimate is adjusted for the individual level effects by the level one equation.

At the third step of the analysis, three district level variables were entered to estimate the expected mean score at the district level for schools with average 8th grade class sizes, average minutes spent on English instruction per day, average percentage of non-highly qualified teachers, and average school enrollments.

(insert Table 3 about here)

Model Results

The baseline model is a simple, fully unconditional random components model. The three equations below express the basic components model.

Fully Unconditional Model
Individual Student Level Model

$$ISATR8_{isd} = \pi_{0sd} + \varepsilon_{isd}$$

School Level Model

$$\pi_0 = \beta_{00} + r_{0sd}$$

District Level Model

$$\beta_{00} = \gamma_{000} + u_{00d}$$

Composite Fully Unconditional Model for ISATR8

$$ISATR8_{isd} = \gamma_{000} + u_{00d} + r_{0sd} + \varepsilon_{isd}$$

The results from estimating this model appear in Table 3, Model 1. We can see that of the total initial individual variation in 8th grade reading test scores, 86.18% of the variation remains at the individual student level within schools ($\rho_i = 126.59 / 146.89 = .8618$). Another 10.66% of the variation is attributable to variance across schools within districts ($\rho_s = 15.66 / 146.89 = .1066$). The smallest amount of variance, 3.16%, is attributable to variation across districts ($\rho_d = 4.64 / 146.89 = .0316$). The reliability estimates (.849 for the individual intercept – school mean, .190 for district means) and the chi-square test that the variance components are zero point to a rejection of the null hypothesis of no higher level variance.

Substantively, we emphasize that this variation aggregates potentially meaningful variance at the teacher and classroom level. Even with this lack of information, the first finding we would like to highlight is that well over 85% of the variability in reading scores across Illinois occurs at the most proximate levels surrounding the students: themselves, their classrooms and

their teachers. This basic result highlights repeated findings that most of the variation in academic achievement occurs at the most proximate learning environments, the individual and one's immediate learning context. As widely disparate as Illinois schools are, most of the variation is at this more local level within schools. Still, there remains significant variation at the school and district level to learn what, if any, school and district characteristics are useful in explaining the distribution of test scores. These values indicate that a multilevel analysis is required for the random intercept coefficients.

(Insert Table 3 about here)

Individual Student Level Models

The four individual level explanatory variables were introduced sequentially in models 2-5, Table 3.

Each of the variables explains a significant portion of the individual level variation. Tests of whether the individual level ρ coefficients are fixed or random led us to fix the coefficients for the variables indicating female and migrant status; the non-white and free lunch program indicator coefficients were made random. This result is presented as model 5 in Table 3. Including these 4 individual level variables explains 8.1% of the individual level reading score variance. Poverty, as indicated by participation in the free lunch program, produces the largest substantive disadvantage in reading score performance, 5.68 points or nearly half (.46) a standard deviation in reading score. Conditioned on other variables in the model, females are expected to score 3.62 points higher than males. We note that the conditional racial gap between non-whites and whites decreases from -4.90 to -3.07 by including indicators for sex, poverty, migration status, and noting its random variation. Thus, the non-white v. white gap is reduced by over a third (37.3%) with just these three additional indicators. The effects of all these statuses on unequal reading score performance are consistent with expectations derived from previous research and accepted working hypotheses that explain these variations.

The intercept term represents the expected reading score conditioned on the following statuses: white, male, non-migrant and poverty status. This coefficient retains significant variation indicating that there remains substantial and significant variation in the conditional averages of reading scores across schools, un-modeled sources of inequality that remain after taking into account major sources of individual variation. We also note that the two random coefficients estimated for non-white and poverty status continue to have significant variation across schools as well. This indicates that the effects of non-white and poverty status vary across schools.

School Level Models

The basic individual level equation was re-estimated in the next set of models (6-9) that sequentially included the administrative school variables of interest in this analysis. We focus in this section on using these school level variables to explain the individual level equation intercept, here interpreted as the expected score of a male, non-migrant, not in the free lunch program for each school. Average 8th grade class size (SACS8) is introduced first in model 6. A negative coefficient is expected, decreasing class-size is expected to raise performance. Instead a positive (.069), and statistically significant effect (p -value $<.001$), is estimated for class-size. Our analysis of state-wide data, unfortunately, contributes to the ambiguity of findings regarding the effects of class-size on performance. This positive result pertains to the conditional reference group: white, non-migrant males not in lunch programs. In fact, this group may be less responsive to class size, and if so in a positive way. Alternative model specifications that change the reference group will be examined in a subsequent analysis. There may be other sources of confounding as well, measuring an explanatory effect from a cross-sectional survey, or mis-specification error that fails to control for an omitted variable that can alter or mask this effect. We just speculate of these issues here, but plan subsequent analysis to wrestle with this issue.

(insert Table 4 about here)

We expect the level of reading test scores to increase in schools that devote more minutes to English instruction (SENGL8). This variable is added in model 7. Again expectations are confounded, a negative coefficient of $-.01$ is estimated, translating to a decrease of one test score point for every 100 minutes of instruction added per day, hopefully an unlikely reality, but also a substantially small effect. Our speculation on this finding involves a possible misspecification of the model using cross-sectional data. If students with poorer reading ability at

time t-1 are clustered in particular schools – as we know they are, then the schools may use information from the dependent variable to assign the value of an independent variable, the length of time spent in English, at a later time, t. This would confound any estimate of the effect of time on reading estimated from a single time point.

