

Accepted Version for Archival. Published version can be cited as:  
Curran, F. C., Little, M. H., Cohen-Vogel, L., & Domina, T. (2018). School Readiness Assessments for Class Placements and Academic Sorting in Kindergarten. *Educational Policy*. Online First.

**School Readiness Assessments for Class Placements and  
Academic Sorting in Kindergarten**

F. Chris Curran\*

*University of Maryland, Baltimore County*

1000 Hilltop Circle

411 Public Policy

Baltimore, MD 21250

curranfc@umbc.edu

\*Corresponding author

Michael Little

Peabody Hall

UNC Chapel Hill

Chapel Hill, NC 27599

mhlittle@live.unc.edu

Lora Cohen-Vogel

Peabody Hall

UNC Chapel Hill

Chapel Hill, NC 27599

lora.cohen-vogel@unc.edu

Thurston Domina

Peabody Hall

UNC Chapel Hill

Chapel Hill, NC 27599

tdomina@email.unc.edu

**Abstract:** Kindergarten readiness assessments are commonly used in schools nationwide. Prior work shows that the use of such assessments for class placement decisions has increased in recent years. This paper uses the Early Childhood Longitudinal Study of 2011 to explore whether the use of readiness assessments for such purpose predicts differential sorting of students across classrooms by prior academic ability. Results from multilevel models as well as other sensitivity analyses suggests that the use of readiness assessments for classroom placements is predictive of slightly higher cross-class ability sorting, particularly in English/language arts. The implications for policy and practice are discussed.

**Keywords:** sorting, tracking, data use, elementary, ECLS

## School Readiness Assessments for Class Placements and Academic Sorting in Kindergarten

### Introduction

In recent years, the use of school readiness assessments has been a ubiquitous feature of the kindergarten experience. The number of states with laws requiring kindergarten students to take readiness assessments has increased substantially over the past decade and a half, with recent estimates suggesting that as many as 26 states require such assessments (Little & Cohen-Vogel, 2017; Council of Chief State School Officers, 2010; Education Commission of the States, 2014; Gewertz, 2014 October 7). Regardless of state laws, administrators in nearly three in four of the nation's elementary schools reported administering a school readiness assessment to students in kindergarten during the 2010-11 school year (Little, Cohen-Vogel, & Curran, 2016).

Kindergarten readiness assessments are instruments designed to assess various aspects of a student's preparedness for the formal school environment. The assessments vary in format, content assessed, length, and time of administration across states and districts. Some are administered before the start of the school year, while others are administered after school begins (Conn-Powers & Peters, 2007; Education Commission of the States, 2014; Gewertz, 2014 October 7; Public Citizens for Children and Youth, 2015). In general, however, the assessments seek to provide educators with more information on the academic and, in some cases, social readiness of students for formal schooling. Data from readiness assessments are reportedly used for various purposes, including to individualize instruction and to identify students who may need additional testing (Conn-Powers & Peters, 2007; Little et al., 2016).

New evidence, however, shows large increases in the reported use of readiness assessment data for one purpose, in particular: to determine classroom placements (Little et al.,

2016). Using data from the Early Childhood Longitudinal Study, Little and colleagues (2016) find that the use of readiness assessment scores for this purpose increased by almost fifty percent between 1998-99 and 2010-11. By 2010-11, 4 in 10 administrators in schools that administered readiness assessments reported using them to place students in classes (Little et al., 2016).

The implications of using readiness assessment data in this way are unclear. On the one hand, it is possible that schools might use scores from the assessments to distribute students evenly between classrooms—in essence, creating heterogeneous classes with regard to students' incoming abilities. On the other hand, schools might use test scores to facilitate homogeneous sorting, potentially segregating lower ability students from their higher performing peers. Given the widely accepted findings that students' peers and teachers have a significant impact on their academic achievement (e.g., Rockoff, 2004; Sanders & Rivers, 1996), it is our contention that decisions about how students are grouped into classes are high-stakes. The implications of sorting in kindergarten are particularly salient given evidence that achievement and experiences during kindergarten are predictive of both proximal and later life outcomes (Chetty et al., 2011; Claessens & Engel, 2013; Claessens, Duncan, & Engel., 2009; Claessens, Engel, & Curran, 2014).

