Testing the Impact of Student Demographics and Educational Input Variables on Standardized Achievement: A Statewide Approach

Abstract

This paper offers a comprehensive approach to the study of the socioeconomic, racial, school finance, and other important building-level input variables on standardized student achievement. Hierarchical Linear Modeling was employed to capture the nested nature of these school inputs. The educational data tested within this paper included an exhaustive accounting of every public school in Missouri. The results of the study largely confirmed the decade's-long research efforts that have explored the relationship between student demographics, school finance inputs, and other impactful educational inputs with standardized achievement. In short, FRL rates and race are confirmed to be considerable determinants of student achievement.

Keywords: achievement; accountability; reform; Hierarchical Linear Modeling

Efforts to successfully reform the nation's public schools are most likely to succeed when leaders possess a more complete knowledge of what matters in the educational process. That is, before school leaders focus on what they are able to control in the instructional reform process, they must remain aware of the many factors which they cannot dictate. Ultimately, the resulting impact of instructional improvement efforts is better understood if these school leaders are cognizant of the influence of these factors upon the wider reform process. That is, building-level inputs and demographic composites presently act as either challenges or opportunities for instructional leaders who seek to positively impact student engagement levels over time.

In this vein, the paper explores the many educational inputs associated with community wealth and statewide funding. As effective efforts are based on test scores, the linkage between these measurable, largely uncontrollable variables are tested in relationship to resulting standardized achievement levels. Finally, the findings are framed according to complex mathematical models that have been designed to accommodate the complex and fluid universe of instructional reform in the modern accountability era.

Perhaps at no other time in the history of United States public education have school systems mirrored the private sector to such an indistinguishable extent in terms of accountability for bottom line results. Standardized test performance levels represent the final product to be "manufactured by" the schools in the accountability era. A problematic feature of these curricular initiatives is the extent to which "teachers view the acquisition of pedagogical

knowledge as unrelated to formal programs that seek to demonstrate or develop that knowledge" (Firestone & Pennell, 1993, p 507).

Concern exists that the demands placed upon schools to achieve adequate growth on high-stakes testing has led to perverse incentives for schools and states to artificially demonstrate academic progress (Linn, Baker, & Betebenner, 2002). As Linn, Baker, and Betebenner (2002) note, a potential unintended consequence of the NCLB legislation may be that state and school education leaders manipulate the standardized testing instruments to inflate test pass rates. Doing so is to the detriment of the students within these same states that the laws were crafted to benefit (Linn, Baker, & Betebenner, 2002). Though teaching methods and resulting classroom quality are likely linked closely to test scores, it becomes important to test the input variables that can be measured in relationship to achievement levels. This provides instructional leaders with an up-front knowledge of the instructional hurdles that they will face on the basis of these factors, many of which they can address, but not directly control.

Review of Literature

1a: Funding and other Key Inputs

The conventional wisdom in educational finance circles is that enhancing the resources allocated to public schools will immediately exact heighted standardized achievement levels. Conflicting data on the topic aside, the empirical question has been rendered moot in an era where nondiscretionary school funds are no longer scarce but are now nonexistent. Fortunately, promising evidence demonstrates that a proportional relationship exists between educational objectives at the building level and student achievement, a factor entirely devoid of pecuniary considerations (Hanushek, 1995). Along similar lines, the IPI is grounded in the expectation of

educational excellence and pedagogical best practices that foster classroom environments which prepare all students to excel. Testing whether such a treatment, which is almost wholly devoid of pecuniary circumstances, is effective in enhancing higher-order student engagement levels across classrooms is an important methodological question. Such a finding can then lay the groundwork for additional empirical inquiry as to whether a student engagement-student achievement relationship can be evidenced to exist. While both research questions alone represent intriguing empirical propositions, they also enable findings that can pragmatically inform school leaders of practices and processes that benefit students in terms of their learning experiences and their performance abilities as they sit to take standardized tests.

An underlying assumption of the school reform movement is that a fundamental alteration of a school's operating practices can affect the quality of a school's educational provision. It is not uncommon for researchers to find that 75% of the school level variance rests outside the control of schools, however, as socioeconomic factors typically dwarf other school performance variables (Heck, 2001). This seems to suggest that school leaders' abilities to dictate test performance might be largely uncontrollable. Indeed, Rumberger and Palardy (2005) stress that the most influential school input, the characteristics of the student body, was the least equitably distributed among schools.

The impressive influence that socioeconomic demographics exhibit on school performance can compel student leaders to undertake more nefarious attempts to demonstrate enhanced standardized test performance. More specifically, school leaders may sometimes consciously make organizational decisions to expel certain students from their populations (Amrein & Berliner, 2003; Rumberger & Palardy, 2005). It is not uncommon for school leaders to be befallen by the suspicion that they "doctored" their performance numbers, an accusation

that appears to be corroborated by studies that find a pronounced discrepancy between high and low stakes testing performance despite reports of rapidly increasing high-stakes test scores (Nozawa et al., 2007). A most notable example of this can be found in a 2002 study conducted by Jacobs, which discovered that high stakes test gains did not translate into low stakes test gains among a large cohort of Chicago public school students (cited in Nozawa et al., 2007).

