A characterization of sorting and implications for value-added estimates

Jessica Alzen[1]
Ben Domingue[2]

2013

[1] School of Education, University of Colorado Boulder
[2] Institute of Behavioral Science, University of Colorado Boulder

A characterization of sorting and implications for value-added estimates

Abstract

Students are non-randomly assigned, or sorted, into classrooms in various ways across and within schools. In this study, we use longitudinal data sets from three districts to investigate a metric for the characterization of sorting at the school level. We analyze whether non-random student assignment is associated with value-added estimates for teachers. The three longitudinal datasets come from large, urban districts but, despite this similarity, we find there is substantial variation in the degree of sorting across school districts. We see more evidence of sorting in districts with higher proportions of ELLs and students from low socioeconomic backgrounds. Despite differences in the characteristics of schools which sort, sorting as we have quantified it is only marginally related to VA estimates.

Key words: ability grouping; accountability; high stakes testing; school/teacher effectiveness

A characterization of sorting and implications for value-added estimates

The wealth of data collected under No Child Left Behind's mandated testing of all students in grades 3-8 provides the longitudinal data necessary to investigate trends in school and district practices. These practices include the deliberate grouping of students into classes (which we refer to as sorting) over time as well as methods of estimating teacher effectiveness such as value-added models (VAMs). VAMs use multiple years of longitudinal rather than cross-sectional data and control for various factors such as prior achievement, English Language Learner (ELL) status, and proxies for socioeconomic status such as qualifying for Free and Reduced Lunch (FRL) to create estimates of teacher effectiveness (Goldhaber, Walch, &Gabele, 2012; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). These estimates are widely, though often controversially, employed to estimate a teacher's causal effect on student learning (Baker et al., 2010; McCaffrey, Lockwood, Koretz, & Hamilton, 2003). One problem with using VA estimates for teacher evaluation is that students are non-randomly assigned, or sorted, into classrooms in ways that are potentially unaccounted for in value-added models (Rivkin, 2007).This non-random assignment may bias VA estimates (Rothstein, 2010). Because such estimates are frequently believed to be measures of teacher quality and may be used for personnel evaluation, this potential bias is worrisome.

Previous literature on sorting focuses on how purposeful sorting of students based on ability may lead to negative effects on students, such as segregation within schools (Oaks 2005; Wells & Serna 2006; Zimmer, 2003). In this study, we investigate variation in sorting practices across and within schools as well as seek to identify characteristics of schools which tend to sort more than others. We also explore whether student sorting may impact teachers through their VA estimates. Despite efforts to control for student characteristics, teachers with higher proportions of special education students, ELLs, and low-income students often receive lower "effectiveness" scores than teachers with lower populations of these students (Baker et al., 2010). Whatever the educational advantages or disadvantages of non-random assignment for *students*, sorting may also have implications for *teachers*.

In order to investigate student sorting, we propose a metric which compares the distribution of student ability within classrooms to the overall distribution of student ability within a school. For a given school, we consider the ratio of the weighted mean of the standard deviations of prior year abilities assigned to teachers to the overall standard deviation of prior year abilities within a school. Sorting is thus defined as something that happens within schools or at the school rather than classroom level. Values close to unity on this sorting metric indicate that classes generally contain the same distribution of student ability as in the school overall. That is, the school does not engage in the practice of ability sorting in a way that pools students of common abilities. Decreases in this metric are evidence of sorting within the school as classes contain more concentrated groups of student abilities than the school as a whole. Under this characterization of sorting, a *school* engages in ability sorting, not classrooms. Classrooms may have higher or lower average student abilities or more or less concentrated groups of student abilities, but we operationalize sorting as a school-level phenomena.

This study focuses on three primary research questions:

1. How much does the proposed metric for sorting vary? Variation is explored both within and across districts.
2. To what extent, if at all, is this variation related to ELL and low SES populations?
3. How is the metric related to value-added estimates of teacher effectiveness?

The paper is organized as follows. In Section 1, we describe the context in which we pose these research questions and present background information regarding methods of characterizing sorting. In Sections 2 and 3, we discuss our data and methods. We then present the results in section 4, followed by discussion and conclusions in Sections 5 and 6.