The percent of teachers not highly-qualified in a school should have a negative effect on reading test performance. In this case, the estimated coefficient is consistent with expectations, -.07 (model 8 in Table 4). However, this substantively small effect also has a relatively large estimated standard error, .04 and a corresponding p-value of .08. Even for the conditional reference group of white male, non-poverty and non-migrant students, the aggregated effects of not-highly qualified teachers negatively impacts reading score performance.

The overall school enrollment has a significant positive effect on average school reading test scores for this conditional reference group of students. Previous research has been ambiguous about this effect as well. If school size is picking up some unmeasured resource variables, this positive result indicates that larger schools do benefit the average performance of this reference group indicated by the intercept and determined by the individual level equation. This effect, .0018, is significant at the .038 level.

Thus, in trying to explain inequalities in expected school performance for this reference group, only one characteristic, the effect of non-qualified teachers, has an effect consistent with working hypotheses. All of the effects are relatively small, but not likely to represent a fuller population where the effects are zero.

Overall, the four school level organizational variables explain 2.3% of the variation in the conditional school means left over from the individual level equations $((12.73 - 12.44)/12.44)$, not an appreciable amount. In this paper, we focus only on the administrative resource variables under the control of school officials, and not the composite, or contextual variables, related to student characteristics that are likely to have even larger effects. That tract will be pursued in a subsequent analysis.

With respect to the results that counter working hypotheses about the effects of class size, school size, and time on English, we speculate that cross sectional studies, regardless of how rich in information, such as the data we are working with, will misspecify causal effects because time is excluded. These school level variables, rather than representing pure administrative resource variables experienced by a totally randomly assigned students, are operating in an administrative context that is anticipating test score results. If schools have already isolated the poor readers into smaller classes, class size will have a positive effect, especially by the subgroup of students referenced by the individual level intercept, school mean: white, male, non-migrant, non-poverty students. The same logic applies to the number of minutes allocated to English instruction.

District Level Models

Three administrative district level attributes are introduced sequentially in models 10-14 (Table 4). The amount of instructional expenditures per pupil (IEPP) is found to have a significant positive effect on explaining the variation in district level reading scores, even after adjusting for individual and school level effects (Model 10). This effect is not huge; it would take a substantial investment to raise district average test scores appreciably. If we used these results to argue for increased expenditures, the equations implied in model 10 imply that to raise district's conditional average tests one half a standard deviation (6.16) for our reference group of schools: those having average school size, English class size, and English instruction time, and zero non-qualified teachers, 6.16 points, a district would have to increase per pupil spending by \$7,080, an unfeasible amount. But according to this result, increased spending at a global district level does increase test performance of the district.

Insert Table 4 about here

The percentage of teachers in the district with M.A. or higher degrees was entered into model 11. Even though this is a crude measure of teacher quality, it is shown to have a significant and positive impact on district average reading test scores (.058). The impact of more highly educated teachers seems more appreciable and policy relevant than the impact for global spending.

The third district level variable, average teacher salary, was also examined separately

(model 12). This expenditure has a statistically significant and positive effect, but again one, like global per pupil expenditures, substantively small and unfeasible. According to this estimate, raising a district's average teacher salary by \$10,000 would raise the district average reading test score by 1.15 points, less than .10 standard deviations (.093). Each district level resource measure showed a statistically significant, yet substantially small effect in the expected direction of working hypotheses.

In Models 13 and 14, we learn only teacher salary retains its statistical significance when multiple district level independent variables are entered. This is largely due to colinearity, the IEPP is a composite measure of instructional expenditures of which salaries are a major component, and salaries are determined on graded scales of longevity and post-graduate credit hours. But the results of this panel of models reveals that average teacher salary is the most robust effect of all of the district level variables considered. Effects attributable to global per pupil spending and the percentage of teachers with advanced degrees overlap with the teacher salary measure. By adding these district level variables, one also concludes that the variation of expected conditional district reading scores across districts is no longer random. The estimate of the district level variance component is 1.69 at the district level with a p-value of .200 in Model 13. The district level resource variable, average teacher salary, makes the district level model deterministic.

One can also note the effect of adding the district level variables has on the significance levels of the school level variables, all now fail to reach .05 significance. We also note that after one introduces the district level equations, the sign of the coefficient for school enrollment becomes negative, indicating that after one controls for the effects of district level resource variables on district level means, larger schools produce lower reading test scores for our reference group of students (p-value = .17, however). The effect of having more non-highly qualified teachers in the school continues to have a similar negative effect on reading test performance, but now with a p-value = .08. The school level effects of class size and minutes spent on English instruction remain counter-intuitive with respect to their sign.

4. Conclusions

This initial analysis of data generated by the Illinois Board of Higher Education permits a few conclusions regarding what can be learned about student performance from such data. It is straightforward and uncontroversial to use this analysis to attribute the variation in performance to different levels, the student, the school, and the district. The analysis estimated that 86.2% of the variation in 8th grade reading scores appears at the individual student level (embedded in unspecified classroom and teacher contexts). A respectable amount of variation (10.7%) in reading test scores was attributable to variation across schools within districts. The least amount of variation in scores was attributable to variation across districts (3.16%).

At the individual level sex, race, poverty, and migrant status indicators accounted for 8.1% of the variance in the reading scores, individual based inequality explained. Furthermore, we saw that the racial gap in performance was attenuated by including poverty and migrant status. Many more individual level characteristics are known to be associated with student performance, but the administratively generated data used in this analysis includes only this very limited set.

Since many of the policy incentives and penalties for student performance are leveled at schools, it is important to identify what, if any, school level characteristics account for unequal performance in 8th grade reading. This paper, in its current version, only inspected the set of school level variables measuring school resources and inputs: average 8th grade class size, average number of minutes spent on 8th grade English instruction, percent of teachers not highly qualified, and school size. This initial foray into school level effects produced somewhat counter-intuitive results with respect to class size and minutes spent on instruction, although it is also apparent that other investigation on these effects leads to ambiguous results. At the school level, the percent of teachers not highly qualified behaves as theory and common sense imply: higher percentages of unqualified teachers lowers performance.

