Comprehensive Partitioning of Student Achievement Variance to Inform Equitable Policy Design

Michael C. Rodriguez & Kyle Nickodem
University of Minnesota

Minnesota Youth Development Research Group
www.mnydrg.com

April 2018

Paper presented at the annual meeting of the
National Council on Measurement in Education, New York, NY.

Citation:

Rodriguez, M.C., & Nickodem, K. (2018, April). *Comprehensive partitioning of student achievement variance to inform equitable policy design*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

Comprehensive Partitioning of Student Achievement Variance
to Inform Equitable Policy Design

Abstract

So that large-scale assessments positively impact teaching and learning, we must more comprehensively investigate (partition) the sources of variance in achievement test scores and evaluate assessment-accountability policy targets. We find less than 20% of variance in a state assessment is between schools (over 80% within), of which 70% is explained by school demographics. Practice and policy implications are explored.

Conceptual Framework

As the Nation's Report Card, the National Assessment of Educational Progress (NAEP, 2017) provides strong long-term trends in the academic achievement of the nation's students. The NAEP program strives to achieve the theme of this year's NCME program: To make assessment a stronger force for positive impact on teaching and learning. It does this primarily by monitoring trends, among which are trends regarding educational disparities. Through the NAEP results, we are able to evaluate progress toward educational equity, particularly regarding the role of test use in informing educational practice and public policy. The use of test results to inform practice and policy is, of course, a modest approach to the use of achievement tests from the federal perspective; whereas a more ambitious position is in the use of standards-based tests as the education form agent, which was a failure as anticipated (Koretz, 2017).

Among the many means of supporting the use of assessment information, NAEP releases periodic reports on special topics of national interest. A recent report, *School Composition and the Black-White Achievement Gap* (Bohrnstedt et al., 2015), closely examined the role of school composition in schools that participated in NAEP, with a strong focus on educational equity. Other researchers make similar cases, investigating the role of school and teaching resources and other school and community characteristics addressing educational equity (Baker, Farrie, & Sciarra, 2015; Hanushek & Woessmann, 2017; Reardon , 2015; Reardon, Kalogrides, & Shores, 2016; Whitehurst, Reeves, & Rodrigue, 2016). Figlio & Karbownik (2017) presented evidence regarding the variability in educational disparities across schools. They argued that because of variability of school success with advantaged and/or disadvantaged student groups, attention should focus on practices within individual schools. A number of other reports make similar cases, arguing for a redistribution of school funding, school resources, teaching resources, and other context variables that play a role in addressing achievement gaps (Baker, Farrie, & Sciarra, 2015). Because school composition is of great concern, particularly in urban areas, we explore the role of school composition with a statewide population-level database of student achievement.

Achievement Gaps

Typically, achievement gaps are defined simply in terms of differences in percent proficient or average score between students of color and White students. Perhaps the most common question regarding the magnitudes of achievement gaps is about the role of socio-economic status (SES; Harwell, Maeda, & Lee, 2004). Ideally, we would have a complex measure of SES, including family income, parental education level and occupation(s), and other resources. But typically we only have participation in free and reduced-priced lunch programs (FRL) as an indicator of SES. Unfortunately, we know there are limitations in this indicator, particularly for students in secondary schools, who are less likely to participate. Nevertheless, it is a powerful explanatory variable when examining variance in test performance, both at the student and the school levels.

As an example, Cotrell et al. (2015) summarized wide-ranging cognitive disparities reported in the research literature and argued that cognitive tests robustly predict job performance. They found evidence to support an established model of educational disparities, based on racially disparate conditions, including family income, maternal education level and verbal ability, availability of learning materials in the home, parenting factors, and child birth outcomes. Based on analysis of the NICHD Study of Early Child Care and Youth Development, they found:

- The Black-White gap has been studied since as least 1922.
- There is no strong theoretical basis for the cognitive ability gap.
- Cognitive tests show large Black-White differences, an average Cohen's *d* (standardized mean difference) of 1.0.

- This gap in cognitive tests is near three times as large as the gap in job performance (although there is substantial evidence supporting the use of cognitive abilities to predict job performance, racial disparities do not follow the same pattern).
- The concept of race implies a history of housing segregation, education segregation, and occupational segregation (among others). Occupational segregation results in income disparities; education segregation results in disparities in maternal education and verbal abilities; housing segregation results in educational & occupational disparities. A cyclical pattern emerges.

It is clear that race/ethnicity embodies complex characteristics that are consistently associated with academic achievement, often the result of exclusionary public policies or structural barriers reducing access to resources and opportunities to learn. Segregation is a long term result of such policies/structures and this plays an important role in understanding variance in student and school-level achievement.

The Importance of Variability

State departments of education (as a result of federal regulations) report and focus on the percent meeting standards and other performance levels; states and local education agencies report and focus on percent meeting standards and average test scores; NAEP reports and focuses on percent at each proficiency level and average test scores. However, in most schools, teachers do not work with a majority of students who achieve at the average – teachers do not target their teaching practices at the percent meeting standards. Teachers and schools (and communities) are struggling to meet the needs of diverse students – particularly students who differ in academic preparation, in academic experience, and in academic achievement. By focusing on the percent meeting standards or the average test score, we limit our understanding of the real challenge in education, which is how to meet the needs of students who vary widely in achievement (knowledge, skills, and abilities). The variability in achievement presents the greatest challenges.

In a comprehensive study of student achievement data from 49 states, the district-level ICCs ranged from near zero to .23 in ELA and mathematics, and were 29% larger in 8th grade compared to 3rd grade (Fahle & Reardon, 2017). Furthermore, the ICCs in mathematics tended to be larger than those in ELA (about 13% larger), yet the correlation between the ICCs in the two subject areas was .94. In a study of TIMMS-R 1999 data, the estimate ICC was only about 6.3% of variance was between schools (Kim & Choi, 2008). OECD (2004) estimated that about 34% of variance in student PISA mathematics scores among 15-year-olds was between schools, but this varied substantially across countries. In the USA, the between-school variance is closer to about 30%.

The Role of School Composition & Resources

Figlio & Karbownik (2017) presented evidence regarding the variability in educational disparities across schools. They argue that because of variability of school success with advantaged and/or disadvantaged student groups, attention should focus on practices in individual schools. They were not able to explain variation in school success levels regarding educational disparities based on kindergarten readiness of students or the overall level of SES within a school. Because of the inconsistency in school success with advantaged or disadvantaged students or both, they argued that accountability policies should focus on school success with specific populations.

International evidence typically illustrates the limited power of school expenditures and class size to explain variation in student achievement, pointing to measures of school quality inputs as having more potential regarding student outcomes (Hanushek & Woessmann, 2017).