Gains on standardized test-taking abilities might not be transferrable to real world applications, of course (Nozawa et al., 2007). Nevertheless, schools dedicated to reforming and improving their practices spend inordinate amounts of time and effort preparing their students for state tests. For instance, schools frequently offer incentives to students as inducements to perform well on these tests (Raymond, 2008). Ultimately, the pressure to demonstrate test score growth can exact a toll on school leaders. Nozawa et al. (2007) find that those teachers who perceive standardized test performance to be of greater importance also reported feeling more professional pressure (Nozawa et al., 2007). As teachers are subjected to increasing levels of organizational stress, this could conceivably diminish the quality of their instruction (Griffith, 2004).

1b: Testing Funding and Other Key Inputs

Unresolved to date is the exact extent to which wealth dictates educational quality. Much more firmly established, however, is that demographic and building-level inputs associated with wealth are strongly correlated to achievement. Of interest in the study, then, is not merely the degree to which designated building-level inputs impact standardized achievement, but whether instructional improvement plans can help mitigate factors that otherwise prove to alarmingly diminish test performance levels. Schools' socioeconomic compositions, along with funding levels, must largely be treated as fixed by educational leaders. Nevertheless, empirical methods

that explore, at the statewide level, whether reforms can help dull these otherwise piercing effects on achievement is a methodological undertaking that has pressing implications for educational leaders in all the nation's public schools.

2a: Effective School Reform Efforts

The effective schools movement has been galvanized by a desire to propel schools out of the doldrums of mediocrity (Cuban, 1998). Heightened levels of teacher motivation and self-efficacy that result from effective school processes must eventually be translated into actionable change initiatives. Ultimately, the nature of school reform processes will determine the effectiveness of these school improvement initiatives. According to Rumberger and Palardy (2005), school processes include school administrators' evaluation of how their schools' inputs are organized and managed, the consideration of the practices that are used within their schools, and the climate that permeates the schools' learning environments. Some school procedures become so pervasive as to best become characterized as the "grammar of schooling" or "school regularities" (Miles & Darling-Hammond, 1998). These school processes are commonly monitored by formal evaluation mechanisms. Thornton et al. (2007) argue that one such form of operational oversight, the frequency of teacher evaluation, is not a sufficient instrument to ensure an effective learning environment for students, however. Instead, schools must focus on how such evaluative information is used by school leaders (Thornton et al., 2007).

As school leaders consider how they might organize schools to become more effective, this invariably entails the consideration of how building-level resources can be more effectively utilized (Miles & Darling-Hammond, 1998). Miles and Darling-Hammond (1998) studied a sample of effective schools and determined that such schools "demonstrate that it is possible to

support student achievement at extraordinarily high levels by managing instructional resources to maximize individual attention for students and learning time for teachers" (Miles & Darling-Hammond, 1998, p.10). As the most influential variable found in those effective schools studied by Moe and Chubb (1990) was the freedom from bureaucratic control, this finding has pressing implications for those policymakers and school leaders who claim that school finance is the foremost determinant of school performance levels (Moe & Chubb, 1990). Cuban (1998) has also written extensively on the mechanics of effective school reform efforts. Cuban (1998) derides the general school reform process, as he laments the capriciousness with which these change reforms are undertaken. School reform efforts are as dependent on the health of the culture and climate within schools as they are on financial considerations.

School leaders oftentimes find it difficult to maintain measured responses to the frenzied pressures of the No Child Left Behind (NCLB) Act. Attempts at school reform, however, require that the faculty believe that change is needed and that they plan for such change appropriately and rationally (Moe & Chubb, 1990; Witte & Walsh, 1990). The NCLB Act leaves school leaders with little choice but to fundamentally alter the nature of their educational instruction. Teacher commitment to meaningful change and the extent to which actors are willing to become actively involved in such change efforts is invariably required of school improvement and reform efforts (Leithwood, Menzies, & Jantzi, 1994). Furthermore, as educators seek to accomplish their objectives, these faculties benefit from continuous communication with one another as they work to ensure the success of their improvement efforts (Ferrara, 2007). While these school change efforts need not be incremental, they cannot be adopted in the form of a shock treatment, either. Indeed, successive approximations that build

upon previous change efforts have been demonstrated to work well (Leithwood, Clipsham, Maynes, & Baxter, 1976).

Statewide results that link the variables associated with educational practices with resulting test scores in the study are only as veracious as the models are sound. With this in mind, key input variables for all districts across the state are included to draw the fullest and most accurate connection between instructional environments and test scores. With such knowledge, school leaders can best design and anticipate the challenges associated with their reform efforts on the basis of the composite profiles of their student populations. There is further reason to believe that model misspecification will result if the school change occurs at a time other than that which researchers suspect. Were this the case, researchers would subsequently incorporate in their models assumptions that are grounded in the supposition that change attempts across grade-levels are uniformly challenging (Thum & Bhattacharya, 2001).