This teacher effect is highlighted and reinforced by the district level equation explaining the variance in district level means. While all measures of spending and teacher quality at the

district level are positively related to higher average district 8th grade reading scores, average district teacher salaries emerges as the most powerful predictor of inequality at the district level, controlling for and conditional upon the models specified at the school and individual levels. Thus, at this stage of the analysis, for one bottom line conclusion, we emphasize the likely effects of teacher quality, as measured by salary, on achievement.

This analysis needs to be extended, hopefully, in the very near future, by including additional school level variables that focus on compositional effects. There is much more to be learned about the school level effects, both organizational and compositional, on reading test performance. Very little of this variation was explained by the first set of organizational variables specified in the models. The planned analysis can also estimate the higher level organizational effects on the two random coefficients at the individual level, the racial gap and poverty indicators. The results of this paper demonstrate that there is significant variation of how poverty and race influence test scores across schools.

We think this paper demonstrates the usefulness of the multi-level modeling approach for analyzing the yearly assessment data generated now by nearly all state boards of education. But our more important conclusion emphasizes that policies based on statistical evidence, causal models and imputation, require time dependent measures. Information about the most proximate levels relevant to children's education: their own characteristics, their classroom's, and their teachers are needed to specify the causal models sound policy should be based upon. The No Child Left Behind Act and state legislatures have created a test and measurement machine that generates data faster than it can be analyzed. And yet without time tracking the characteristics of students, teachers, and classes in a manner that protects the privacy and professional rights of all concerned, models will be misspecified and, if used to inform policy, could lead to misdirected efforts. With a relatively slight modification in the production of the data, highly valid insights can be gleaned.

References

- Burtless, Gary, Ed. 1996. *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington DC: Brookings Institution Press.
- _____. 1996. "Introduction and summary." Pp. 1-42 in Gary Burtless (Ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington DC: Brookings Institution Press.
- Ehrenberg, Ronald., Daniel D. Goldhaber, and Dominic J. Brewer. 1995. "Do teachers' race, gender, and ethnicity matter? Evidence from the national educational longitudinal study of 1988." *Industrial and Labor Relations Review* 48:547-561.
- Ferguson, Ronald F. 1991. "Paying for public education: New evidence on how and why money matters." *Harvard Journal on Legislation* 28:465-98.
- Friedkin, N.E. & and J. Necochea. 1988. "School system size and performance: A contingency perspective." *Educational Evaluation and Policy Analysis* 10:237-249.
- Hanaway, Jane, and Joan E. Talber. 1993. "Bringing context into effective schools research: Urban-suburban differences." *Educational Administration Quarterly* 29:164-186.
- Hanushek, Eric, A. 1996. "School resources and student performance." PP. 43-73 in Gary Burtless, Ed. *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington DC: Brookings Institution Press.
- Hedges, Larry, V, Richard Lane, and Rob Greenwald. 1993. "Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes." *Educational Researcher* 23:5-14.
- Hedges, Larry, V., and Rob Greenwald. 1996. "Have times changed? The relations between school resources and student performance." Pp. 74-92 in Gary Burtless, Ed. *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington DC: Brookings Institution Press.
- Kane, D.O., and T.J. Staiger. 2002. "The promise and pitfalls of using imprecise school accountability measures." *Journal of Economic Perspectives* 16:91-114.
- Murnane, richard J., and Frank Levy. 1996. "Evidence from fifteen schools in austin, Texas." Pp. 93-96 in Gary Burtless, Ed. *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington DC: Brookings Institution Press.
- Phillips, Meredith. 1997. "What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school." *American Educational Research Journal* 34:633-662.
- Raudenbush, Stephen W, Brian Rowan, and Yuk Fai Cheong. 1993. "Higher order instructional goals in secondary schools: Class, teacher, and school influences." *American Educational Research Journal* 30:523-553.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2002. "[Teachers, Schools, and Academic Achievement](http://edpro.stanford.edu/eah/papers/basic.july2002.PDF)". Retrieved July 31, 2004.
- Sutton, Alice and Irina Soderstrom. 1999. "Predicting elementary and secondary school achievement with school-related and demographic factors." *The Journal of Educational Research*. 92: 330-338.
- Tobias, Justin L. 2004. "Assessing assessments of school performance: The case of California." *The American Statistician* 58:55-63.
- Witte, J.F., and D.J. Walsh. 1990. A systematic test of the effective schools model." *Educational Evaluation and Policy Analysis* 12:188-212.
- Yang, Jung-Ho. 2000. "The Effects of School Community on Students' Academic Learning Growth: A Multilevel Analysis of NELS:88 for High Schools." *Dissertation Abstracts International, A: The Humanities and Social Sciences, 2000*. 61:1794-A.