In Minnesota, a series of policy shifts contributed to housing and educational segregation, particularly in the Twin Cities urban area (Orfield, 2015). In addition, MN is known for its innovation in charter school and school choice policy design. However, charters are another source of segregation, where 45 of the 50 most racially concentrated schools in the Twin Cities are charter schools (Orfield, 2017). "The evidence suggests that, under the most favorable set of assumptions for poverty academies, racial integration is more likely to produce academic benefits for nonwhite and low-income students than the creation and maintenance of segregated charter schools" (Orfield, 2017, p. 3). Nationally, charter schools are also seen as being "generally more racially and economically segregated than traditional public schools" (Whitehurst, Reeves, & Rodrigue, 2016, p.6). Differences in school racial composition are associated with meaningful (but small) differences in student achievement, but perhaps with the exception of those charter schools with strong academic focus and "no-excuses" philosophies (Whitehurst, Reeves, & Rodrigue, 2016).

Reardon (2015) conducted a number of focused studies on school segregation and racial academic achievement gaps. Among the volumes of findings, he argued that disparities in school poverty rates between White and Black students are “consistently the most powerful correlate of achievement gaps” (p. 1). This implies that high-poverty schools are less effective than low-poverty schools. In addition, there is significant variation in achievement gap magnitudes across the country – economics, demographics, segregation, and school conditions explain about three-fourths of the geographic variation in gaps; the strongest correlates being racial/ethnic differences in parental income, parental education, and racial/ethnic segregation (Reardon, Kalogrides, & Shores, 2016, p.1). Finally, Fahlea and Reardon (2017) recently reported that nearly 90% of the variation in district-level ICCs across states can be explained by variation in district racial and economic segregation and the composition of students in school districts. Moreover, “states with high levels of white-black and economic segregation have, on average, more between district variation” represented in ICCs (Fahlea & Reardon, 2017, p. 17). This should be expected, since as districts (or schools) become more homogeneous as a result of segregation, the proportion of variance due to districts will increase (larger ICCs); and paradoxically, if we adequately measure that segregation resulting in larger between-district variance, we will ultimately be able to account for it (resulting in a higher variance-accounted-for statistics).

Test Score Use

As with most state tests, the MCAs were developed in response to the requirements in the Elementary & Secondary Education Act (ESEA, initially established in 1965). ESEA was reauthorized as NCLB (2001), required standards-based testing for evaluating school quality, and imposed severe sanctions on poorly performing schools. ESEA was recently reauthorized as ESSA (2015), retaining the use of standards-based tests without punitive consequences for poorly performing schools. ESSA currently requires states to identify the bottom 5% performing schools based on academic achievement, growth, and a number of other indicators (however, being on “the list” is punitive to members of the communities of those schools, with secondary negative consequences such as home sales and enrollment). We acknowledge that the MCAs provide the key evidence of academic achievement and progress for schools. With this information, the state identifies the bottom 5% of performing schools in order to provide comprehensive supports and interventions. As required by federal regulations, the tests must be aligned to grade-level state academic content standards (standards-based tests), so that the resulting inferences can be made regarding the success of the school in providing access to state content standards (i.e., high quality educators provide high quality instruction regarding state standards and students learn them).

The Minnesota Office of the Legislative Auditor (2017) recently completed an audit of the state testing programs. They found that use of test scores at the local level varies widely, where many educators and school leaders feel unprepared to interpret most testing data they receive. Policy shifts in Minnesota have contributed to housing, employment, and educational segregation, particularly in the Twin Cities (Orfield, 2015). Moreover, there is an increasing trend toward culturally-specific charter schools, as well as several other trends with education and assessment policies. These are relevant contexts for understanding variability in achievement.

Our interest in partitioning variance in student test scores is largely due to the use of these scores as school-level success/quality indicators. The federal accountability target is school, focusing on stakes for schools, not students. We maintain that the vast majority of variance in student performance is within schools, rather than between schools, which may suggest that the accountability targets are misplaced.

Research Questions

- How much variation in student achievement is within versus between schools?
- How much variation is a function of student (within school) and school (between school) demographic characteristics?
- How much additional variation is a function of school quality indicators, including those included as options through ESSA?

Answers to these questions have implications for educational, assessment, and accountability policies.

Methods

Data Sources

Data were linked from four sources, including (a) 2014-2015 Minnesota Comprehensive Assessment and Minnesota state demographic database, (b) 2013-2014 National Center for Education Statistics School Universe Survey Common Core of Data, (c) 2011-2012 Office of Civil Rights Data Collection, and (d) 2013 Minnesota Student Survey. The variables and descriptions from each data file are listed in Table 1.

Minnesota Comprehensive Assessment (MCA). The MCAs are the state's school accountability tests for students in grades 3-8 and high school (https://education.mn.gov/MDE/fam/tests/). They are summative measures of student achievement of the Minnesota Academic Standards in mathematics, reading, and science. They are administered annually in the spring. "The State uses the aggregated test scores to report to the public and the U.S. Department of Education how Minnesota students are performing in school. Statewide test data help the State evaluate the progress schools are making in reducing achievement gaps among student groups" (MN Department of Education, 2016, p.1).

We received the statewide MCA score data file at the student-level, with information about the school of enrollment. Only students who were in a school on October 1, with MCA scores in that school, were included, as this is the inclusion rule for school accountability purposes. MCA scale scores (i.e., linearly transformed IRT scores) were used as the outcome variable. Student variables used for required student-group reporting were included as predictors in the models, such as English Language Learner, Title I, Special Education, Gender, Free/Reduced-Priced Lunch, Race/Ethnicity (American Indian, Asian/Pacific Islander, Hispanic, Black, White).

Common Core of Data (CCD). The Public Elementary/Secondary School Universe Survey CCD was obtained from the National Center for Educational Statistics (NCES; https://nces.ed.gov/ccd/). The CCD provides basic information and descriptive statistics for all schools, their students, and their teachers. Several school-level variables were obtained from the CCD for MN schools, including Full-Time-Equivalent (FTE) staff counts, Title-I status; indicators for magnet and charter schools; proportion of students receiving FRL, and city/urban/suburban location of school. In addition, "regular" public schools (School-Type 1; 67% of all schools) were included in these analyses, eliminating the inclusion of special education schools (11.9%), vocational schools (0.4%), and alternative schools (20.7%). This included 1698 schools statewide (out of 2521 in the CCD file).

Civil Rights Data Collection (CRDC). We received the Office of Civil Rights Data Collection (CRDC) data files for MN, including data collected from USA school districts as mandated by the Civil Rights Act of 1964, the Education Amendments of 1972, the Rehabilitation Act of 1973, and the Department of Education Organization Act. At the CRDC website (https://ocrdata.ed.gov/), this effort is described as "wide-ranging education access and equity data collected from our nation's public schools". This data file was used to compute several school-level variables, including student-teacher ratio, total expense (combined salaries and other expenses), expense per student, teachers with under 2 years of experience, proportion of certified teachers, proportion of students taking at least one Advanced Placement course, number of sports teams offered and proportion of students participating in school athletics, and discipline variables.