In this paper, we explore the relationship between administrators' reported use of kindergarten readiness assessments for class placement and what we refer to here as “cross-class sorting” by academic ability. To do so, we draw on data from the newest iteration of the Early Childhood Longitudinal Study (ECLS-K:2011). Before presenting the findings, we review the literature on kindergarten readiness assessments and ability sorting in education, present our guiding framework, and describe the data and methods used in the study.

### **Prior Literature**

## **Kindergarten Readiness Assessments**

We begin our review of the literature by providing an overview of the relevant research on kindergarten readiness assessments. Kindergarten readiness assessments vary widely in the specific form that they take but are typically designed to measure students' academic skills at the outset of formal schooling. Some assessments consist of brief inventories of early academic skills (e.g., number of letters and numbers known). Others are extensive formative assessment systems, administered over a number of days or weeks, that capture children's readiness across a range of developmental domains (Little & Cohen-Vogel, 2017; Snow & Van Hemel, 2008). This latter conception of readiness assessments is considered best practice in the field and was adopted by the U.S. Department of Education in its Race to the Top Early Learning Challenge program (Merrill, et al., 2018; Snow & Van Hemel, 2008).

Existing research on kindergarten readiness assessment is variable in approach. Some studies focus on the technical details entailed in the assessment of young children, such as validity and reliability (Epstein, Schweinhart, DeBruin-Parecki, & Robin, 2004; Scott-Little & Niemeyer, 2001). Others have sought to define the appropriate form and use of kindergarten readiness assessments. A 2008 National Research Council (NRC) report stressed the need to develop high-quality assessment tools to measure children's school readiness along a holistic range of developmental domains. The report, which included over 50 experts from around the United States, cautioned against the use of readiness assessments for high-stakes decisions, such as evaluating or delaying children's entry into kindergarten. Instead, the report stressed that readiness assessment data should be used to provide information to teachers so they may identify the needs of children, individualize instruction, and drive program improvement (Snow & Van Hemel, 2008). Despite this caution, recent studies indicate that data from kindergarten readiness

assessments are used in high-stakes ways that are inconsistent with NRC recommendations (Little et al., 2016; Little & Cohen-Vogel, 2017).

Little and colleagues (2016) documented the use of kindergarten readiness assessments using nationally representative data, and the findings served as a key motivation for the present study. In that study, Little and colleagues (2016) analyzed data from the Early Childhood Longitudinal Study- Kindergarten Class of 2010-11 to document (1) how prevalent the use of kindergarten readiness assessments are and (2) how school leaders reported using the data from these readiness assessments. During the 2010-11 academic year, 72 percent of school administrators reported that their school used some form of kindergarten readiness assessment. Of those schools using kindergarten readiness assessments, 41 percent of school administrators reported that they use the data to determine children's class placements. Further, this finding marks an almost 50 percent increase (13 percentage points) from the previous administration of the same survey in the 1998-99 school year. These results motivate our investigation into the link between kindergarten readiness assessments and sorting. We now turn to review the literature on ability sorting.

### **Ability Sorting**

Ability sorting represents the practice of grouping students by academic ability. In the United States, ability sorting has typically taken one of two forms: within-class sorting or cross-class sorting (Cohen-Vogel & Rutledge, 2009). *Within-class sorting* refers to the practice of enrolling students of different abilities in the same classroom but then grouping students by ability for various aspects of instruction or practice (Gamoran, 1992). Within-class sorting is very common in elementary schools, particularly for reading instruction (Loveless, 1998), and is often referred to simply as "ability grouping" (Loveless, 2013). A number of studies, including

some using the original Early Childhood Longitudinal Study, find generally positive relationships between within-class sorting and student outcomes in elementary school (McCoach et al., 2006; Kulik, 1992; Slavin, 1987; Tach & Farkas, 2006), though other studies find the relationship to vary for students of different abilities or in different contexts (Hong et al., 2012; Kulik, 1996, Lou et al., 1996). For instance, Garrett and Hong (2016) find some negative effects on English language learners in classrooms that predominantly utilize homogenous within-class ability sorting.