2b: Testing Effective Reforms

The socioeconomic composite of a student body, along with a host of other inputs that rest outside the control of school leaders, can markedly impact student achievement. Yet, these same variables are not likely to influence the overall extent of reform effectiveness as severely. Such levels of reform progress, in addition to the broader considerations of educational quality, revolve around test score performance. Naturally, then, variables that impact test scores are of interest to researchers and educational leaders. More specifically, a discussion of the educational factors that remain an everyday reality that school leaders must address, but cannot manipulate, helps to frame a realistic assessment of instructional improvement in public schools. Quantifying the extent to which uncontrollable inputs govern test score performance levels is an important first step in both planning and executing school improvement efforts. Before reform

goals can be fully designed and efficaciously incorporated at the school level, educators must acknowledge these challenges and design programs that address achievement growth on the basis of demographics and other uncontrollable input variables.

3a: Statistical Methods

Hierarchical Linear Modeling (HLM) represents an especially attractive methodology upon which to properly address the complete bundle of easily measurable but complexly interrelated variables. The structurally and spatially nested nature in which student learning and school processes are configured is duly accounted for by HLM Models. A three-level Hierarchical Linear Model can account for classroom engagement that is nested within distinctive districts and Regional Professional Development Centers (RPDCs). Exceptionally large standardized factor loadings need not be evidenced to establish significant empirical findings. Marginal increases in student achievement, at an even lesser pecuniary cost, are always a welcome prospect in public education. While the methodology itself may be of little interest to school leaders or policymakers, the interactions of the many consequential and oftentimes confounding building level variables may prove to be of far greater salience to such an audience. As not all teachers possess the same level of competence, those students situated within effective or ineffective classrooms, a determination entirely outside of their control (Waxman, Huang, Abderson, & Weinstein, 1997), can materially affect the level of such students' learning as well as their prospects for success in their future academic and professional undertakings.

3b: Testable Statistical Models for This Study

The factors that determine the standardized achievement levels across schools' student populations are varied and typically interrelated. Testing the relationship, both in terms of classroom behaviors and resulting test performance, invites a methodology that encompasses the layered universe that embodies the educational processes in today's public schools. HLM, as a statistical method that incorporates nested variables at the school, district, and region levels, allows for educational policy propositions involving the extent to which input levels influence standardized achievement outcomes to be fully tested in these more encompassing models. As a consequence, the quantifiable influence of identified input variables inform school leaders of their presence and impact on these desirable school improvement processes. Also made known is the time and ambitiousness required of the reform plans relative to school districts that are not required to contend with these variables.

Methods

Statistical Model Configurations

The structurally and spatially nested nature in which student learning and school processes are configured can be duly accounted for by HLM Modeling. The empirical consideration of the site-level variables and their contemporaneous interactions with one another in a more holistic manner is made possible according to the HLM methodology.

To adequately account for the nesting of building-level resource inputs in the greater environmental context, the introduction of a third level to the model that incorporates the region level (level three) can additionally be considered by the researcher as he attempts to account for the structure inherent in student learning. Furthermore, knowledge of more elaborately constructed HLM models enabled the researcher to immediately evaluate the proportion of

variance explained among each of these levels to determine whether a parsimonious pruning of entire levels of the HLM Models is warranted.

Level-One School Variables: Level-One of the HLM models employed in the study contained the resource variables shown by the research to impact the nature of both student learning and achievement.   The data associated with such variables that are linked with each school setting can be introduced into the multilevel statistical study at Level One of the HLM models.   Several building-level school inputs are also imbedded in this level of the HLM models. The racial composition of study populations (pct_white), teacher salary (tchr_sal), the proportion of certified teachers (cert), the proportion of students eligible for free and reduced lunch (FRL1), the student teacher ratio (stu_tchr), and the administrator teacher ratio (admin), are accounted for at Level One. Each of these variables, to varying degrees, coincide with student engagement behaviors in dictating both the nature of student learning in the classroom and on standardized tests.

Level-Two School District Variables: School districts comprise the second level of the multilevel statistical study that incorporates student engagement data from within and across classrooms. The schools that provided IPI classroom data were located within Missouri school districts in all corners of the state. While not categorically the case, anecdotal evidence and more cursory observations suggest that school districts exhibit a pronounced and inescapable influence on the health and effectiveness of the schools that operate within them.   The demographic data provided by Missouri's Department of Elementary and Secondary Education is quite exhaustive. For the purposes of this study, traditional socioeconomic, and controllable and uncontrollable educational resources and input factors were collected and recorded for the corresponding school districts containing the schools that provided data for the current research undertaking. More

specifically, the per pupil expenditure levels (Fund), the racial composition (Pct_white), the free

and reduced lunch rate at the district level (FRL), teacher salary (tchr_sal), student teacher ratio

(stu_tchr), the administrator teacher ratio (admin), the proportion of certified teachers (cert),

teacher experience, and the dropout rate of both white and black students, and the proportion of

students that pursue further education (follow) are included at Level Two. These variables, both

in isolation and acting in concert, can govern both student engagement and standardized

achievement levels with considerable impact at times.