Minnesota Student Survey (MSS). The MSS is developed by the MN Departments of Education, Health, Human Services, and Public Safety (https://education.mn.gov/MDE/dse/health/mss/). It is administered every three years and 85% of MN school districts participate, administering the survey to their students voluntarily in grades 5, 8, 9, and 11. From this survey, indicators related to school engagement were obtained and included in these analyses, including student mobility, participation in after-school activities, rates of skipping class and rates of skipping school (full days without an excuse).

Of current interest regarding the consideration of social and emotional learning (SEL) factors, a set of three developmental skills and three developmental supports are measured, which can be considered SEL measures (see Rodriguez, 2017, for a full technical report on these measures). *Commitment to Learning* includes items regarding student engagement in class, preparation for learning, time spent on homework, being achievement oriented and valuing the role of being a student–generally caring about school. *Positive Identity* includes having a sense of control of one's life, feeling good about self and future, dealing well with disappointment and life's challenges, and thinking about one's purpose in life. *Social Competence* embodies the abilities to say no to dangerous/unhealthy things, build friendships, express feelings appropriately, resist bad influences, resolve conflicts without violence, accept differences in others, and recognize the needs and feelings of others. *Empowerment* includes having a sense of safety at home, at school, and in the neighborhood; feeling valued; being included in family roles; and having responsibilities. *Family/Community Support* involves being able to talk with mothers (if available) and feeling cared for by parents, other adult relatives, friends, adults at school, and adults in the community. Finally, *Teacher/School Support* includes the perception that adults at school treat students fairly and listen to students; that youth feel cared for by teachers at school. In addition, scores concerning being bullied (as a victim), engaging in bullying behavior (as a perpetrator), family violence, and mental distress were obtained from the MSS and included in these analyses.

The measurement and inclusion of SEL measures in school accountability has been a recent topic of interest, inspired by the so called 5th allowable indicator in the Every Student Succeeds Act (S. 1177, 2015). ESSA allows states to include

> measures of school quality, climate, and safety, including rates of in-school suspensions, out-of-school suspensions, expulsions, school related arrests, referrals to law enforcement, chronic absenteeism (including both excused and unexcused absences), incidences of violence, including bullying and harassment. (p. 1848)

However, the currently approved ESSA plan in MN does not include aspects related to school climate or SEL, as chronic absenteeism was chosen to meet the federal requirements.

Table 1
Description of Each Variable Included in Each Model

Model	Variable	Level	Variable Description
FRL	FRL	1	Indicator for whether student receives free or reduced-priced Lunch
	FRL.M	2	Proportion of students in a school that receive FRL
FRL+Race	AI	1	Indicator for American Indian student
	API	1	Indicator for Asian/Pacific Islander student
	HIS	1	Indicator for Hispanic student
	BLK	1	Indicator for Black student
	AI.M	2	Proportion of students in a school who are American Indian
	API.M	2	Proportion of students in a school who are Asian/Pacific Islander
	HIS.M	2	Proportion of students in a school who are Hispanic
	BLK.M	2	Proportion of students in a school who are Black
	FRL•AI	1	Indicator for whether student is both American Indian and receives FRL
	FRL•API	1	Indicator for whether student is both Asian/Pacific Islander and receives FRL
	FRL•HIS	1	Indicator for whether student is both Hispanic and receives FRL
	FRL•BLK	1	Indicator for whether student is both Black and receives FRL
	FRL•AI.M	2	Proportion of students in a school who are both American Indian and receive FRL
	FRL•API.M	2	Proportion of students in a school who are both Asian/Pacific Islander and receive FRL
	FRL•HIS.M	2	Proportion of students in a school who are both Hispanic and receive FRL
	FRL•BLK.M	2	Proportion of students in a school who are both Black and receive FRL
All Demographics	LEP	1	Indicator for student's limited English proficiency classification
	SPED	1	Indicator for student's special education classification
	FEMALE	1	Indicator for female
	LEP.M	2	Proportion of students in a school who are classified as SPED
	SED.M	2	Proportion of students in a school who are classified as LEP
	FEM.M	2	Proportion of female students in a school

Table 1 (cont.)

All CCD	FTE	2	Full-time equivalent teachers
	MAGNET	2	Indicator for magnet school
	CHARTER	2	Indicator for charter school
	Diversity	2	Measure of diversity within a school based on number of ethnicities and proportion of ethnicities represented
	STratio	2	Student-to Teacher ratio

CRDC - Teachers	Prop.fte.cert	2	Proportion of FTE teachers in a school that are certified
	Avg.teach.salary	2	Average teacher salary
	Prop.fte.absent	2	Proportion of FTE teachers in a school who were absent more than 10 days of the school year
	Prop.teach.under2	2	Classroom teachers in their first or second year of teaching

CRDC - Resources	Expense.student	2	Expense per student
	Prop.out.sus	2	Proportion of students in a school receiving one or more out of school suspensions
	Prop.in.sus	2	Proportion of students in a school receiving one or more in-school suspensions
	Tot.teams	2	Total sports teams offered at the school
	Prop.athletes	2	Proportion of athletes in a school from all possible athletes in the school

Note. The last two variables regarding sports and athletes are only included in the high school model, since these were only available for high schools.

CRDC - Academics	Num.ap.courses	2	Number of AP courses offered by the school
	Prop.ap.course	2	Proportion of students in a school who took at least one AP course`
	Num.class.math	2	Number of advanced Math courses offered by the school
	Num.class.biology	2	Number of Biology courses offered by the school
	Num.class.physics	2	Number of Physics courses offered by the school
	Num.class.chemistry	2	Number of Chemistry courses offered by the school

Note. These variables are only included in the high school model, since these were only available for high schools.

Table 1 (cont.)

MSS - Engagement	Mobility	2	Proportion of students in a school who moved schools during the year
	Activity3	2	Proportion of students in a school who participate in a combination of number of activities or number of days summing to 3 or more
	Activity.intensity	2	Average days per week the average student in a school participates in an extracurricular activity
	Skip.class	2	Proportion of students in a school who skipped school in the last 30 days
	Skip.school	2	Proportion of students in a school who skipped a class in the last 30 days
MSS - Assets	CTL	2	Commitment to learning
	Em	2	Empowerment
	PI	2	Positive identity
	SC	2	Social competence
	TSS	2	Teacher-school support
	FCS	2	Family/Community Support
MSS - Challenge	Bullied	2	Getting bullied
	Bully	2	Bullying others
	FV	2	Family violence
	MD	2	Mental distress

Note. MSS variables are included in models for grades 5, 8, and high schools, as they are only measured in those grades; skipped school, skipped a class, and mental distress are only included in grades 8 and high school, since they are only measured in those grades.