Table 1: Summary Statistics for District Level Variables

Variable Name	Variable Label	Districts Reporting	Mean	Median	Stan. Dev	25% - 75%	Coef. Var.
DW	DISTRICT - WHITE %	890	83.8	94.8	23.3	79.1-98.1	27.8
DB	DISTRICT - BLACK %	890	7.9	1.2	18.2	.5 -4.1	230.3
DH	DISTRICT - HISPANIC %	890	6.0	1.7	10.9	.4 -6.5	181.6
DA	DISTRICT - ASIAN %	890	2.1	.6	4.7	.2 -1.7	221.2
DN	DISTRICT - NATIVE AMERICAN %	890	.1	0	.3	0 - .2	197.5
DENR	DISTRICT TOTAL ENROLLMENT	890	2277.5	917	14,515	452 -1895	637.3
PDLOIN	LOW-INCOME DISTRICT %	890	26.3	22.7	18.9	10.0 -35.0	74.7
PDINV	PARENTAL INVOLVEMENT DISTRICT %	890	96.4	98.3	5.6	95.1 -100	5.8
TRUANTD	CHRONIC TRUANTS RATE DISTRICT %	890	1.5	.5	3.0	0 -1.8	196.7
DACS8	AVG CLASS SIZE - DISTRICT (GR8)	413	20.9	21.3	5.3	17.8 -24.8	25.4
DENGL8	MIN PER DAY ENGL (GR8) DISTRICT	414	91.1	90.0	19.3	82.0 -99.0	21.2
DPWT	% WHITE TEACH - DISTRICT	893	96.8	100	9.1	97.7 -100	9.4
DTOTTCH	TOTAL TEACH FTE COUNT- DISTRICT	893	143.3	62.0	837.2	34.0 -124	584.3
DAVTE	AVG TEACH EXP. - DISTRICT	893	14.5	14.5	2.7	12.7 -16.3	18.9
DMAD	MA+ DEGREE - DISTRICT	893	36.7	34.6	18.3	23.4 -49.7	49.8
ELPTRDT	PUPIL-TEACHER RATIO -ELEM ALL(DIST TYPE)	789	18.3	18.7	.4	17.9 -18.7	2.18
ELPTRDS	PUPIL-TEACHER RATIO -ELEM ALL(DIST SIZE)	789	16.9	17.2	1.7	17.1 -17.2	10.2
DAVTSAL	AVG TEACHER SALARY DISTRICT	893	44,671	42,830	9405.7	38207 -49694	21.1
DCNHQT	% CLASSES NOT TAUGHT BY HIGHLY QUALIFIED TEACHERS	890	.78	0	3.5	0 - 0	445.2
IEPP	2001-02 INSTRUCT EXPEND PER PUPIL DISTRICT	893	4487.9	4254	1072.3	3785 -4918	23.9
AREADD8A	ALL GR8 READ DISTRICT ACADEMIC WARNING	396	.3	0	.8	0 - .1	254
EREADD8A	ALL GR8 READ DISTRICT EXCEEDS	396	10.4	9.0	8.3	4.3 -14.4	79.6

Notes: 25%-75% are the values of the variable that include the middle 50% of cases. The Coefficient of Variation is a crude measure of inequality; it is the ratio of the standard deviation to the mean multiplied by 100, thus it is the amount of variation as a percentage of the mean. Higher values indicate higher variation relative to the mean, hence higher inequality.

Table 2: Summary Statistics for School Level Variables – 8th Grade Schools

Variable Name	Variable Label	Schools Reporting	Mean	Median	Stan. Dev	25% - 75%	Coef. Var.
SW	SCHOOL - WHITE %	3919	65.9	82.9	36.1	38 – 96.5	54.8
SB	SCHOOL - BLACK %	3919	19.00	2.9	31.1	.8 – 21.9	163.8
SH	SCHOOL - HISPANIC %	3919	12.3	2.9	21.7	.6 – 11.8	176.4
SA	SCHOOL - ASIAN %	3919	2.7	0.7	5.5	0 – 2.5	204.3
SN	SCHOOL - NATIVE AMERICAN %	3919	0.1	0	0.4	0 - .2	245.2
PSLEP	L.E.P. SCHOOL %	3919	4.6	0	10.6	0 – 3.1	228.7
PSLOIN	LOW-INCOME SCHOOL %	3919	37.1	28.5	30.4	11.9 – 56.5	82.1
PSPINV	PARENTAL INVOLVEMENT SCHOOL %	3919	96.4	00.0	7.6	96.0 - 100	7.9
PSATTR	ALL ATTENDANCE RATE SCHOOL %	3919	94.4	95.1	2.7	93.9 – 95.8	2.9
PSMOBI	MOBILITY RATE SCHOOL %	3919	16.9	13.9	13.2	8.5 – 22.0	77.9
TRUANTS	CHRONIC TRUANTS RATE SCHOOL %	3919	1.7	0.3	4.0	0 – 1.6	239.1
PSDROT	DROPOUT RATE SCHOOL %	666	4.9	2.9	7.7	1.5 – 5.1	155.9
SACS8	AVG CLASS SIZE - SCHOOL (GR8)	1416	22.9	23.2	5.6	19.0 – 26.8	24.6
SENGL8	MIN PER DAY ENGL (GR8) SCHOOL	1418	93.7	92.0	21.8	84.0 - 103	23.3
SCNHQT	% CLASSES NOT TAUGHT BY HIGHLY QUALIFIED TEACHERS(SCHOOL)	3918	2.0	0	5.6	0 - 0	285.2
STEPC	% TCHRS WITH EMERGENCY OR PROVISIONAL CREDENTIALS(SCHOOL)	3891	2.0	0	4.6	0 – 2.0	226.1
SENR211A	TOTAL SCHOOL ENROLLMENT	3818	161.1	131.0	122.6	77.0 - 208	76.1
MECP2S03	2003-COMPOSITE 2003 SCHOOL COMPOSITE PERCENT FOR MEETS STANDARDS	83	81.3	83.7	10.4	77.5 – 88.2	12.7
AREADS8A	ALL GR8 READ SCHOOL ACADEMIC WARNING	1393	0.5	0	1.1	0 - .4	227.6
BREADS8A	ALL GR8 READ SCHOOL BELOW	1393	38.1	37.2	16.6	26.3 – 50.0	43.4
MREADS8A	ALL GR8 READ SCHOOL MEETS	1393	53.4	55.1	12.3	45.8 – 61.9	23.1
ERADSS8A	ALL GR8 READ SCHOOL EXCEEDS	1393	8.0	5.9	8.4	1.9 – 11.4	105.2

Notes: 25%-75% are the values of the variable that include the middle 50% of cases. The Coefficient of Variation is a crude measure of inequality; it is the ratio of the standard deviation to the mean multiplied by 100, thus it is the amount of variation as a percentage of the mean. Higher values indicate higher variation relative to the mean, hence higher inequality.