1. Context and Background

The sorting of students into classrooms based on factors such as teacher strengths, parent preferences, and student achievement is a relatively common practice (Harris, 2009; Rivkin, 2007). Depending on school climate, more or less effective teachers may be disproportionately assigned students who are more difficult to teach for a variety of reasons: teacher favoritism, matching of student and teacher personalities, or deference to persistent parents (Rothstein, 2010). While these assignments may be beneficial to classroom instruction and school climate, they may create bias in estimates of teacher effectiveness. Little work has attempted to identify if particular kinds of schools engage in the practice more than others. Further, sorting practices may vary widely across districts. Understanding if and how sorting practices vary systematically within and/or across districts provides insight into school practices and may have notable relationships to teacher effectiveness rating as well.

Previous investigation applying what is known as the “Rothstein Falsification Test” suggests that a teacher’s current VA estimate can be indicative of her students’ prior year achievement (Rothstein, 2010; Author, 2011). Since it is illogical to think that a 5th-grade teacher has any influence over her students’ 4th-grade achievement, these predictions suggest that the placement of 4th-grade students into 5th-grade classrooms could have implications for the estimates from certain VA models. The existence of this type of bias is concerning since VA estimates are increasingly used in rating systems that are tied to high-stakes decisions. The induced bias in VA estimates and related policy decisions may essentially stack the deck against teachers assigned to certain types of classrooms. It is important to understand, or at least acknowledge, the mechanisms by which student ability is distributed among classrooms so that any potential influences are considered when determining uses of VA estimates.

Prior studies regarding student sorting into classrooms have taken various approaches to investigate the practice. The Rothstein Falsification Test has been applied several times since it was first suggested (Author, 2011; Koedel& Betts, 2009). Clotfelter, Ladd, &Vigdor (2006) use descriptive statistics to examine the relationship between teacher and student characteristics and employ several χ^2 tests to determine if student assignment to classrooms is statistically independent of student gender, race, FRL status, previous attendance at the same school, prior test score, or parental education level. Boyd et al. (2009) control for both mean and standard deviation of classroom prior student achievement in their study relating teacher preparation to student achievement to capture any potential relationships regarding the spread of student achievement. Collins &Gan (2013) use a ratio of the overall standard deviation of prior abilities within a school to the weighted mean of the standard deviation of the prior year abilities within a classroom and assigned to a particular teacher. This method of characterizing sorting is similar to the method we employ in this paper.

The method of sorting discussed here is widely applicable since it does not depend on a specific VAM and focuses at the school level rather than classroom. In contrast, the Rothstein Falsification Test depends upon a specific VA model. This metric would also be difficult to compare across districts. We investigate sorting as a school-level phenomenon, so the methods employed by Clotfelter, Ladd, &Vigdor (2006) and Boyd et al. (2009) are insufficient as they focus on the classroom level. The method we choose to employ in this paper is similar to that of Collins &Gan (2013) and adds relevant analysis by attempting to characterize and describe sorting as a school-level phenomenon that is not reliant on a specific VA specification.

Sorting as it relates to teacher effectiveness estimates is the topic of much debate regarding VA methodology. However, we believe that the implications of the concept have not yet been fully developed for several reasons. Sorting is a policy decision made at the school-level that affects classroom composition for multiple teachers. Value-added estimates are typically teacher-level estimates. Claiming that sorting influences VA estimates necessitates additional theorizing regarding the details of causal mechanism through which this school-level policy is influencing classroom-level outcomes. One plausible mechanism would be that teachers with classes made up of disproportionate quantities of ELL students, for example, could expect to have downwardly biased VA estimates. While it is unclear why this would happen, previous research has found it in some instances (Baker et al., 2010). This is notable because VA estimates are designed to control for student characteristics. An alternative possibility is that ability grouping actually makes the teaching process easier since less differential instruction is needed (Collins &Gan, 2013). If the latter is true, one might expect to see systematically higher VA estimates for teachers in those schools that practice sorting. Additionally, it is important to consider the fact that the types of bias induced by sorting probably differ as a function of local context and further analysis of how sorting manifests itself differently across a variety of such contexts is an important consideration for this study. Although our research is not able to completely resolve these issues, we think that this framing of the issue is crucial to a deeper understanding of the relationship between sorting and estimates of teacher effectiveness.