Analytic models. Hierarchical Linear Modeling (HLM) was employed in order to partition variance in MCA scores within and between schools while also examining the specific student- and school-level characteristics that account for that variation (Peugh, 2010; Raudenbush & Bryk, 2002). All student-level characteristics were group mean centered within school while all school-level characteristics were grand mean centered. The centering procedures result in the student-level and school-level characteristics being uncorrelated, and thus, the regression coefficients are the unbiased within-school (i.e., student-level) and between-school effects (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002).

Using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) in *R* (R Core Team, 2017) a series of models built sequentially were estimated to explore the amount of variance explained uniquely by each set of variables. The series of models were run separately for mathematics and reading scaled scores and by each of the grades 3 to 8 and high school.[1] Additionally, each model was run once using full-information maximum likelihood (FIML) estimation and once with restricted maximum likelihood (REML). FIML is needed for comparing nested models directly whereas REML has been shown to produce less biased variance estimates (Peugh, 2010; Raudenbush & Bryk, 2002). Both purposes were of interest in the present study. Ultimately, like the simulations run by Kreft & de Leeuw (1998), the estimators produced similar results which is unsurprising given that the data used is nearly the entire population. The first model in the series was the unconditional model. The unconditional model with the Level 1 and Level 2 models combined is

$$SSCORE_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

where $\mathrm{SSCORE}_{\mathrm{ij}}$ is the MCA scaled score for student *i* in school *j*, γ_{00} is the mean intercept (school grand mean), u_{0j} is the random effect for school *j*, and r_{ij} is the residual for student *i* in school *j* (student random effect). The distribution of the student residuals and school random effects are assumed to be $r_{ij} \sim N(0, \sigma^2)$ and $u_{0j} \sim N(0, \tau_{00})$, respectively. The intraclass correlation (ICC) is the proportion of the total variance in $SSCORE_{ij}$ that is between-school variation (τ_{00}) and is represented by $\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)}$.

After the unconditional model, the models shown in Table 1 were run with each set of variables added to the antecedent set. Most HLM software, including *lme4*, are unable to account for missing data at Level 2 (i.e., school-level) and although the CCD, CRDC, and MSS datasets attempted to collect information from all schools there was non-response and the schools that did not provide responses varied by dataset. Thus, as variables from the CRDC, for instance, were added to the models that already included the MCA demographic and CCD variables, greater numbers of schools contained missing data, which reduced the sample used in the model. Differing sample sizes means that direct comparisons of the full model to the previous nested models can no longer be made appropriately. To include as many school-level variables as possible while still being able to make comparisons across models, three different samples were created: one with no missing data from the CCD variables, one with no missing data from the

[1] The reading MCA is only taken by grade 10 students and the mathematics MCA is only taken by grade 11 students.

CCD and CRDC variables, and one with no missing data from the CCD, CRDC, and MSS variables. This means that the first sample, which we will call the CCD sample, only the unconditional model, FRL model, FRL+Race, All Demographic, and All-CCD models could be fit. For the CRDC sample (the second sample, which contained fewer schools than the CCD sample), all of the models for CCD sample were fit as well as the Teacher, Resources, and Academics models. Lastly, all of the models from Table 2 were fit with the third and smallest sample – the MSS sample.

From a theoretical standpoint, in order to explain as much of the between-school variation as possible we would fit the models with random effects for every student-level variable, including interaction variables. Practically, however, doing so would create overfitting and convergence issues. To balance the theoretical desire with the practical limitations we first ran a series of pilot models whereby for each of the samples (CCD, CRDC, and MSS) the full model was run once with the only random effect being the intercept (i.e., u_{0j}) and then a separate model for each student-level variable with the variable added as a random effect. A deviance test then determined whether adding the student-level variable as a random effect significantly improved model fit. If the deviance test for a student-level variable as a random effect was significant at $p < .01$ across all grades, then it was included as a random effect in the final series of models.[2] The pilot models produced similar significant random effects across the three samples, six grades, and two subjects. For all three samples with MCA reading as the outcome, FRL, API, BLK, LEP, and SPED were significant as random effects for all grades while the pilot models with MCA mathematics also included HIS as a significant random effect. With group mean centering at Level 1, grand mean centering at Level 2, and random effects for certain Level 1 variables, the FRL+Race model with MCA mathematics as the outcome, for example, is written as

[2] If the random effects used in the model were allowed to differ by grade then we would not be able to appropriately compare the variance explained across grades.

Level-1 Model for FRL, Race/Ethnicity, and Interactions (denoted with •)

$$\begin{aligned} SSCORE_{ij} = {} & \beta_{0j} + \beta_{1j}(FRL_{ij} - FRL.M_j) + \beta_{2j}(AI_{ij} - AI.M_j) + \beta_{3j}(API_{ij} - API.M_j) + \\ & \beta_{4j}(HIS_{ij} - HIS.M_j) + \beta_{5j}(BLK_{ij} - BLK.M_j) + \beta_{6j}(FRL{\bullet}AI_{ij} - FRL{\bullet}AI.M_j) + \\ & \beta_{7j}(FRL{\bullet}API_{ij} - FRL{\bullet}API.M_j) + \beta_{8j}(FRL{\bullet}HIS_{ij} - FRL{\bullet}HIS.M_j) + \\ & \beta_{9j}(FRL{\bullet}BLK_{ij} - FRL{\bullet}BLK.M_j) + r_{ij} \end{aligned}$$

Level-2 Model

$$\begin{aligned} \beta_{0j} = {} & \gamma_{00} + \gamma_{01}(FRL.M_j - \overline{FRL.M}) + \gamma_{02}(AI.M_j - \overline{AI.M}) + \gamma_{03}(API.M_j - \overline{API.M}) + \\ & \gamma_{04}(HIS.M_j - \overline{HIS.M}) + \gamma_{05}(BLK.M_j - \overline{BLK.M}) + \gamma_{06}(FRL{\bullet}AI.M_j - \overline{FRL{\bullet}AI.M}) + \\ & \gamma_{07}(FRL{\bullet}API.M_j - \overline{FRL{\bullet}API.M}) + \gamma_{08}(FRL{\bullet}HIS.M_j - \overline{FRL{\bullet}HIS.M}) + \\ & \gamma_{09}(FRL{\bullet}BLK.M_j - \overline{FRL{\bullet}BLK.M}) + u_{0j} \end{aligned}$$

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

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

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

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

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

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

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

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

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

Random-effect coefficients from Level-1 were not modeled at level-2, as the purpose of the models were to explain variation in student-level achievement (SSCORE$_{ij}$, within schools) and school-level achievement (β_{0j}, between schools). In some cases, as noted above, level-2 random effects (u_j) were not estimable or nonsignificant and were removed (fixed), and this was done so consistently across grades to support model comparison.

The series of HLM models were run sequentially for each grade and each test subject (mathematics and reading). The sequence of model is as it appears in Table 1. Each subsequent model includes all variables from the antecedent model.