Table 3: Multi-level Modeling Results Predicting 8th Grade Illinois Students' ISAT Reading Score, Individual Level Equation Specifications

Model:	1	2	3	4	5
Individual level coefficients					
Intercept π_0	156.87	157.66	155.89	156.85	156.69
NW*		-4.09	-4.07	-3.77	-3.07
FEM			3.60	3.54	3.62
FRL*				-4.96	-5.68
MIGR					-3.39
School Level Coefficients on expected school average (π_0)					
SACS8					
SENGL8					
SCNHQT					
SENR211A					
District level coefficients on expected district average (β_0)					
IEPP					
DMAD					
DAVTSAL					
Variance Components					
σ^2	126.59	122.87	119.65	116.34	116.34
$\tau^2_{\pi_0}$	15.66	12.69	12.35	12.85	12.77
$\tau^2_{\beta_0}$	4.64	3.45	3.83	2.66	2.77
deviance (par. Estimated)	1134764 (4)	1130789 (9)	1126906 (10)	1122800 (17)	1122797 (18)

*These individual level coefficients were specified as random coefficients at school and district levels in Model 2 and all subsequent models; intercepts were specified as random. All other coefficients were fixed. Model 5 explains 8.1% of the original individual level variation in scores.

- NW: Student is nonwhite (0,1=yes)
- FEM: Student is female (0,1=yes)
- FRL: Student enrolled in free lunch program (0,1=yes)
- MIGR: Student identified as a migrant (0,1=yes)
- SACS8: School's average 8th grade class size (grand mean adjusted)
- SENGL8: School's average minutes spent on 8th grade English (grand mean adjusted)
- SCHHQT: School's percent teachers not highly qualified (grand mean adjusted)
- SENR211A: School's full time enrollment (grand mean adjusted)
- IEPP: District's Instructional expenditure per pupil (grand mean adjusted)
- DMAD: District's percent of teachers with MA+ degrees (grand mean adjusted)
- DAVTSAL: District's average teacher salary (grand mean adjusted)

Table 4: Multi-level Modeling Results Predicting 8th Grade Illinois Students' ISAT Reading Score, Individual & School Level Equation Specifications

Model:	6	7	8	9
Individual level coefficients				
Intercept π_0	156.92	156.71	156.78	156.81
NW*	-3.10*	-3.09*	-3.06*	-3.08*
FEM	3.62	3.62	3.53	3.62
FRL*	-5.70*	-5.71*	-5.70*	-5.73*
MIGR	-3.37	-3.39	-3.40(1.64)	-3.37(1.78)
School Level Coefficients on expected school average (π_0)				
SACS8	.069	.068	.07	.05
SENGL8		-.01	-.01	-.01
SCNHQT			-.08	-.07
SENR211A				.0018
District level coefficients on expected district average (β_0)				
IEPP				
DMAD				
DAVTSAL				
Variance Components				
σ^2	116.34	116.34	116.32	116.33
$\tau^2_{\pi_0}$	12.73	12.55	12.39	12.44
$\tau^2_{\beta_0}$	2.55	2.56**	2.48**	2.22**
deviance (par. Estimated)	1122784 (19)	1122779 (20)	1122758 (21)	1122754 (22)

* These individual level coefficients were specified as random coefficients at school and district levels in Model 7 and all subsequent models; intercepts were specified as random. All other coefficients were fixed.

** Variance term has p-value > .05.

P-values are included in parenthesis when they are greater than .001.

- NW: Student is nonwhite (0,1=yes)
- FEM: Student is female (0,1=yes)
- FRL: Student enrolled in free lunch program (0,1=yes)
- MIGR: Student identified as a migrant (0,1=yes)
- SACS8: School's average 8th grade class size (grand mean adjusted)
- SENGL8: School's average minutes spent on 8th grade English (grand mean adjusted)
- SCHHQT: School's percent teachers not highly qualified (grand mean adjusted)
- SENR211A: School's full time enrollment (grand mean adjusted)
- IEPP: District's Instructional expenditure per pupil (grand mean adjusted)
- DMAD: District's percent of teachers with MA+ degrees (grand mean adjusted)
- DAVTSAL: District's average teacher salary (grand mean adjusted)

Table 5: Multi-level Modeling Results Predicting 8th Grade Illinois Students' ISAT Reading Score, Individual, School, & District Level Equation Specifications