Level-Three Regional Professional Development Center Variables: Not unlike many states

across the nation, Missouri is comprised of several disparate regions. Impoverished urban

centers in Kansas City and St. Louis are surrounded by more affluent suburban districts that

demonstrate standardized test performance levels that are reflective of these socioeconomic and

demographic endowments. In addition to the two metropolitan, suburban regions of the state are

themselves surrounded by rural regions and small towns/cities. In the technical sense, these

RPDC regions are artificial constructs that assume the form of the district averages of several

demographic and achievement variables. As many of the RDPC's across the state are represented

by several dozen districts, this district average that comprises the RPDC construct amounts to

more than a redundant demographic layer upon which to analyze by employing Hierarchical

Linear Modeling. The several districts within the study were nicely dispersed across the region,

creating averages that are statistically representative of regional demographic, controllable and

uncontrollable inputs, and student achievement. The geography and economic makeup of these

areas are disparate, provide meaningful across-region differences to be methodologically

captured. Here, the FRL rates of regions, along with their racial compositions, were included in

Level Three of each model.

Data Collection Source

The Missouri Department of Education's (DESE) Web Site served as the principal source of secondary data collection for this study (Department of Elementary and Secondary Education, 2008). School districts' and school buildings' demographic, as well as other pertinent teacher and administrator characteristics, are available from the state education department's Web Site. The availability of these data allow for the pairing of the IPI schools with non-treatment schools (which were entirely devoid of the IPI practices) schools that are representative of the typical Missouri public school (DESE, 2008).

Results

Standardized Test Proficiency – 3 Level Model

The HLM output for the three-level state model reveals a significant and impactful relationship between the racial and socioeconomic variables and standardized achievement levels. As the percentage of students who perform proficiently on Missouri's standardized MAP test serves as the state benchmark for compliance with the federal No Child Left Behind Accountability standard, that outcome variable is necessarily considered. The intercept for statewide proficiency on the Communication Arts test was 30.44 while it was found to be 34.37 for mathematics. These values do not include students who perform at an advanced level, a separate performance category that will be considered momentarily. The HLM analysis reveals that the percentage of students who receive free and reduced lunch (FRL) at the building level is negatively related to proficiency levels. Additionally, the percentage of African American students is also negatively related to proficiency at the building level, although to a lesser extent. At the district level, FRL was not found to be statistically significant in effecting the Communication Arts dependent variable, but was found to be slightly negative for Mathematics.

Finally, at the regional level (RPDC), the percentage of free and reduced lunch students (FRL) and minority students (black) evidenced a positive statistical relationship with standardized test proficiency levels for communication arts, while no statistically significant relationship was found for mathematics performance levels. While very moderate, and only evidenced on communication arts, this unexpected directional finding is not without a plausible explanation, as the racial composition of the region is preponderantly Caucasian (greater than the national average). Demonstrated, therefore, is a regional effect that would exhibit a demonstrable but inconsequential effect on student proficiency levels.

Standardized Test Advanced Performance – 3 Level Model

Investigating the relationship between the socioeconomic variables and standardized test performance is important for several reasons. Advanced performance on standardized tests can serve as evidence of instructional excellence insofar as students are able to master concepts and testing practices that will serve them well in their pursuit of postsecondary education. Additionally, advanced performance serves as an indicator of a more promising pool of eventual workforce candidates, signaling to prospective employers and corporations the nature and level of intellectual capital within a state or region. The findings from the three-level HLM model were illustrative on several counts. The slope for FRL and percentage of African American students, which appear to be slight, but are nevertheless noteworthy given the fact that only 12.31 percent of students performed advanced on Communication Arts, while an even fewer 9.75 percent did so on the mathematics component of the test. At level two of the model, only statistically significant findings were evidenced with respect to mathematics. At the third level, FRL was not statistically significant, while the percentage of African American students evidenced positive slopes. While the magnitude of this slope is more pronounced, and is in the

unexpected direction, it is again worth noting that the percentage of minority students within an RPDC is quite low and is not given to vagarious fluctuations that would make this finding disruptively unexpected.

Standardized Test Below Basic Performance – Three-Level Model

Considering the percentage of students who perform below the basic proficiency level is also an important empirical enterprise. An intercept value of 9.64 was found for communication arts and 13.43 for mathematics, but values representing considerable segments of the testing population. When considering the below basic performance as an outcome variable, an unexpected relationship emerged in the first level (school level) of the model that is quite interesting: while the FRL and black predictors were found to magnify the effects on communication arts performance levels in the expected direction, the results were found to be mixed for mathematics performance. At the very least, these numbers suggest that students' lackluster performance on standardized tests cannot be narrowly attributed to FRL. Moreover, while the racial composition of test takers exhibits a small effect on communication arts, the impact is nearly three times greater on mathematics underperformance. The district level yielded less statistically significant findings, although the district slope for FRL on mathematics achievement suggests that the FRL and racial socioeconomic indicators are robustly influential in determining the level of low performing students on mathematics, while this relationship is considerably less remarkable for communication arts.