The second type of ability sorting, *cross-class sorting*, refers to the practice of dividing students into homogenous classrooms based on ability. Often, and particularly in the upper grades, this practice involves placing students into different courses, such as routing lower ability students into remedial mathematics and higher ability students into college preparatory mathematics (Lucas, 1999; Oakes, 2005). The practice of cross-class sorting, particularly when coupled with exposure to different curricula, is commonly referred to as “tracking” or schoolwide “cluster grouping” (Lucas, 1999; McCoach et al., 2006; Oakes, 2005; Matthews et al., 2013).

This study focuses on the latter form of sorting, namely cross-class ability sorting. We use the term “cross-class ability sorting” rather than “tracking” to refer to this practice, as “tracking” tends to connote exposure to different curricula, a practice that is uncommon at the kindergarten level. Some scholars conceptualize sorting students by ability as promoting inequality (Oakes, 2005; Oakes & Guiton, 1995). They argue that students, once placed in lower ability groups, may be denied opportunities to move into higher ability groups, making schools sources rather than disruptors of social stratification (e.g., Condrón, 2007; Lou et al., 1996). Others, however, promote ability sorting as a technique for optimally challenging students of

different developmental levels and abilities (Allan, 1991; Kulik, 1992). They contend that ability grouping benefits students by allowing teachers to adapt “the depth, pace, or difficulty level of a lesson to meet the needs of an individual or a small group of students (Passow, 1962; Tomlinson, 1995; 1999)” (McCoach et al., 2006, 340).

The evidence suggests that the practice of cross-class ability sorting is common in schools but that there is significant variation in the use of this practice across schools and districts (Collins & Ga, 2013; Dieterle et al., 2015; Horvath, 2015). For instance, Dieterle and colleagues (2015) show wide variation in the use of cross-class ability sorting in elementary schools across a single state and find that certain factors, such as serving higher percentages of students with limited English proficiency, are predictive of greater sorting. Similarly, it is possible (though not yet empirically examined) that other contextual factors, such as the extent to which preschool settings are formalized into the public school system, may also contribute to variation in sorting practices across elementary schools.

Cross-class ability sorting may also explain other sorting patterns. A large body of evidence documents the sorting of students by race and family socioeconomic status (e.g., Mickelson & Everett, 2008; Oakes, 2005). Recent research, however, shows that factors such as race become significantly less predictive of sorting after conditioning on prior academic achievement (Dieterle et al., 2015; Kalogrides & Loeb, 2013) and that structural components related to tracking may explain some differences in race and socioeconomic sorting (Mickelson & Everett, 2008).

The research on the impacts of cross-class ability sorting is mixed. Systematic reviews of the research on ability sorting find no relationship between the use of cross-class ability sorting and student achievement (Slavin, 1987; Steenbergen-Hu et al., 2016). In part, this may be

because the practice can be implemented in different ways, some of which may promote positive student outcomes while others may be harmful. For instance, Nomi and Allensworth (2014) use the term “skill based sorting” to describe the practice of sorting students by skill or ability yet still exposing students to the same curriculum. In theory, such a practice allows for customization of instruction to meet the students’ particular needs without sacrificing rigor. Indeed, some work examining cross-class ability sorting finds positive relationships with student achievement (Collins & Gan, 2013; Matthews et al., 2013).

In contrast, however, cross-class ability sorting is often implemented in a manner in which lower ability students are not exposed to the same rigor of instruction or the same quality of instructor. Even when exposed to similar instruction and quality of instructor, exposure to high concentrations of lower achieving peers may also negatively affect students in the lowest skills classrooms. Kalogrides and Loeb (2013) examine sorting in both elementary and high schools, finding that lower performing students are both more likely to be placed with a less experienced teacher and more likely to be exposed to lower achieving peers. Given evidence on the returns to teacher experience (Clotfelter et al., 2006; Harris & Sass, 2011; Kane et al., 2008; Ladd & Sorensen, 2016) and the potential influence of peers on achievement (Gottfried, 2014; Hanushek et al., 2003; McEwan, 2003), this finding suggests that cross-class ability sorting may negatively affect lower performing students.

Indeed, work has demonstrated that tracking can propagate differences between groups and widen gaps in educational attainment between more and less affluent students (e.g., Gamoran & Mare, 1989). Other studies have shown a relationship between tracking and students’ levels of self-esteem, though the relationship varies based on the type of tracking and ability level of the student (Chmielewski et al., 2013). In short, depending on its implementation,

the use of cross-class ability sorting may have both positive and negative influences on student outcomes.