2. Data

This study uses longitudinal data sets from three distinct districts in three different states. The districts are anonymized here as Bayview Public Schools (BPS), Central Unified School District (CUSD), and Norton Public Schools (NPS). For all districts we use scale scores in math and reading while in BPS we also utilize scale scores in writing. We include analysis of multiple subjects because prior analyses have suggested that there might be differential evidence for sorting in math and reading (Author, 2011). Table 1 provides basic demographic information for elementary students in the three districts.

We concentrate on grade 5. This decision was motivated in part by data restrictions (primarily the desire to have multiple prior scores, which eliminated lower grades since testing typically starts in grade 3). We use 5th grade students from CUSD during 2005-2009 and 5th grade students from BPS and NPS in 2012 (this requires data from 2003-2009 for CUSD and 2010-2012 for BPS and NPS to get at least two prior scores for each student). Each of the three districts serves large, urban areas and student populations of over 15,000 students. BPS and CUSD serve student populations more similar in proportions of ELLs and students from lower socioeconomic status than NPS. Roughly 40% of the student sample in BPS and CUSD are identified as ELLs, while NPS serves far fewer ELL students (5% in our sample). Nearly all of the students in CUSD come from lower SES households (as indicated by Title I status), and about three-quarters of the students in BPS qualify for free and reduced lunch. However, NPS has 27% of their student population receiving Title I services. While the three districts are not geographically proximate, there is far greater similarity in several respects between BPS and CUSD, while NPS differs quite a bit from these two.

3. Methods

We focus on a school-level index to measure sorting. For a given school, consider the ratio of the weighted mean of the standard deviations of prior year abilities assigned to a teacher (weights are the number of student assigned to each teacher) to the overall standard deviation of prior year abilities within a school:

$$\frac{\sum_j N_j \sigma_{Y_{ij,t-1}}}{\sigma_{Y_{i,J,t-1}}}. \qquad (1)$$

In this equation, j indexes teachers within school J, N_j is the number of students associated with teacher, and $\sigma_{Y_{ij,t-1}}$ and $\sigma_{Y_{i,J,t-1}}$ are standard deviations of prior scores for teacher j and school J respectively. This measure provides information on how dissimilar the distribution of student prior achievement across classes is from the distribution one would observe if students were randomly grouped into classes. This metric is analogous to an F-statistic, which compares between group variance (distribution of ability between classes in a school) to within group variance (distribution of ability within the school overall). Hence, schools that engage in purposeful sorting of students based on prior achievement should have values on this measure below unity, the ratio expected when classes are randomly formed. We chose to employ this metric rather than an F-statistic because we are more interested in the intuitive properties of the distribution of prior ability rather than a statistical test.

Following an investigation of the distribution of this metric, we then compare this index to three different teacher-level VA estimates. The initial VA specification (referred to as VA1) includes two years of prior student achievement, mean prior achievement for the class, ELL status, and an indicator for low SES. We use qualifying for free-and-reduced lunch in BPS and Title I status in CUSD and NPS as indicators for SES. This choice of model specification is motivated in part by earlier research (Author, 2011) but modified to allow for differences in data availability across districts. The other two VA specifications (VA2 & VA3) are derivatives of the first: VA2 only includes one year of prior achievement at the student level, and VA3 retains two years of prior achievement, but drops mean prior achievement for the class. We consider the school effects as fixed rather than random and estimate all models with custom software available from the authors.

We then compare our sorting index to teacher-level VA estimates in order to answer our final research question. We use the following multilevel model to see if any variation in VA estimates is dependent on sorting.

$$VA_{jJ} = \beta_1 + \beta_2 \text{sort}_J + \zeta_J + \varepsilon_{jJ}\ . \quad (3)$$

Note that we allow for variability in VA estimates at school *J* via the random intercept ζ_j. The key estimate is the estimate of β_2 which identifies whether variation in VA estimates is related to our sorting metric. We also use this model to estimate the ICC and analyze the amount of variability within versus between schools in each district.