_____

Insert Table 1 approx. here

_____

Insert Table 2 approx. here

FRL Rates – Level 1

The FRL predictor values were all found to be in the expected direction, varying

negatively with successful test performance and positively with below proficient performance for

communication arts (while it was not found to be statistically significant for below proficient

mathematics rates). It should be noted that the FRL predictor slopes for advanced mathematics

and communication arts outcome variables was nearly as elevated as was the case for the FRL

predictor slopes for mathematics and communication arts proficiency. Yet, the advanced

performance rates were only one-third of the proficiency level. This demonstrates the more

heightened impact that FRL can exhibit on advanced standardized test performance levels.

Percentage White Students –Level 1

As the percentage of African Americans was considered as the racial predictor in the

prior demographic model, the percentage of white students was used here, which serves as an

internal check to ensure the appropriate directionality of the predictor. The percentage of white

students can also be employed as a predictor for the purpose of comparing the impact that this

racial group exhibits on the test performance levels as compared to their African American

counterparts.   Noteworthy is the predictor slope for the percentage of Caucasian students that

comprise the student body ("White") that is imbedded in the below basic math performance

outcome variable model. Interestingly, the magnitude of the slope is fully three times greater

than for below basic performance on communication arts, and over twice as strong as on

proficiency levels of communication Arts and mathematics. More unexpected was the finding

that the "White" predictor was statistically insignificantly related to advanced mathematics and communication arts performance.

Discipline Rates

The extent to which disciplinary infractions within schools effect standardized test performance levels was also tested. The findings were again in the expected direction. The extent to which these disciplinary rates impact below basic performance is also considerably greater: between 4-8 times as impactful (compare, for instance the discipline slope for below basic mathematics with advanced mathematics performance. A similar pattern is displayed with communication arts and mathematics proficiency outcome variables. As such, it appears that the discipline rates of Missouri public schools can be influential in predicting standardized test performance levels.

Teacher Certification

The Teacher certification predictor ("tch_cert") was also found to be in the expected direction. It is again the case that the magnitude that the teacher certification predictor manifested in relation to the below basic performance was between 4-7 times greater than was the case for advanced proficiency outcome models. While current state and federal laws preclude school districts from employing a large segment of uncertified teaching staff, these findings suggest that even a small proportion of uncertified teachers might exhibit a disproportionately undesirable impact on standardized achievement levels. More specifically, the predictor slope for advanced math outcome was more depressed, while for below basic mathematics it was a considerably more consequential.

Administrator FTE

The number of administrators within a building does not appear to impact student achievement to the extent that intuition would otherwise suggest. The mixed results fail to provide conclusive evidence that including additional administrators on the school leadership team will exact an ameliorative effect on standardized achievement levels.

Teacher Salaries

The findings for the average teacher salary predictor were also mixed; however, evidence suggests that enhancing funding by a significant magnitude could marginally enhance gains in student achievement were the funding increases highly robust. While the findings appear to suggest that enhanced funding would disproportionately impact communication arts performance, this finding was not reproduced when below basic communication arts was tested. Nevertheless, the data offer evidence of the importance of teacher salaries in reducing below basic standardized test performance.

Student Teacher Ratios

Student teacher ratio findings, when combining level one and level two predictor slopes, appear to be impactful on student achievement when these ratios are drastically manipulated. The overall slope suggests notable test score fluctuation were student-teacher ratios to fluctuate by 5 students per teacher or greater. The overall effect on communication arts was less pronounced, but would also be significant were student-teacher ratios to fluctuate by 10 ratio units or greater. The impact of student teacher ratio on below basic MAP performance was also found to be in the expected direction. Hence, for mathematics below basic performance, significant alterations in the student teacher ratios will be moderately impactful, while this would not appear to effect below basic communication arts performance levels to a similar extent.

Local Funds

Local funding, especially in the middle of a pronounced recessionary environment, also becomes of interest in the student achievement calculus. Here, the findings are relatively straightforward in terms of expected directionality, but are rather weak in magnitude. For instance, advanced mathematics and communication arts outcome variables are accompanied by positive but tempered funding slope magnitudes. Below basic performance in both subject areas is even more depressed. Hence, changing the funding mixture to more heavily favor local funding is not likely to impact student achievement levels in an appreciable manner.

Teacher Experience

Teacher experience predictor values displayed a moderately positive relationship with advanced and proficient communication arts and mathematics proficiency levels. There was not, however, the same statistically significant relationship with either of the below basic achievement outcome variables. As such, it appears that significant variations in teacher experience levels (tch_exp) will begin to exhibit demonstrable gains in standardized student achievement levels.

Dropout Levels

Student dropout levels, while influencing student achievement levels in the expected direction, were inconsequentially slight in the predictor magnitude values. Both findings suggest that radically enhanced dropouts rates would have to materialize before standardized test performance metrics would be impacted. Dropout levels for African American students, while statistically significant in some instances, evidenced magnitudes that were in many instances undetectably slight as to the impact of this occurrence on student achievement levels.

FRL in the Three Level Context

The level-three FRL predictor values were, unexpectedly, found to be reversed in directionality from Levels one and two of the HLM models. Accordingly, the impact of FRL on student achievement may be less pronounced than the Level one predictor appears to suggest. While these magnitudes very much comport with intuition, the direction of such magnitudes do not. The same goes for the communication arts below and advanced performance models.