1. Unconditional (no explanatory variables)
2. Race/ethnicity only
3. FRL with race/ethnicity × FRL interaction terms
4. Additional student status variables (gender, limited English proficiency, special education)
5. CCD school variables
6. CRDC teacher variables
7. CRDC school resource variables and school academic variables
8. MSS student engagement variables (at the school level)
9. MSS SEL measures (at the school level)

After running all of the models the ICCs were calculated from the unconditional models to determine the amount of total variance in MCA scores due to within-school and between-school differences. This was done to account for the reduced sample size as subsequent data files added to the models contained missing data for some schools (described above). The reduction in within-school variance (σ^2) and reduction in between-school variance (τ_{00}) was calculated for each model, based on its unconditional variance estimates, to determine the extent to which the variables in the model explained variation in MCA scores.

Results

As a brief summary of some initial exploratory analyses of the achievement data, we found the following:

1. Students in each racial/ethnic group vary substantially – in most grades, students in each group achieve the lowest and highest possible scores, with the exception of American Indian students in mathematics grades 5-8 who do not obtain the highest possible score.
2. School averages vary substantially and the average scores of high-poverty schools (about 18% of schools) are much lower than the average scores of low-poverty schools (about 23% of schools), with little overlap in distributions (poverty level was defined with NCES, 2016, definitions).
3. Within each racial/ethnic group, students participating in FRL score much lower than students not in FRL (with Asian students even more so).
4. After accounting for FRL status, gaps between White students and students of color are reduced, but still large.
   a. For American Indian, Black, and Latino students in FRL, gaps are 35% smaller.
   b. For these students <u>not</u> in FRL, the gaps are 43% smaller.
5. On average, American Indian, Latino, and Black students <u>not</u> in FRL score at levels near White students in FRL, generally across grades and subjects.

Regarding school composition, we found:

6. On average, White students attend schools that are 5% Black, whereas Black students attend schools that are 30% Black.
7. Nearly 94% of the state's White students attend schools with 0-19% Black students, whereas 42% of Black students attend such schools (this includes 85% of MN schools). Another way to say this is: 58% of Black students attend schools with 20% or more Black students, compared to 6% of White students).

These results suggest that there is a connection between FRL and race/ethnicity, but racial/ethnic achievement disparities remain after accounting for FRL (with a reduction of about 1/3). In addition, there is a significant amount of segregation, as expected.

Partitioning Variance

The results of the initial unconditional models, for the purpose of the baseline partitioning of variance are in Tables 2 (mathematics) and 3 (reading). Recall that the unconditional model is:

$SSCORE_{ij} = \gamma_{00} + u_{0j} + r_{ij}$ where $r_{ij} \sim N(0, \sigma^2)$ and $u_{0j} \sim N(0, \tau_{00})$ and $\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)}$.

Table 2
Partitioning Variance in Mathematics by Grade

Grade	# Students	# Schools	τ_{00}	σ^2	ρ
3	59825	862	46.96	202.52	**.19**
4	60221	854	60.67	254.65	**.19**
5	58139	826	31.48	141.85	**.18**
6	56326	594	38.05	156.60	**.20**
7	57068	509	21.17	108.44	**.16**
8	55867	505	29.70	154.92	**.16**
11	52982	430	48.96	234.05	**.17**

Note: τ_{00} is between-school variance, σ^2 is within-school variance, ρ is the ICC.

Table 3
Partitioning Variance in Reading by Grade

Grade	# Students	# Schools	τ_{00}	σ^2	ρ
3	59717	862	62.48	350.78	**.15**
4	60107	854	37.72	195.23	**.16**
5	57846	826	32.77	169.48	**.16**
6	56203	594	41.76	254.99	**.14**
7	57163	509	40.75	260.84	**.14**
8	56139	505	38.56	260.53	**.13**
10	55390	430	25.95	189.30	**.12**

Note: τ_{00} is between-school variance, σ^2 is within-school variance, ρ is the ICC.

Most notable, the between-school variance is relatively small, ranging from 16 to 20% for mathematics and from 12 to 16% in reading; these are lower than ICCs reported by researchers investigating other testing programs, as reviewed above. However, consistent with prior studies, the ICCs are larger for mathematics than reading.

Modeling Within and Between-School Variance

Based on the series of HLM models of variability, we can summarize the findings here:

1. Generally, more than 80% of the variance (80 to 88%) in the mathematics and reading achievement scores is within-school variance.
2. FRL and race/ethnicity are powerful explanatory variables regarding school variance, accounting for 64% of the between-school variance in mathematics on average (58% to 69% across grades), and about 73% in mathematics on average (66% to 77% across grades).
3. Inclusion of LEP, SPED, Gender and Race × LEP interactions (All Demos model) more than double the amount of within-school variation explained for both mathematics and reading as compared to the amount explained by just FRL and race/ethnicity. However, these additional variables only accounted for 1% to 2% additional variance between schools on average.
4. Demographics explain about 20% of the within-school variation of mathematics and reading scores; whereas they explain 66% to 74% of between-school variance.
5. The amount of within school variation explained by the models does not vary much by grade.
6. CCD variables did not explain much additional between-school variance (and no within school variation) beyond what was explained by demographics (about 0 to 2% additional variance across grades).
7. CRDC teacher variables explained an additional 0 to 1% between-school variance. CRDC school resource variables explained an additional 0 to 2% between-school variance. CRDC academic variables (available for high schools only) explained an additional 1% between-school variance. It's important to note that these values are the incremental between-school variance explained, such that the CRDC variables explained less than 1% of the 12 to 20% of variance between schools.
8. The measures of SEL and other challenges measured by the MSS explained an additional 0 to 3% of between-school variance, restricted to grades 5, 8, and high school test scores, more so in reading scores than mathematics scores.

Tables 4 and 5 contain the unconditional ICCs and the unique and incremental variance explained by the addition of each set of variables regarding mathematics scores (Table 4) and reading scores (Table 5).

Table 4

Unique Between School Variance in Mathematic MCA Scores Explained by Model and Data Source for Each Grade

	CCD							CRDC							MSS		
Grade	3rd	4th	5th	6th	7th	8th	HS	3rd	4th	5th	6th	7th	8th	HS	5th	8th	HS
ICC	.19	.19	.18	.20	.16	.16	.17	.17	.17	.16	.17	.14	.15	.15	.14	.13	.11
Model																	
FRL	.65	.60	.62	.63	.58	.52	.60	.63	.56	.58	.57	.53	.47	.59	.56	.48	.48
FRL+Race	.04	.04	.02	.02	.04	.06	.04	.05	.04	.01	.02	.06	.08	.03	.01	.05	.06
All Demos	.01	.00	.02	.00	.00	.02	.08	.01	.00	.01	.00	.00	.01	.07	.03	.04	.06
CCD	.00	.00	.01	.00	.01	.02	.01	.00	.00	.01	.00	.02	.03	.01	.00	.03	.01
Teacher								.00	.00	.00	.00	.01	.00	.00	.00	.00	.00
Resources								.00	.00	.00	.02	.02	.03	.01	.00	.03	.01
Academic														.01			.03
Engagement															.00	.02	.03
Assets & Challenges															.00	.02	.00
Total Explained	.70	.64	.65	.64	.63	.62	.72	.69	.60	.63	.60	.63	.62	.72	.62	.67	.68

Note. CCD = Common Core of Data (NCES); CRDC = Civil Rights Data Collection, including Teacher, Resources, and Academic variables; MSS = Minnesota Student Survey, including Engagement and Assets & Challenges variables; ICC = Intraclass correlation; FRL = free or reduced-priced lunch; models are defined in Table 1. Unique variance explained in each column sum to Total Explained within rounding error.