4. Results

Ability Sorting

Figure 1 shows histograms of the sorting metric based on each subject in each district, respectively. We compute the metric for a school separately each year, so schools appear multiple times in the distribution of the sorting index for CUSD as there are multiple years of grade 5 data. Recall that values on this metric closer to unity indicates less evidence of sorting, while lower values indicate more evidence of sorting within the district. Each of the histograms indicates values greater than one as well. These are typically situations in which schools, perhaps small schools, have a few larger classes which are more varied in ability distribution than the school overall.

Similar to the general demographics of the three districts used in this study, Bayview and Central appear to be more similar in sorting practices than Norton. While Central appears to sort more on the whole based on the IQR than either of the other two districts, both Central and Bayview demonstrate much larger variation in sorting than Norton. There also appears to be more sorting based on reading scores in Central and Bayview. Sorting practices are so minimal in Norton that differences in sorting based on subject is negligible.

The differences in sorting practices between Norton and both Central and Bayview lead to questions regarding why there are such extreme differences. One possible reason could simply be less variability in prior achievement in Norton than in Central and Bayview. Unfortunately, since these districts are in three different states, it is not possible to know if this is true or not since the scores do not come from a common scale. Another potential reason for the differences in sorting practices within schools may be the fact that students are sorted into schools as well. If district lines are drawn such that students are already homogenously grouped into schools, this could lead to less sorting within schools.

In order to identify the amount of sorting between schools compared to within schools, we ran an F-statistic by subject on prior achievement for each district (See Table 2). Since there appears to be so little evidence of sorting in Norton, we might expect the district to have a much larger F-statistics than Bayview and Central. This would suggest that the ratio of between group variability to within group variability is much larger in Norton and could be a reason as to why there is less sorting into classrooms. That is, if students are already sorted by ability between schools, there is little need to sort them within schools. However, what we actually see is lower F-statistics in Norton for both math and reading than the other two districts. District lines appear to sort students into schools as indicated by our metric more so in Bayview and Central than they do in Norton. The reasons for these differential practices are unavailable from the current analysis, but provide fodder for further research regarding sorting practices and their relationships to school and district characteristics.

Characteristics of Sorting

Having established that sorting practices vary across districts, we now inquire whether the degree of sorting within a district (across schools) varies as a function of school demographics. Table 3 contains correlations between the sorting metric and the proportion of ELL and low SES students within a school. The correlations show little evidence of strong linear relationships between the sorting metric and ELL or SES status. That is, schools with higher proportions of ELLs or students from low SES backgrounds do not tend to sort more than those with lower proportions.

Prior work (Author, 2011) suggests a relationship between ELL status and sorting on prior ability. In Bayview and Central, larger percentages of ELL students within a school are associated with increased sorting (lower values on the sorting index). Given that there are so few ELL students in Norton, the failure to find a relationship there is not surprising. It is possible that the correlation obscures non-linear relationships between these two quantities. Figure 2 depicts the sorting metric as a function of the percentage of ELL students within a school. The thick line is based on a LOESS regression of the sorting index on the % off ELL students within a school. We see that both Bayview and Central include schools which serve close to 0% as well as 100% ELLs. Compared to schools that serve no ELL students, there is typically a slight increase in sorting for those schools that have small populations of ELL students. This increase is more dramatic in Reading. This finding seems plausible as districts often prioritize language intervention for ELLs. Intensive focus on developing English skills is often considered a primary way of improving student achievement across disciplines.

The proportion of ELLs in a school seems to have some relationship to sorting. Schools which consist of nearly all ELL students as well as those with close to no ELL students tend to sort less on ELL status than those schools with more linguistically diverse student populations. This finding makes sense given that if either none of the students or all of the students are ELLs, then it is impossible to sort based on this characteristic. In schools with linguistically mixed student bodies, classes may be formed based in part on language status.

Sorting and VA estimates

Finally, we explore the relationship between sorting and value-added estimates using multilevel modeling (Equation 3). The multilevel model results for all three of our VA specifications in each district and subject are presented in Table 4. We report the main parameters from these models: the ICC, estimated slope coefficient on the sorting parameter, and the standard error of that estimate. Recall that VA1 controls for two years of prior student achievement as well as the mean prior for all students in a given class. VA2 uses only one year of prior achievement alongside mean prior achievement, and VA3 retains two years of prior achievement, but drops the mean prior. The ICC indicates that variation in VA estimates due to school assignment is fairly consistent across VA models, averaging roughly 0.3 across all three VA specifications. On average, we see under one-third of the distribution in VA estimates due to school assignment, that is, there is more variability in teacher VA estimates within schools than across schools. This is relatively consistent across all districts and subjects.