It is important to consider these slopes in the three-level methodological context, and not simply as discrete predictors that are isolated at individual levels. Theoretical arguments can be made to weight the school level predictors more heavily, given the questionable veracity of RPDC findings due to the weak predictive power of the level. The "raw FRL" predictive power of all three level FRL predictors combined, however, reveal a less significant relationship between the population of FRL students with student achievement levels than any singular achievement level might suggest. When the additive nature of the FRL predictor inclusion in the three level model is accounted for, the slope for advanced math and communication levels becomes slightly positive. Below basic mathematics performance levels remain in the expected direction, and more robustly so, while in the communication arts models the FRL predictor is found in the unexpected direction. The key target outcome variable for NCLB compliance, math and communication arts proficiency rates of public schools, remains unchanged and impactful for mathematics, while it is considerably more diminished and trivial for communication arts.

Across-Level Variance Computations

Calculating the proportion of variance across levels, also known as computing the Intraclass Correlation Coefficient (ICC), involves the simple process dividing the calculated variance for a given level by the overall variance associated with the model. Such computations

provide a meaningful empirical context within which to compare the extent of standardized test score variance that is attributable to disparities in building, district, and regional level educational inputs. Also of interest is the extent to which the model specification at each level successfully accounts for overall observed variance at each level. By computing the difference in specified models that contain independent predictor variables with empty models that are bereft of such predictors, the extent to which these educational input predictors account for standardized test performance variance can be ascertained.

Variance Explained by the Three Level Models

        As is depicted in Tables Three through Seven below, the variance explained at level one was comparatively low in relation to the district and regional levels. Nevertheless, the variance appears to be most widely attributable to across school differences, which accounts for 79-94% of model variance. Across district variance, on the other hand, accounts for the remaining 6-21% of model variance. Given the very low to nonexistent variance explained at the RPDC level, pruning the model into two levels may be not only appropriately parsimonious but methodologically warranted.

_____

Insert Table 3 approx. here
_____

_____

Insert Table 4 approx. here
_____

————————————————

Insert Table 5 approx. here

————————————————

————————————————

Insert Table 6 approx. here

————————————————

————————————————

Insert Table 7 approx. here

————————————————

Discussion

Explanation of Variance – Three Level Model

The variance findings from the three-level model that incorporated the many educational input variables associated with that effect standardized test achievement levels were revealing: 79-94% of the overall variance exists between schools while the remaining 6-24% of variance is attributable to the district level. Such a finding is very much intuitive, as the inputs found at the school level should be most impactful on determining how that given school's population of students performs on the standardized testing instrument. More surprising, however, is the relatively insignificant predictive power of the level-one independent variables. The level-one predictors, embedded in the models to account for variance at the school building level, captured only 10-45% of the overall level one variance. The second and third levels within the models exhibited a far greater explanatory power, with the district level predictors accounting from between 34-89% of the computed variance. The regional level (Level three) exhibited very little initial variance, so the tabulated explanatory power of this level is tenuous at best. Hence, it

appears that while between-school variance predominates the models' explanation for variance distributions, the actual range with which each level varies is most aptly accounted for by the socioeconomic, demographic and other school input predictors at the district level.

A consideration of an identical three-level structure that incorporates only demographic predictors at each level (FRL rate and the percentage of African American students) can be used to conclude how much of the variance is accounted for by the socioeconomic and racial variables, as opposed to other school inputs. The level-one variance that was accounted for was considerably less robust: here the model that accounted for the most variance, 15%, is considerably less than the model that was outfitted with key input variables (45%). Accordingly, it appears that at the school level, the racial and socioeconomic composition of student populations accounts for only one third of the variance that can otherwise be explained when these predictors are accompanied by additional school finance and classroom/teacher characteristics at the building level. On the other hand, school demographic predictors were found to account for 34-74% of the variance at the district level, a sum only slightly less than for the model that was more fully specified with the "key educational input" predictors. As expected, 70-94% of the models' variance existed between schools, while the remaining variance was accounted for at the district levels. Only very slight amounts of the model variance appears to be attributable to regional differences, another finding that squared nicely with the *a priori* assumptions about the standardized test performance of students.

Practical Meaning of These Variance Findings

As expected, the greatest variation in student test performance, when considered with a host of influential predictor variables, occurs at the school building level, as opposed to between-district or regions. The more inquiring of readers might question why the Intraclass Correlation

(ICC) value is not greater at the school district (in other words, 100% of the variance explained at the building level). While these empirical findings are suggestive of the importance of the school building level's role in standardized test performance levels of students, several qualifiers are in order: 1) many school districts are balkanized to such an extent that district-level initiatives that are soundly constructed educational policies are not translated at the site level in the spirit in which they were crafted, and 2) the school building level demographics serve as a proxy for the home environments of schoolchildren. In other words, very low test performance may be far less a product of inept educational leadership, and far more ascribable to destructive home lives that deleteriously impact student achievement despite proper building-level educational practices.