Model:	10	11	12	13	14
Individual level coefficients					
Intercept π_0	156.79	156.66	156.52	156.51	156.66
NW*	-3.13*	-3.12*	-3.17*	-3.17*	-3.17*
FEM	3.62	3.62	3.62	3.62	3.62
FRL*	-5.71*	-5.72*	-5.74*	-5.73	-5.73*
MIGR	-3.33	-3.33	-3.30	-3.30	-3.30
School Level Coefficients on expected school average (π_0)					
SACS8	.07	.038	.028(.02)	.027(.019)	.030(.019)
SENGL8	-.01	-.005(.019)	-.009(.005)	-.008(.005)	-.008(.005)
SCNHQT	-.078(.04)	-.078(.044)	-.078(.044)	-.077(.044)	-.079(.044)
SENR211A	.001(.0008)	.001(.0007)	.001(.0008)	-.001(.0008)	.001(.001)
District level coefficients on expected district average (β_0)					
IEPP	.00087	--	--	--	.000181(.00018)
DMAD		.058	--	.014(.011)	.01(.01)
DAVTSAL			.000115	.000133	.000118
Variance Components					
σ^2	116.34	116.34	116.34	116.35	116.35
$\tau^2_{\pi_0}$	12.16	11.61	11.41	11.14	11.37
$\tau^2_{\beta_0}$	1.50**	1.73**	1.45**	1.69**	1.35**
deviance (par. Estimated)	1122717 (23)	1122709 (23)	1122677 (23)	1122675 (24)	1122674 (25)

* These individual level coefficients were specified as random coefficients at school and district levels in Model 7 and all subsequent models; intercepts were specified as random. All other coefficients were fixed.

** Variance term has p-value > .05.

Standard errors are included in parenthesis when p-value > .05.

- NW: Student is nonwhite (0,1=yes)
- FEM: Student is female (0,1=yes)
- FRL: Student enrolled in free lunch program (0,1=yes)
- MIGR: Student identified as a migrant (0,1=yes)
- SACS8: School's average 8th grade class size (grand mean adjusted)
- SENGL8: School's average minutes spent on 8th grade English (grand mean adjusted)
- SCHHQT: School's percent teachers not highly qualified (grand mean adjusted)
- SENR211A: School's full time enrollment (grand mean adjusted)
- IEPP: District's Instructional expenditure per pupil (grand mean adjusted)
- DMAD: District's percent of teachers with MA+ degrees (grand mean adjusted)
- DAVTSAL: District's average teacher salary (grand mean adjusted)