When interpreting the estimates for school sorting, it is important to recall that a lower value on the sorting metric indicates more sorting. So, negative coefficients indicate that teachers in schools with lower scores on the sorting metric (and hence more sorting) tend to have higher value-added estimates, and those instances with positive estimates indicate situations where less sorting is related to lower value-added estimates. In all models, the coefficient on school sorting is quite small and not significantly different from zero, so we see no notable relationship between any of our VA estimates and sorting metrics across models, districts, and subjects. Sorting as we have characterized it does not appear related to our estimates of teacher effectiveness and estimates of teacher effectiveness do not appear dependent on the school in which a teacher works. Both of these findings support prior work suggesting that VA estimates are not sensitive to student or teacher assignment.

5. Discussion

We find noticeable variability in sorting across both Bayview Public Schools and Central Unified School District, but little evidence of sorting overall in Norton Public Schools. While sorting has weak correlations with proportions of ELLs and students from lower-SES backgrounds across the districts, those districts that serve significantly larger populations of ELLs and students from low SES backgrounds do tend to sort on prior achievement far more than NPS, which has noticeably smaller proportions of these subgroups. Sorting appears to happen more frequently based on reading achievement than other subjects in both Bayview and Central, and the little sorting that occurs in Norton is indistinguishable between subjects.

It also appears that schools with more linguistically diverse student populations tend to sort more based on ELL status than those schools that are more homogenous regarding ELL status. Even NPS follows a similar pattern of sorting less at schools with higher and lower proportions of ELLs in the district. The differences in sorting patterns across these districts parallel the differences in demographic make-up of the student bodies warrant further investigation to understand more about the motivation behind the sorting practices (or lack thereof) within these locations.

Our proposed sorting metric is related to only small changes in VA estimates, and those estimates are not significantly different from zero. It seems that these VA specifications, based on prior work in the field, adequately control for the distribution of prior student achievement into classes within a school. While this finding corroborates other findings, it is important to continually investigate potential sources of bias in VA estimates as they are increasingly used in high-stakes decisions. It is also important to remember, however, that the sorting metric employed here only accounts for prior student ability. Accounting for additional sorting variables (i.e. discipline records, SES, gifted and talented status, and special education status) could uncover greater influence of non-random sorting on VA estimates.

The variability in the sorting metric implies there is something happening at the school level to sort students, but the mechanism by which this happens is not consistent across schools or districts. Although researchers may want to look at schools and districts to gain an understanding of sorting practices, this does not appear to greatly influence VA estimates. Particularly notable is that the district with very low proportions of ELLs and students from low-SES backgrounds seems to engage in sorting in a very minimal capacity while the districts studied here with relatively high proportions of the same sorts of students tend to sort more and have more variability in sorting across the district.

6. Conclusion

Ability sorting, as explored here, is a commonly recognized practice. The degree to which this occurs varies both within and between school districts. More investigation is needed to understand why certain schools and districts choose to sort students based on ability and why others do not and the influence these sorting practices may have on students and teachers.

This inquiry finds little evidence of sorting influencing teachers through their VA estimates across three urban districts in three different states, but the investigation was limited to a single sorting metric. Focusing on this relationship between VA estimates and distribution of student ability is timely as increasing numbers of personnel decisions, pay for performance programs, and rewards are being linked to teacher evaluations. Further investigation is needed to understand how sorting may influence school or classroom practices. It is also important to identify additional mechanisms for sorting practices to ensure that other characterizations of sorting are not needed in value-added models.

If nonrandom sorting of students has an impact on teacher VA estimates are not taken into account, education could be shortchanged. High-quality teachers may be less willing to work in schools or with students that are likely to lower their effectiveness ratings. As VA estimates are increasingly related to high-stakes decisions, current or potential teachers may abandon the profession altogether. Evidence of the existence and influence of non-random sorting is seen in other work, but the specific mechanisms by which this practice influences VA estimates remains unclear. This uncertainty should incite researchers investigate additional sorting mechanisms and cause decision-makers to carefully consider their inclusion and use of VA estimates for teacher evaluation and the stakes to which these estimates are tied.