Table 5

Unique Between School Variance in Reading MCA Scores Explained by Model and Data Source for Each Grade

	CCD							CRDC							MSS		
Grade	3rd	4th	5th	6th	7th	8th	HS	3rd	4th	5th	6th	7th	8th	HS	5th	8th	HS
ICC	.15	.16	.16	.14	.14	.13	.12	.13	.14	.14	.12	.12	.11	.10	.12	.09	.08
Model																	
FRL	.70	.72	.74	.69	.72	.64	.57	.68	.70	.72	.64	.68	.60	.54	.69	.57	.49
FRL+Race	.04	.04	.01	.03	.05	.04	.09	.04	.04	.01	.03	.06	.05	.10	.00	.07	.15
All Demos	.00	.01	.00	.00	.00	-.01	.07	.00	.00	.00	.00	-.01	-.01	.04	.03	.00	.01
CCD	.01	.00	.01	.01	.02	.02	.01	.01	.00	.01	.00	.02	.03	.01	.00	.03	.00
Teachers								.00	.00	.01	.00	.00	.01	.01	.01	.00	.01
Resources								.00	.00	.00	.02	.01	.01	.00	.00	.00	.00
Academics														.01			.02
Engagement															.01	.02	.04
Assets & Challenges															.01	.03	.03
Total Explained	.75	.77	.76	.72	.78	.69	.75	.73	.75	.75	.70	.76	.67	.70	.77	.72	.74

Note. CCD = Common Core of Data (NCES); CRDC = Civil Rights Data Collection, including Teacher, Resources, and Academic variables; MSS = Minnesota Student Survey, including Engagement and Assets & Challenges variables; ICC = Intraclass correlation; FRL = free or reduced-priced lunch; models are defined in Table 1. Unique variance explained in each column sum to Total Explained within rounding error.

Exploring the Independent Role of School-Level Factors

Since FRL and race/ethnicity were such powerful explanatory variables, we wanted to explore the independent role of the school-level factors absent the demographic variables (considering schools to come from the same population, essentially without the effects of segregation and unequal distribution of student characteristics). The resulting variance explained for the sequential inclusion of the variables in models without demographics are contained in Tables 6 (mathematics) and 7 (reading). Those findings can be summarized as:

1. When demographic variables are not included, school-level factors explain 29 to 60% of the between-school variance in mathematics and 23 to 66% for reading (including CRDC and MSS variables).
2. School-level factors tend to explain more between-school variance at the higher grades for mathematics, but not necessarily for reading.
3. Teacher variables explain less variance during middle school (grades 6 to 8) than elementary or high school.
4. We explored the data further and found that there is a confounding among the set of demographic variables and these school resource variables. One possible reason that these variables add little to no incremental variance explained to the full models is that they vary with student demographics – schools with more students in FRL and more students of color are associated with lower levels of resources. So when school demographics are controlled, the potential impact of resources has been explained (accounted for).

Table 6

Unique Between-School Variance in Mathematic MCA Scores Explained by Level 2 Only Model and Data Source for Each Grade

	CRDC							MSS		
Grade	3rd	4th	5th	6th	7th	8th	HS	5th	8th	HS
ICC	.17	.17	.16	.17	.14	.15	.15	.14	.13	.11
Model										
Teachers	.23	.20	.23	.15	.17	.14	.37	.32	.24	.45
Resources	.09	.08	.08	.15	.17	.22	.11	.05	.21	
Academics							.04			.06
Engagement								.09	.13	.05
Assets								.04	.01	.02
Challenges								.01	.01	.02
Total Explained	.32	.29	.31	.30	.34	.36	.52	.51	.60	.60

Note. CRDC = Civil Rights Data Collection, including Teachers, Resources, and Academics variables; MSS = Minnesota Student Survey, including Engagement, Assets and Challenges variables; ICC = Intraclass correlation; models are defined in Table 1.

Table 7

Unique Between School Variance in Reading MCA Scores Explained by Level 2 Only Model and Data Source for Each Grade

	CRDC							MSS		
Grade	3rd	4th	5th	6th	7th	8th	HS	5th	8th	HS
ICC	.13	.14	.14	.12	.12	.11	.10	.12	.09	.08
Model										
Teachers	.21	.20	.20	.14	.11	.08	.21	.38	.24	.28
Resources	.12	.11	.13	.17	.17	.15	.07	.08	.16	
Academics							.05			.12
Engagement								.15	.17	.14
Assets								.04	.04	.04
Challenges								.01	.00	.03
Total Explained	.33	.31	.32	.31	.27	.23	.32	.66	.60	.61

Note. CRDC = Civil Rights Data Collection, including Teachers, Resources, and Academics variables; MSS = Minnesota Student Survey, including Engagement, Assets and Challenges variables; ICC = Intraclass correlation; models are defined in Table 1.

Exploring the Independent Role of School-Level SEL Measures

Given the recent interest in the possible role of SEL measures, we explored their associations with academic achievement without the influence of student demographics and other school resources. It appears that a similar case exists here as with school resources. There is some a priori alignment between student demographics and SEL, and between SEL and school resources, as with student demographics and school resources. Once school resources are accounted for, the SEL measures provide little to no additional explanation of achievement variance. In Table 8, we see that in 5th grade mathematics achievement scores, student demographics explain 57% of the variance between schools, whereas SEL measures explain 47% of the variance between schools. However, once school demographics are accounted for, the measures in combination explain an additional 7.3% of the between-school variance. However, once school resources are accounted for, this drops to zero. The findings are similar for reading and the other two grades where SEL measures were available.

Table 8
Percent of Between-School Variance Explained by SEL, With and Without Student Demographic Information; 5th Grade Mathematics

Variables	Percent of variance explained
Student demographics alone	57.4%
Social & emotional learning alone	46.9%
Above & beyond student characteristics	7.3%
Teacher/School Support	2.6%
Bullied	4.8%
Commitment to Learning	1.1%
Positive Identity	5.3%
Social Competence	6.7%

Summary of Variance Partitioning and Variance Explained

In an attempt to represent the many models graphically (98 models were estimated in total), we included the first 84 models in Figure 1. These models do not include the models involving the MSS student engagement and SEL measures, as those only included grades 5, 8, and high schools, but moreover, did not explain additional variance between schools. Figure 1 illustrates the extent to which the variance in mathematics and reading achievement scores is a function of within and between-school variance and the extent to which each model explains both within and between-school variance. The purple-shaded regions, near the top of each stacked-bar, represents the volume of between-school variance, the green-shaded regions represent the volume of within-school variance. Notice that the green bars are substantially larger (accounting for the 80%+ variance within schools). Within each color, the darker color accounts for the explained variance.