Random Effects Models and their Constituent Predictors – Three-Level Models

FRL

The racial and the socioeconomic composition of schools were not treated as synonymous measures, but rather as distinctive (and distinctly testable) educational inputs. That is, neither variable was considered to be a broad, encapsulating proxy of the other. Such a finding is borne out empirically, as the free-and-reduced lunch rate of student populations within schools, districts, and regions, exhibited far less influence on standardized achievement than did the racial composition of these students. Providing some quantification of these broader empirical assertions is in order. Take first the schools' FRL rates: it appears that mathematics performance is most impacted by the FRL levels of student populations, as the predictor slope associated with the proficient outcome variables was -.15 and .08 when associated with the below basic proficiency outcome variable. Hence, were the FRL population to balloon by twenty percent within a school, below proficiency standardized test performance within a school

would increase by nearly 5 passage-rate points. Communication arts performance levels do not appear to be impacted to a similar extent, however, as below basic and proficiency outcome variables were accompanied by FRL predictor slopes of a more diminished magnitude (-.04 and -.03, respectively).

More interesting is the finding that the advanced mathematics outcome variable is positively impacted by the FRL levels, although this is slight (FRL predictor slope of .05). While these findings combine the school, district, and regional FRL predictor slopes, an investigation solely at the school building level reveals slope values which range from -.11 to -.15 with the corresponding proficient and advanced performance outcome variables. The below basic communication arts model was found to have an FRL predictor slope of only .09, however. Using communication arts performance levels as an example, the same twenty percent increase in FRL populations would yield a 4.40 decrease in the percentage of students that perform at advanced levels on the communication arts segment of the MAP test.

Race

The percentage of white students was employed as the predictor of student racial composition on standardized test performance levels. The predictor magnitude for below basic mathematics and communications arts was determined to be -.09, while slopes of .07 and .06, respectively, were evidenced for advanced communication arts and mathematics outcome variable models. These slopes, while seemingly insignificant of their face, are actually quite substantial given the percentage of white students within Missouri schools. As such, the difference between an inner-city and a rural school might well be a raw difference of 90% white students (10% white students in St. Louis, 100% white students in rural southwest Missouri). Remarkably, this translates to a 14.40 point discrepancy between urban and rural schools based

only on racial composition when considering communication arts performance

(below+advanced) and 13.50 points for mathematics performance (below + advanced).

Unquestionably, this leaves the race measurement as the most impactful input variable in

predicting student achievement levels. Hence, prior research suggesting the importance of race

on student achievement is very much validated by an exhaustive study of the racial composition

of every school within Missouri.

School Finance

Teacher salaries (tch_sal) and student-teacher ratios (stu_tchr), variables that are both

principally predicated upon school funding, appear to be considerably more influential on

student achievement than are administrator FTE and the proportion of local funds that schools

receive. More specifically, the advanced communication arts outcome variable was

accompanied by a tch_sal predictor slope of .0004, while that value was found to be .0001 for the

model in which advanced mathematics was the designated outcome variable. Such findings are

less than compelling, however, as radically augmented levels of teacher salaries would have to

be supplied to begin to see accompanying standardized achievement successes. As a result, a

$10,000 increase in teacher salaries would yield only a 4 point increase in advanced

communication arts, and a mere 1 point increase in the advanced mathematics performance rate.

Student-teacher ratios are also largely dependent upon the state's per pupil funding

formula. Mathematics performance levels were impacted to a noteworthy extent by the student-

teacher ratio within schools and their corresponding districts: the models contained predictor

slopes of -.30 for advanced mathematics and .20 for below basic mathematics. Accordingly,

every 10 unit increase in the student-teacher ratio, which is certainly very large, creates a 5 point

increase in undesirable standardized achievement performance.

Teacher Experience

Teacher experience also appears to moderately impact standardized student achievement

levels. The district level (level-two) predictors for communication arts and mathematics

proficiency were found to be .22 and .30, respectively. Nearly identical values were evidenced

for advanced communication arts and mathematics, as well (.23 and .28, respectively).

Accordingly, 4.5 percentage point increases in desirable communication arts and 5.8 percentage

point increases in mathematics performance can result from enhancing a school faculty's

experience level by an average of ten years.

For the reader's convenience, several of these key findings are reproduced in tabular form below:

_____

Insert Table 8 approx. here

_____

By way of statistical analysis, this study was designed to determine the extent to which

uncontrollable input variables, principally predicated around community wealth, impacts

achievement levels across all public schools within states.   Though instructional leaders can

undertake reform plans in an effort to bolster test performance, it is wise to first consider the

hurdles they face in undertaking such a challenge. In short, empirical findings that account for

the challenges to school improvement on the basis of building-level inputs and demographic

factors can allow for a more informed and appropriately benchmarked instructional improvement

scheme. This, in turn, positions school leaders to most appropriately enhance student learning

and standardized achievement over time.

Advanced standardized achievement is appreciably impacted by the racial and socioeconomic composition of student populations. Among the most compelling of the study's findings involved test score discrepancies linked to race alone. 50% disparities in the socioeconomic and racial compositions of student populations account for 7 point advanced math performance and 8.5 point communication arts achievement discrepancies. Also noteworthy is the determination that 50% disparities in the socioeconomic and racial compositions of student populations account for 4.5 point below basic math performance and 5.5 point below basic communication arts deviance.