References

Author, 2011.

Baker, E., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., et al. (2010). *Problems with the use of student test scores to evaluate teachers*. *Economic Policy*. Washington, D.C. Retrieved from http://www.epi.org/page/-/pdf/bp278.pdf

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440. doi:10.3102/0162373709353129

Clotfelter, C. T., Ladd, H. F., &Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. Retrieved from http://www.nber.org/papers/w11936.pdf?new_window=1

Collins, C. A., &Gan, L. (2013). *Does sorting students improve scores? An analysis of class composition*. National Bureau of Economic Research. NBER Working Paper 18848. Retrieved from http://www.nber.org/papers/w18848

Goldhaber, D., Walch, J., &Gabele, B. (2012). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. Center for Education Data and Research. Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202012-6_Does%20the%20Model%20Matter.pdf

Harris, D. N. (2009). Would accountability based on teacher value-added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy*, 4(4), 319–350.

Koedel, C., & Betts, J. R. (2009). Does student sorting invalidate Value-Added Models of teacher effectiveness? An extended analysis of the Rothstein Critique. Retrieved from http://economics.missouri.edu/working-papers/2009/WP0902_koedel.pdf

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Rand Corporation. Retrieved from http://www.rand.org/pubs/monographs/MG158.html

Oakes, J. (2005). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Rivkin, S. G. (2007). Value-Added analysis and education policy. National Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.urban.org/uploadedpdf/411577_value-added_analysis.pdf

Rothstein, J. (2010). Teacher quality in educational production: tracking, decay, and student achievement. *Quarterly Journal of Economics*, (February), 175-214.

Wells, A. S., & Serna, I. (1996). The politics of culture: Understanding local political resistance to detracking in racially mixed schools. *Harvard Educational Review*, *66*(1), 93-118.

Zimmer, R. (2003). A new twist in the educational tracking debate. *Economics of Education Review*, *22*, 307-315. doi:10.1016/S0272-7757(02)00055-9

Table 1.Enrollment& Demographics-Based on grade 5 students in specified years with at least two prior test scores.

	Year	Unique students	Unique teachers	Unique schools	% ELL	% Low SES*
Bayview	2012	4,203	269	86	42	*72*
Central	2005-2009	178,142	3,940	481	41	90
Norton	2012	8,721	422	92	5	26

*Note: In Central&Norton, Title I is used to identify low SES. In Bayview, Title I was unavailable, so FRL is used instead.

Figure 1. Sorting metric distribution in each district and subject respectively: On this metric, lower values correspond to more ability sorting across classrooms.