Figure 1 illustrates a few high-level findings:

1. The vast majority of achievement score variance is within-school variance (80% or more).
2. FRL and race/ethnicity explain a significant portion of between-school variance (63 to 74% on average) and little within-school variance (5 to 10%).
3. With the additional student demographic variables, a slight additional variance is explained between-schools, but the within-school variance explained is more than doubled.
4. Additional school-level variables, including information about teachers, resources, academics, explain hardly noticeable additional variance between schools.

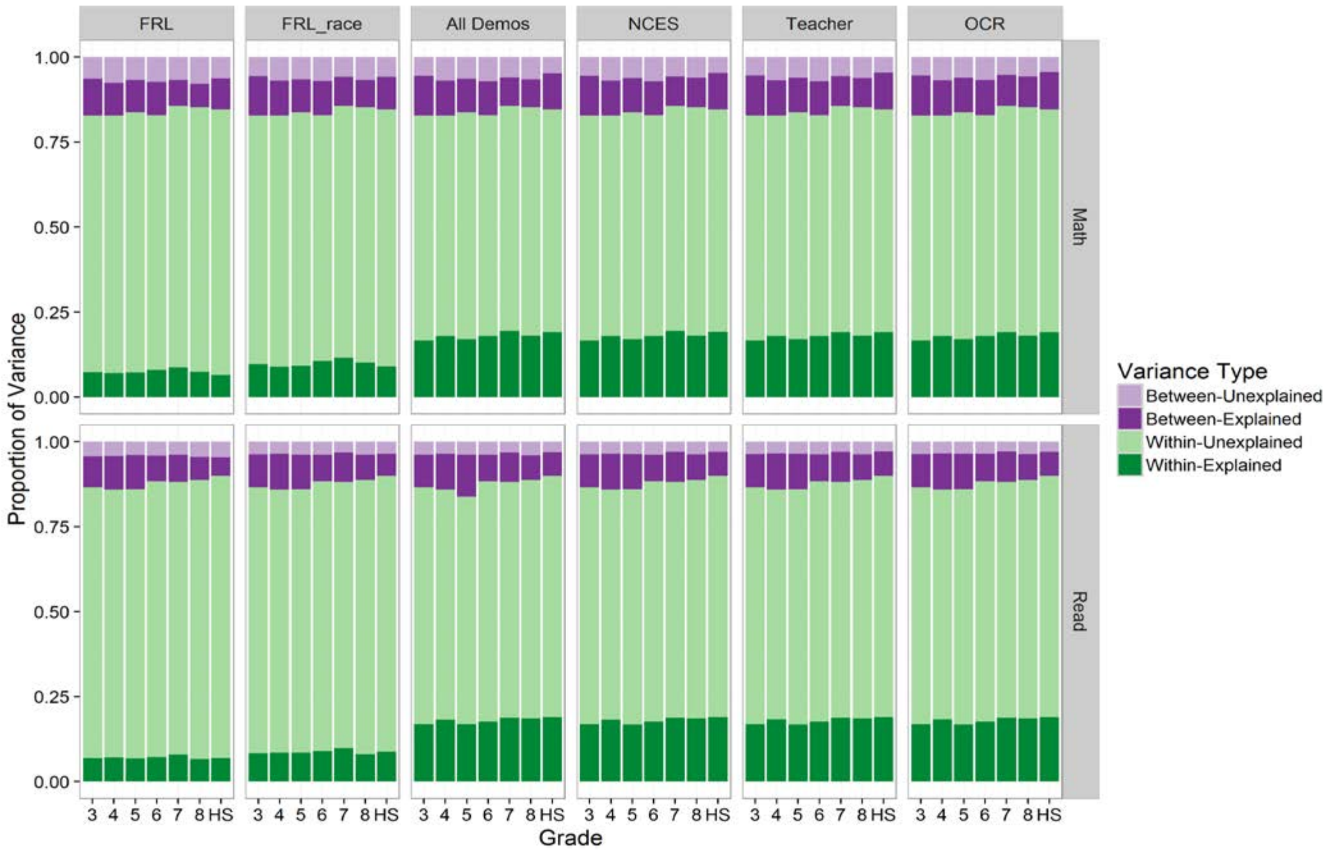

Figure 5. Between and within school variance in Math and Reading MCA scores explained and unexplained by each model by grade.

Each subsequent block includes all previous variables:

Student-level variables (group-mean centered) and school-level averages
FRL: Free/reduced price lunch indicator (proportion at school-level).
Race: Four race/ethnicity indicators (proportions at school-level) + FRL×Race interactions.
All Demos: Gender + Special Ed. Status + ELL Status (proportions at school-level).

Additional school-level variables
NCES: The CCD variables, including teacher FTE, magnet school indicator, charter school indicator, student-teacher ratio.
Teacher: the CRDC variables regarding teachers, including proportion of certified teachers, average teacher salary, proportion FTE teachers absent more than 10 days during the year, proportion of teachers in 1st or 2nd year of teaching.
OCR: Additional CRDC school resource variables, including expenses per student, proportion receiving out-of-school suspensions, proportion receiving in-school-suspensions, total number of sports teams offered, proportion of athletes in school.

The following school level variables were added to the above models with no noticeable change:
CRDC academic variables, MSS student engagement variables, MSS measures of SEL and challenges.

Implications

The measurement of academic achievement is driven by two primary drivers or realities. First, individuals differ and the research on individual differences has inspired much of what we do in educational and psychological measurement. Second, public education is an important societal investment, as it supports the training and preparation of people to contribute to their family, community, and the world. For as long as we have engaged in the measurement of educational outcomes, there also has been a need to engage in accountability-associated decision making; we believe that decisions are better made when they are informed. As we strive to improve the educational preparation and outcomes of students at all levels of education, we similarly strive for educational equity, meeting the unique learning needs of all students. In fact, at some level, the inspiration behind some elements of ESEA, NCLB, and ESSA (federal accountability legislation) is based in equity concerns, essentially to close achievement gaps.

Partitioning variance in student achievement is an important way to understand where individual and group differences occur and to what we might attribute variation in those differences. Sources of variation (those things that explain variation) potentially become policy targets. Much of the federally funded research in education is devoted to finding those malleable factors that might influence individual and group differences in achievement. This partitioning of variance is particularly helpful when student and school-level indicators that are malleable are found to explain variation in achievement. Although many student-level indicators are not necessarily malleable, they do help us identify groups of students who may need additional supports – the practices and policies regarding the support of students with different characteristics are malleable.