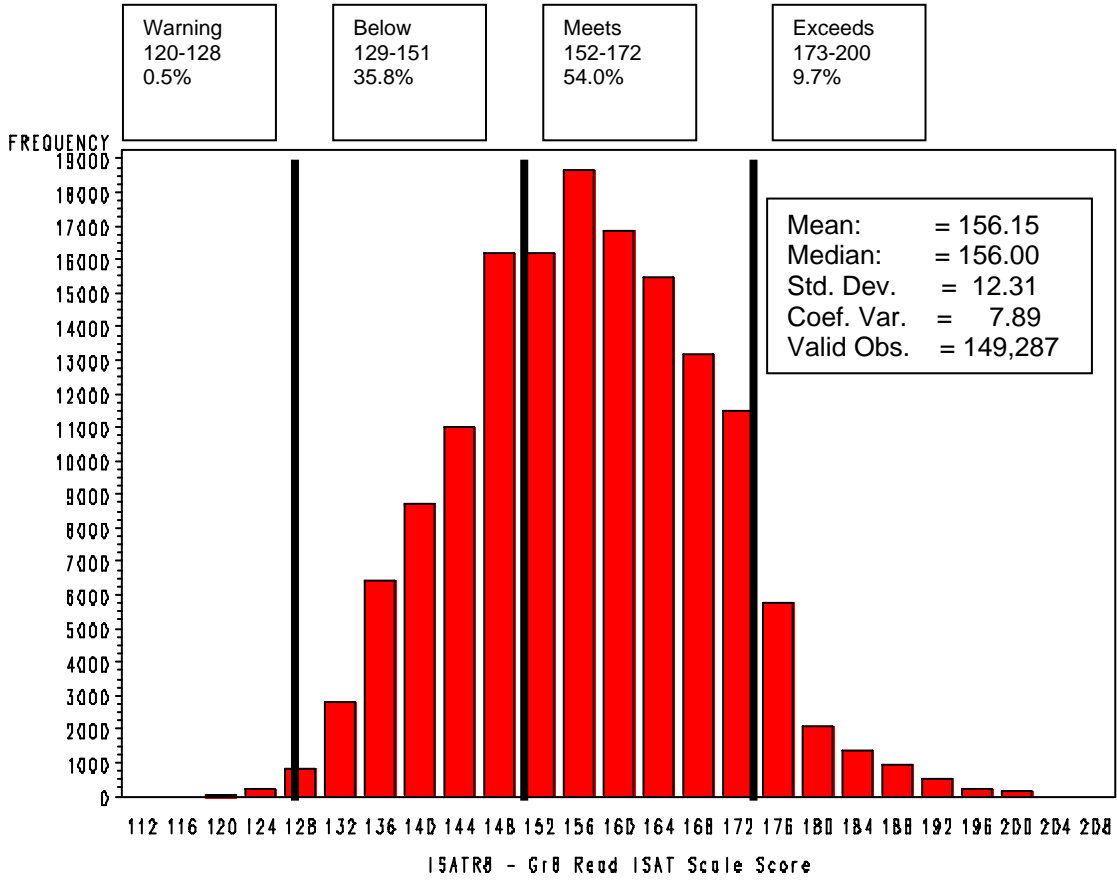


Figure 1: Distribution of 8th Grade ISAT Reading Scale Test Scores, 2002-03

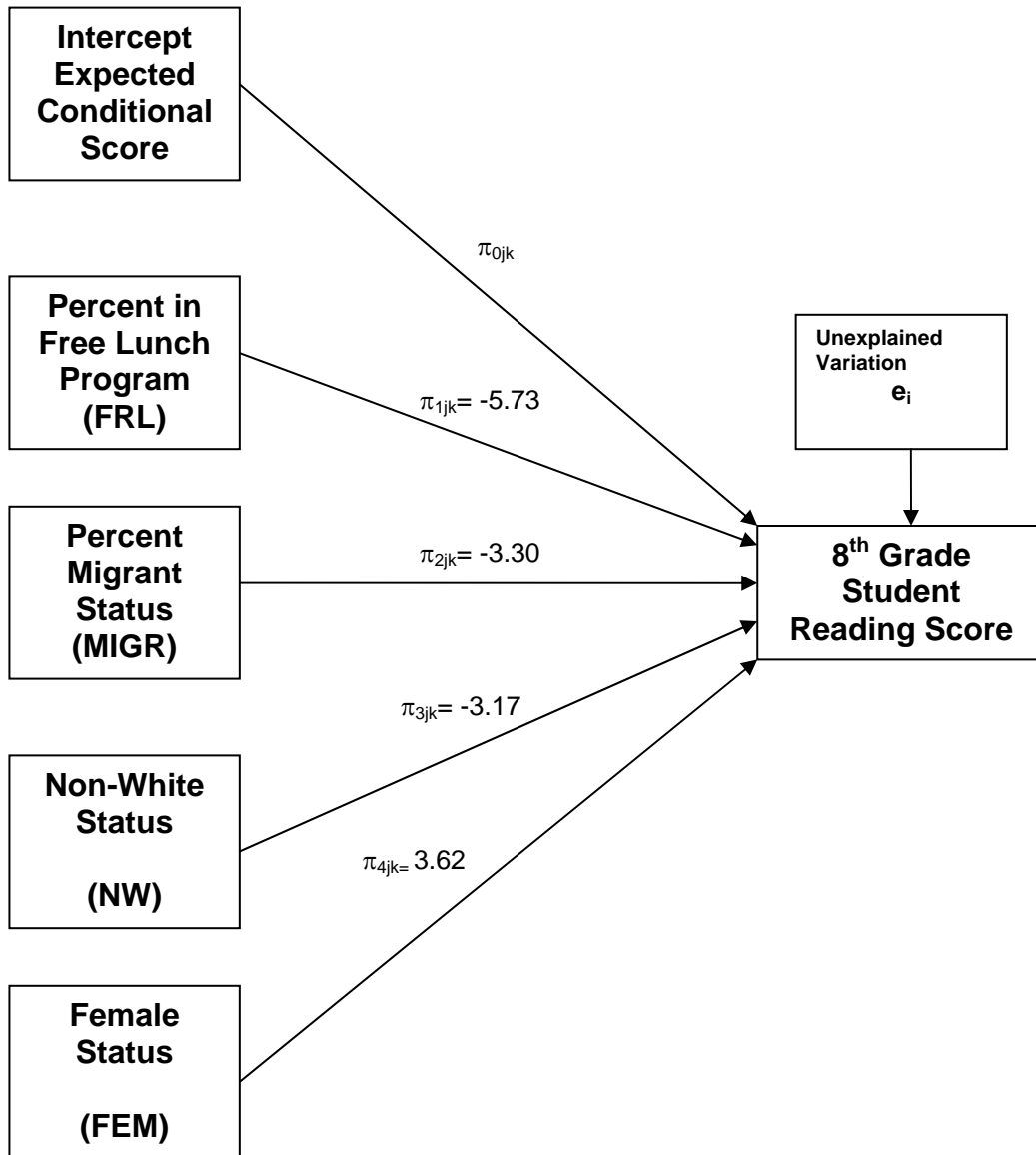


Figure 2: Individual Level Effects on 8th Grade Reading Scores
(Estimates from Model 13)