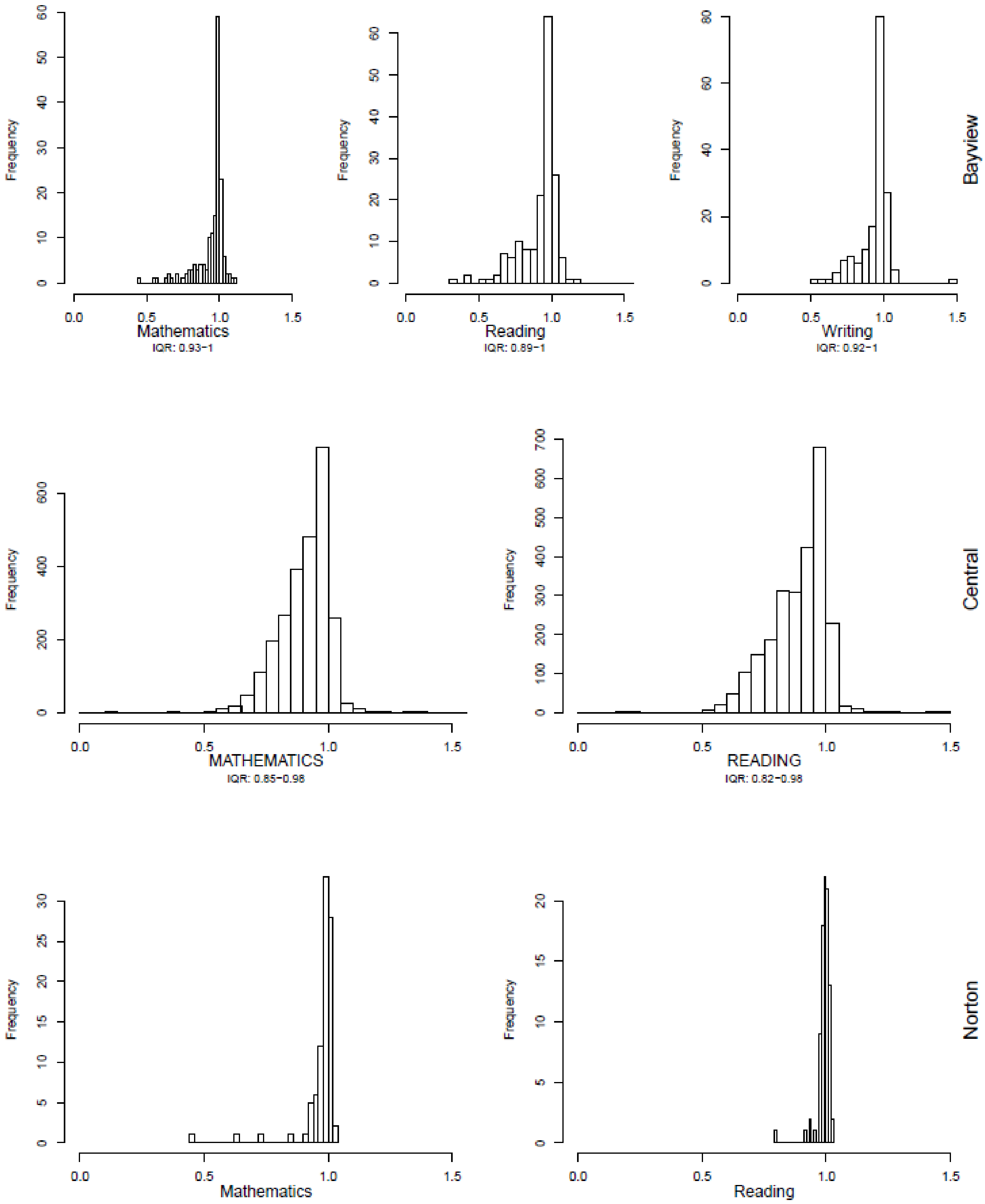

Table 2. Prior achievement F-statistics

	Bayview	Central	Norton
Math	18	15	14
Reading	16	18	12
Writing	21	NA	NA

Table 3. Correlations between sorting metrics and student characteristics by subject

	Bayview		Central		Norton	
	FRL	ELL	Title I	ELL	Title I	ELL
Math	.08	.08	-.08	-0.11	.-0.02	-0.12
Reading	-.09	-0.08	-.09	-.160	-0.05	0.04
Writing	-.13	-.15	NA		NA	

Figure2. Sorting and proportion of ELL students in each district and subject respectively: Line indicates a LOESS fit line. These suggest that, in general, schools with the lowest and highest proportion of ELLs do not tend to sort as much as those with a more linguistically heterogeneous student body.