Given the consistency of ICCs across grades and subject areas, there are policy implications regarding the focus of school supports. When the vast majority of variation in student achievement is within schools, rather than between schools, the focus might better serve school and student needs if focused within school, rather than school differences, which is the current focus of federal and state school accountability policies. These findings and recommendations are consistent with those of many others (e.g., Baker, Farrie, & Sciarra, 2015; Borhnstedt, 2015; Figlio & Karbownik, 2017; Orfield, 2017; Yeh, 2017).

From these results, the promise of SEL measures to contribute meaningfully to accountability policy indicators of school quality is limited. The SEL measures employed here do not appear to contribute additional explanatory power regarding school-level achievement.

Overall, these findings have implications for equitable distribution of key education resources across schools or to focus equitable use of resources within schools to reduce achievement gaps and improve educational achievement for all students. We find that school resources are important, but once we account for school demographics, the role of resources is negligible (because they are distributed inequitably in a way that is associated with school demographics).

We need to call into question the practice of identifying schools as the target for federal education accountability. It is clear that the majority of variance in achievement occurs within schools, and the smaller portion that occurs between schools (less than 20%) is mostly explained by within-school demographics (about 70% or more). Less than 7% of variance in achievement remains between schools, little of which is explained by school level teacher characteristics, school resources, school advanced academics, or SEL (indicators valued in the current ESSA requirements).

References

Baker, B.D., Farrie, D., & Sciarra, D.G. (2016). *Mind the gap: 20years of progress and retrenchment in school funding and achievement gaps*. Princeton, NJ: Policy Information Center, Educational Testing Service. Retrieved from http://www.ets.org/research/pic

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). *School Composition and the Black–White Achievement Gap* (NCES 2015-018). U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Cottrell, J.M., Newman, D.A., & Roisman, G.I. (2015). Explaining the Black-White gap in cognitive test scores: Toward a theory of adverse impact. *Journal of Applied Psychology, 100*(6), 1713-1736.

S. 1177, Public Law 114-95, 114 Cong. (2015). *Every Student Succeeds Act* (enacted). Retrieved from https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.

ESSA (2015). *Every Student Succeeds Act* (S. 1177, Public Law 114-95). General information can be found at https://www.ed.gov/ESSA

Fahle, E.M., & Reardon, S.F. (2017). *How much do test scores vary among school districts? New estimates using population data, 2009-2013*. CEPA Working Paper No. 17-02. Stanford, CA: Center for Education Policy Analysis, Stanford University. Retrieved from http://cepa.stanford.edu/wp17-02

Figlio, D., & Karbownik, K. (2017). Some schools much better than others at closing achievement gaps beteween their advantaged and disadvantaged students. *Evidence Speaks Reports, 2*(19), 1-8. Retrieved from https://www.brookings.edu/research/some-schools-much-better-than-others-at-closing-achievement-gaps-between-their-advantaged-and-disadvantaged-students/

Hanushek, E.A., & Woessmann, L. (2017). School resources and student achievement: A review of cross-country economic research. In M. Rosen, K.Y. Hansen, & U. Wolff (Eds.), *Cognitive abilities and educational outcomes*. Cham, Switzerland: Springer International.

Harwell, M.R., Maeda, Y., & Lee, K. (2004, April). *Replicating and extending White's (1982) meta-analysis of the relationship between SES and student achievement*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Kim, J., & Choi, K. (2008). Closing the gap: Modeling within-school variance heterogeneity in school effect studies. *Asia Pacific Education Review, 9*(2), 206-220.

Kreft, I. G. G.., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.

MN Department of Education. (2016). *Frequently asked questions: Why statewide test results matter*. Roseville, MN: Author. Retrieved from https://education.state.mn.us/mdeprod/groups/educ/documents/hiddencontent/bwrl/mdm0/~edisp/mde034287.pdf

Minnesota Student Survey Interagency Team. (2013). *2013 Minnesota Student Survey statewide tables* (White paper). Retrieved from http://education.state.mn.us/MDE/dse/health/mss/

NAEP. (2017). *National Assessment of Educational Progress*. Washington DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/nationsreportcard/

National Center for Education Statistics. (2015). *2013-14 Public Elementary/Secondary School Universe Survey Data* [Data file and codebook]. Retrieved from https://nces.ed.gov/ccd/pubschuniv.asp

National Center for Education Statistics. (2016). *The condition of education 2016*. Washington DC: Institute of Education Sciences, US Department of Education. Retrieved from http://nces.ed.gov/pubs2016/2016144.pdf

NCLB (2001). *No Child Left Behind Act* (Public Law PL 107-110). General information can be found at https://www2.ed.gov/nclb/

OECD (2004). How student performance varies between schools and the role that socio-economic background plays in this. In Organisation for Ecomomic Co-Operation and Development, *Learning for tomorrow's world: First results from PISA 2003* (pp. 159-205). Paris, France: Author. Retrieved from https://www.oecd-ilibrary.org/

Office of the Legislative Auditor. (2017). *Standardized student testing. Evaluation Report*. St. Paul, MN: Author, State of Minnesota. Retrieved from http://www.auditor.leg.state.mn.us/ped/2017/studenttesting.htm

Orfield, M. (2015). *Why are the Twin Cities so segregated?* Minneapolis, MN: Institute on Metropolitan Opportunity, University of Minnesota. Retrieved from https://www.law.umn.edu/institute-metropolitan-opportunity/school-integration-and-segregation

Orfield, M. (2017). *The Minnesota school choice project. Part I: Segregation and performance*. Minneapolis, MN: Institute on Metropolitan Opportunity, University of Minnesota. Retrieved from https://www.law.umn.edu/institute-metropolitan-opportunity/school-choice

Peugh, J. J. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*, 85-112.

R Core Team. (2017). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reardon, S.F. (2015). *School segregation and racial academic achievement gaps.* CEPA Working Paper No. 15-12. Stanford, CA: Center for Education Policy Analysis, Stanford University. Retrieved from http://cepa.stanford.edu/wp15-12

Reardon, S.F., Kalogrides, D., & Shores, K. (2016). *The geography of racial/ethnic test score gaps*. CEPA Working Paper No. 16-10. Stanford, CA: Center for Education Policy Analysis, Stanford University. Retrieved from http://cepa.stanford.edu/wp16-10

Rodriguez, M.C. (2017). *Technical report on developmental skills, supports, and challenges from the 2013-2016 Minnesota Student Survey* (White paper). University of Minnesota, Minneapolis, MN: Minnesota Youth Development Research Group. Retrieved from http://www.mnydrg.com/research

Whitehurst, G.J., Reeves, R.V., & Rodrigue, E. (2016). *Segregation, race, and charter schools: What do we know?* Washington DC: Center on Children and Families, Brookings Institution. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/10/ccf_20161021segregation_version-10_211.pdf

Yeh, S.S. (2017). *Solving the achievement gap: Overcoming the structure of school inequality*. New York, NY: Palgrave Macmillan.