Despite the potentially important impact of cross-class ability sorting on student outcomes, little research has examined the factors that predict the use of this form of sorting in elementary schools. This is true despite increases in the amount of information collected on the ability of incoming kindergarten students. In past decades, elementary schools may have been dependent on parental input or informal input from preschool providers for gauging incoming student ability. Kindergarten readiness assessments potentially provide school personnel with increased information on student ability, allowing for easier implementation of cross-class ability sorting. Coupled with a heightened focus on the use of data to guide decision making, such information on incoming student ability may potentially serve to alter the process by which students are placed into classrooms. In the next section, we turn to a formal theoretical framework that details the potential mechanisms by which the use of readiness assessments may impact classroom placement decisions.

### **Using Data to Sort or to Mix: The Push and Pull of Theory**

Those who predict that sorting students by ability will lead to differentiation in curricula, pacing, and depth of content coverage which, in turn, will lead to performance improvements for all students have been emboldened by recent developments in data collection and analytics. Theoretically speaking, new and expanded student-level data systems and the reports generated from them offer teachers better information about individual student needs, thus improving the potential for appropriate curricular differentiation (Cohen-Vogel & Harrison, 2013).

Based on the assumption that schools use assessment data to facilitate effective curricular differentiation, one might expect to see more cross-class ability grouping in schools that report

using readiness assessments in the classroom assignment process. We label this expectation the *sorting hypothesis*. We test it by examining the extent to which students at similar levels of incoming achievement cluster together in the same classrooms in schools that use readiness assessments for classroom placements.

It is not a given, however, that the use of readiness assessments for classroom placements increases the clustering of students in classrooms. Research suggests that even as districts are gradually improving supports for data use (e.g., professional development), many educators still feel unprepared to use data to engage in activities that range from interpreting test scores to adjusting the curriculum (Gallagher et al., 2008; Means et al., 2009; Murnane et al., 2005; Sharkey & Murnane, 2003) (see also Wachen et al., 2016). Lack of preparedness or suspicions about validity from assessment data may limit behavioral responses to data availability. Further, professional organizations in the field of education express a great deal of caution about the use of these assessments for classroom placements. Indeed, just how schools might use data produced by kindergarten readiness assessments was the subject of a 2008 report by the National Research Council (NRC). As we discussed above, the NRC report cautioned against using assessments for high-stakes decisions, such as delaying children's entry into kindergarten (Snow & Van Hemel, 2008). We submit that cross-class sorting may similarly be high stakes. These institutional constraints may condition whether data—like those generated from school readiness assessments—are used to sort students into classes on the basis of ability. Institutional constraints, like organizational norms, routines, and power structures, are so strong in fact that educational policy is said to rely on changing particular aspects of schools without altering school as an organizational form (Ogawa, 2009). Theorists (see, e.g., March & Olsen, 1984;

Powell & DiMaggio, 1991) posit that organizations possess powerful conserving forces, forces that in the words of Mark Smylie “make persistence paramount to change” (1995, 6).

In terms of using data for classroom placement, school leaders would arguably be confronting widely-held institutional norms that emphasize fair treatment. Indeed, in a study by Osborne-Lampkin and Cohen-Vogel (2014), principals reported using test score data as they assigned students to classes; they did so, however, not to create homogenous classrooms based on ability but instead to ensure within-class heterogeneity and a sense of “fairness” among teachers. According to the authors, “educators in the study used student demographic and performance information to populate classrooms that *intentionally spread high, middle, and low performers equally across grade-level teachers*” (p. 204). Similarly, recent qualitative work on the use of data in elementary schools documents the use of such data to create within-class heterogeneity in classrooms in the upper elementary grades (Park & Datnow, 2017). Accordingly, we articulate a competing null hypothesis, which holds that the use of readiness assessments in classroom placements is not associated with differential distributions of student achievement across kindergarten classrooms.