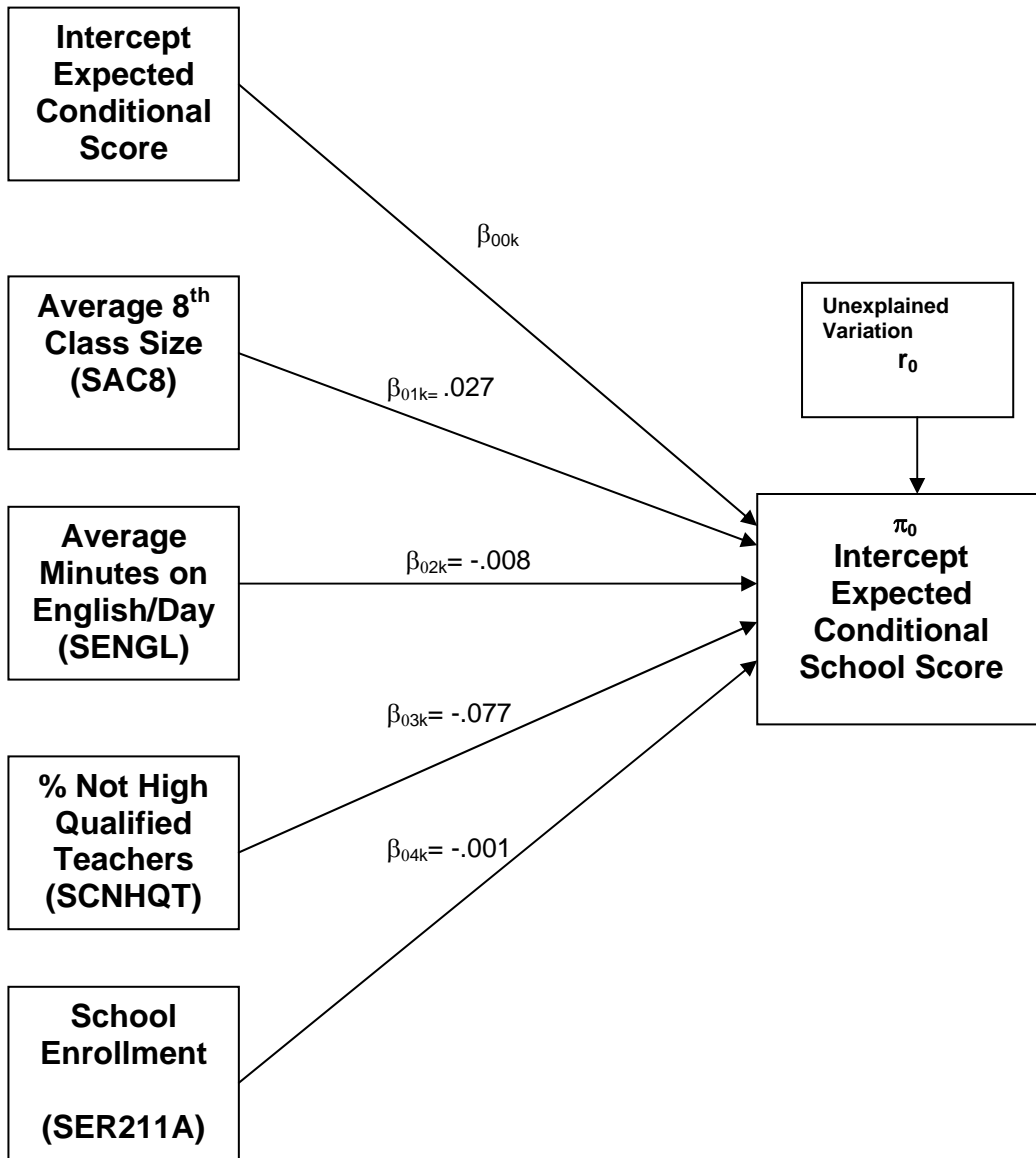


Figure 3: School Level Effects on School's Expected 8th Grade Reading Scores (Estimates from Model 13)

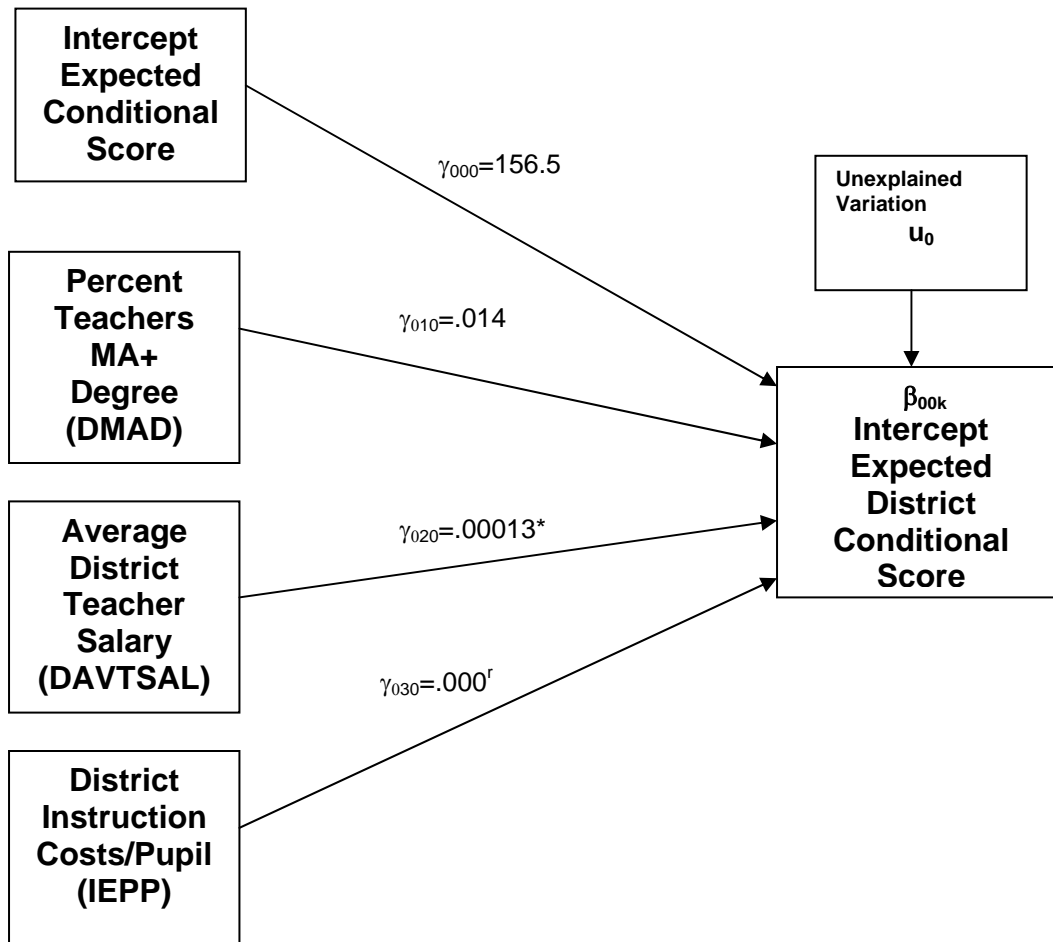


Figure 4: District Level Effects on District's Expected 8th Grade Reading Scores (Estimates from Model 13)

Notes:

ⁱ The I.S.B.E. has the right of prepublication review of this article as a condition allowing us to analyze the complete assessment file. All the analysis, opinions, and conclusions expressed herein are solely those of the university based authors.

ⁱⁱ In early 2005, the Illinois General Assembly voted to eliminate testing in writing and social studies for two reasons, budget shortfalls, and the tests were not mandated in the No Child Left Behind Act.