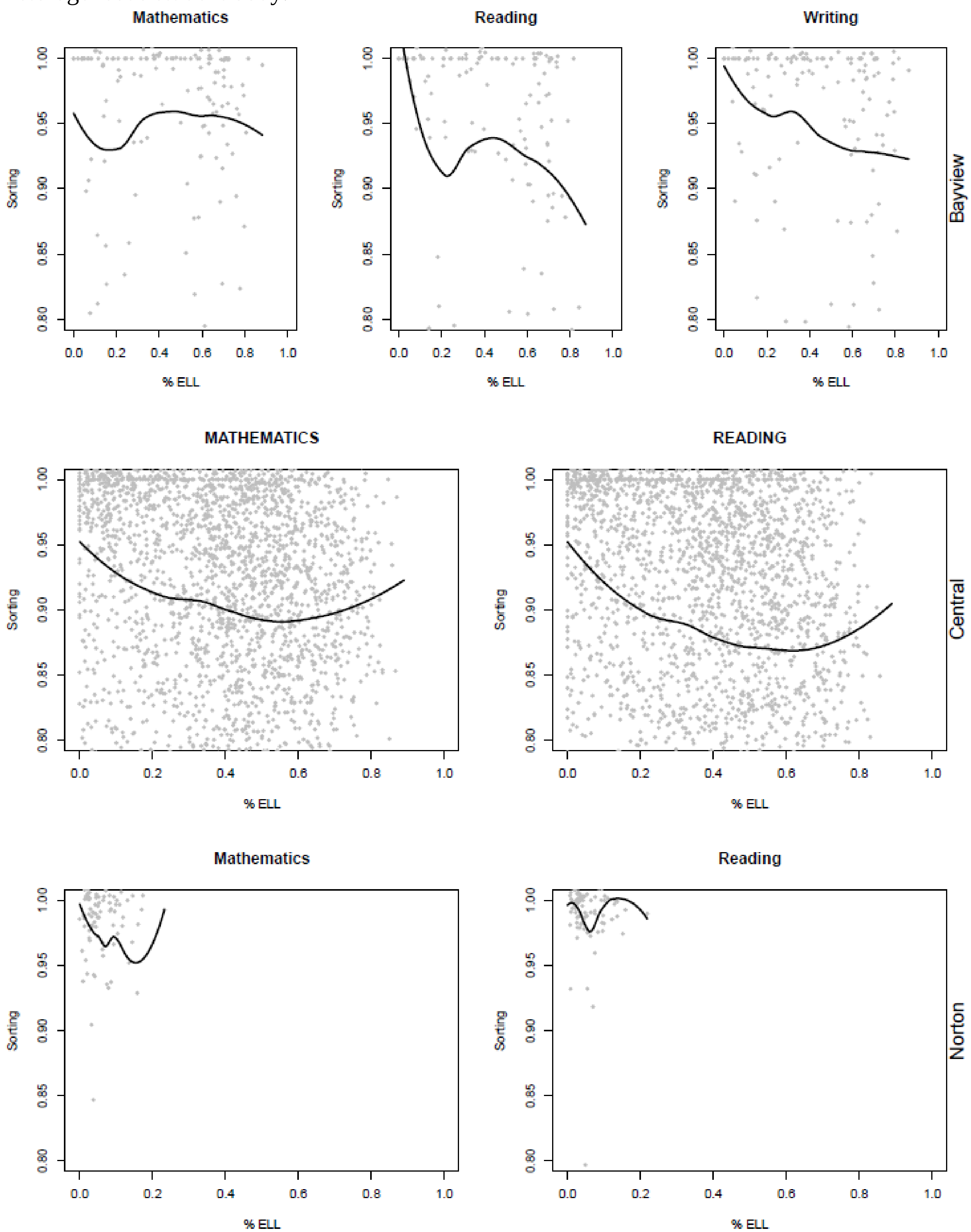

Table 4.HLM Results

VA Specification	Subject	District	ICC	Estimate	SE
VA1	Mathematics	Bayview	0.42	-0.05	0.10
	Mathematics	Central	0.21	0.01	0.03
	Mathematics	Norton	0.25	0.19	0.14
	Reading	Bayview	0.46	0.00	0.05
	Reading	Central	0.22	0.01	0.01
	Reading	Norton	0.21	-0.18	0.22
	Writing	Bayview	0.47	0.08	0.11
VA2	Mathematics	Bayview	0.36	-0.12	0.10
	Mathematics	Central	0.21	0.04	0.03
	Mathematics	Norton	0.25	0.14	0.14
	Reading	Bayview	0.33	-0.01	0.06
	Reading	Central	0.21	0.01	0.01
	Reading	Norton	0.24	-0.23	0.24
	Writing	Bayview	0.32	0.02	0.10
VA3	Mathematics	Bayview	0.42	-0.09	0.10
	Mathematics	Central	0.20	0.00	0.03
	Mathematics	Norton	0.25	0.20	0.14
	Reading	Bayview	0.43	-0.01	0.05
	Reading	Central	0.21	0.01	0.02
	Reading	Norton	0.24	-0.18	0.23
	Writing	Bayview	0.43	0.08	0.11