Confirmed by the study, then, is the proposition that the socioeconomic and racial composition of student populations do not effect schools' achievement levels equally. From a consideration of two-level HLM modeling, FRL rates were found to be twice as impactful as race on advance Communication Arts performance but many times weaker than was race on below basic Communication Arts. 50% disparities in the FRL rate would decrease advanced communication by 5.5 points, and advanced mathematics performance by 3 points. 50% disparities in the FRL rate would increase below basic communication arts by 1 point, but a similar 50 percent disparity in race would increase below basic communication arts by 4.5 points. Results pertaining to school inputs directly impacted by school funding are mixed. Nevertheless, such inputs are clearly not capable of fundamentally transforming student performance, as 10 unit decrease in student-teacher ratios would only enhance advanced mathematics performance by 2.4% , while slightly diminishing advanced communication arts performance (-.2%).

The results of the statewide study indicate the pronounced influence of race and socioeconomics on standardized achievement levels. As the composition of student populations

can differ radically from one district to the next, this study informs school leaders with greater

quantitative exactitude of how the composition of student populations position schools relative to

surround district and state averages. Before school leaders focus their attention on reform, they

must know what obstacles they face in promoting both instructional improvement and outcomes

excellence. Working from such knowledge, these instructional leaders can then design

improvement plans and formulate goals that are appropriately exacting. Given what is now

known about the descriptive factors of student achievement levels on state testing instruments,

such information is sure to be of great help to a great many educators.

References

Amrein, A. L., & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32-38.

Cuban, L. (1998). How schools change reform: Redefining reform success and failure. *Teachers College Record, 99*(3), 453-477.

Ferrara, D. L. (2007). The school improvement and transformation system. *Education Planning,16*(1), 18-30.

Firestone, W.A. & Pennell J.R. (1993). Teacher commitment, working conditions, and differential incentive policies. *Review of Educational Research, 63*(4), 489-525

Freeman, D. J. (1989). State guidelines promoting teaching for understanding and thinking in elementary schools: A 50-state survey. *Educational Evaluation and Policy Analysis, 11*(4), 417-429.

Griffith, J. (2004). Relation of principal transformational leadership to school staff job satisfaction, staff turnover, and school performance. *Journal of Educational Administration, 42*(3), 333-356.

Hanushek, E. (1995). The quest for equalized mediocrity: School finance reform without consideration of school performance," in Lawrence O. Picus and James L. Wattenbarger (ed.). *Where does the money go? Resource allocation in elementary and secondary schools* Thousand Oaks, CA: Corwin Press, pp. 20-43.

Heck, R. H. (2001). Direct and indirect writing Assessments: Examining issues of equity and utility. *Educational Evaluation and Policy Analysis, 23*(1), 19-36.

Leithwood, K., Clipsham, J. S., Maynes, F., & Baxter, R. P. (1976). Curriculum change at the system level: A four-year mathematics project. *Curriculum Theory Network, 5*(3), 219-

245.

Leithwood, K., Menzies, T., & Jantzi, D. (1994). Earning teachers' commitment to curriculum reform. *Peabody Journal of Education, 69*(4), 38-61.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). *Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001.* Los Angeles, CA: UCLA.

Miles, K. H., & Darling-Hammond, L. (1998). Rethinking the allocation of teaching resources: Some lessons from high-performing schools. *Educational Evaluation and Policy Analysis, 20*(1), 9-29.

Missouri Department of Elementary and Secondary Education, (DESE). (2008). Website. http://dese.mo.gov

Moe, T., & Chubb, J. (1990). Letting schools work. *City Journal.* Retrieved 1/12/2008, www.city-journal.org/article01.php?aid=1632

Nozawa, Y., Waltman, K, & Lai, E.R. (2007). Discrepancies between test scores from high-stakes and low-stakes contexts: An investigation of the effects of school-level characteristics and test-preparation practices. Center for Evaluation and Assessment. University of Iowa. Presented at National Council for Measurement in Education.

Raymond, M. (2008). Paying for A's: An early exploration of student reward and incentive programs in charter schools. Center for Research on Education Outcomes. Stanford University.

Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal, 42*(1), 3-42.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 11). Thousand Oaks, CA: Sage.

Thornton, B., S. (2007). A systems approach to school improvement: Program evaluation and organizational learning. *Education, 128*(1), 48-55.

Thum, Y. M., & Bhattacharya, S. K. (2001). Detecting a change in school performance: A Bayesian analysis for multilevel join point problem. *Journal of Educational and Behavioral Statistics, 26*(4), 443-468.

Waxman, H. C., Huang, S. Y., Abderson, L., & Weinstein, T. (1997). Classroom process differences in inner-city elementary schools. *The Journal of Educational Research, 91*, 49-59.

Witte, J. F., & Walsh, D. J. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis, 12*(2), 188-212.

Zhang, Y., Shkolnik, J., & Fashola, A. (2005). Evaluating the implementation of

comprehensive school reform and its impact on growth in student achievement, *American Institutes for Research.* Washington, D.C.