Together, lessons from theory provide little in the way of clear hypotheses about what we will find as we examine whether schools that report using data for class placements do so in ways that increase cross-class ability sorting. Evidence in support of the *sorting hypothesis* would argue that better information about individual students is seen by educators as improving the potential for appropriate curricular differentiation through ability grouping. Alternatively, evidence in support of the *null hypothesis* would be indicative of the powerful conserving forces in schools that lead educators to use student assessment data in ways that make classrooms more heterogeneous.

## Methods

### Data Source

We use data from the newly released ECLS-K:2011. The ECLS-K:2011 is an ongoing longitudinal study that began with a nationally representative sample of approximately 18,000 kindergartners from about 970 schools in the 2010-2011 school year. Administered by the Institute of Education Sciences (IES), the ECLS-K:2011 includes a wide variety of variables on the school and classroom environment, student achievement, and student and family characteristics (Tourangeau et al., 2015). Importantly, for the purpose of this study, the dataset includes indicators of whether a school used a kindergarten readiness assessment and, if so, whether the assessment results were used for class placements. Combined with student standardized achievement score measures, this dataset allows for an exploration of the relationship between readiness assessment use for class placements and classroom sorting by academic ability.

The ECLS-K:2011 employed a complex, multi-stage sampling design, where 90 counties or groups of counties were first sampled, followed by public and private schools within each PSU, and, finally, kindergarten children in each school. The data used in the present study were collected in the fall and spring of the 2010-11 academic year. We used sampling weights provided in the dataset in order to account for differential probabilities of selection into the sample and to adjust for the effect of nonresponse (Tourangeau et al., 2015).

In order to account for the potential bias introduced by large amounts of missing data in analyses of complete-case data, we performed multiple imputation to estimate missing values. Multiple imputation is a method that replaces missing values with multiple simulated values to complete a dataset. Once estimated, analyses are conducted that adjust the parameter estimates

for the uncertainty of replacing missing values (Rubin, 1996). Specifically, we imputed missing values using the chained equations methodology and generated 10 imputed datasets (Cameron & Trivedi, 2009; Royston & White, 2011).

**Readiness assessment use.** The primary measure of readiness assessment use for class placements comes from the school administrator survey that was fielded in the fall of 2010. Each administrator was asked whether his or her school administers a readiness assessment; if so, the respondent described whether or not the assessment data are used for each of the following six purposes: (1) to determine eligibility for enrollment when a child is below the cut-off age for kindergarten, (2) to determine children's class placements, (3) to identify children who may need additional testing (i.e. for a learning problem), (4) to help teachers individualize instruction, (5) to support a recommendation that a child delay entry, and (6) other. Our primary independent variable is a binary indicator that represents schools that reported using a readiness assessment *and* using the data from it to determine children's class placements. In our analyses, we compare these schools to those that gave a kindergarten readiness assessment but did not use it for class placement decisions as well as a wider group of schools that also includes those that did not give a readiness assessment at all. Descriptive statistics on the key independent variables are shown in Table 1.

A limitation of our measure of readiness assessment use for class placement is that we do not have specific information about the type of readiness assessment used, the extent to which the assessment measures traditional academic skills, or whether the assessment is universal or targeted in terms of administration. Some states administer a readiness assessment to a random sample of entering kindergarteners (Stedron & Berger, 2010). A second limitation is that since we do not have data on the timing of assessments, we are unable to distinguish between schools

in which assessments occur before the start of instruction in kindergarten and those that occur after the start of instruction, and are thus less likely to influence kindergarten classroom assignments. Additionally, beyond state sanctioned readiness assessment policies, individual schools may decide whether or not to administer certain screening assessments if a learning disability is suspected. Each of these data limitations likely imposes a conservative bias on the observed relationship between assessment use and classroom placement practices. That said, to our knowledge, these are the only nationally representative data linked to children that capture information about schools' use of readiness assessments.

**Measures of student achievement.** Measures of student achievement come from direct one-on-one student assessments in math and reading. The mathematics component of the cognitive assessment measured skills in conceptual knowledge, procedural knowledge, and problem solving. Specifically, the assessment captured achievement in number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and patterns, algebra, and functions. All children were given a common set of routing questions at the beginning of the assessment and were then assessed with a set of high, medium, or low, difficulty questions. Because all children did not receive the same set of questions, we used vertically aligned item response theory (IRT) scores for analysis (Tourangeau et al., 2015).

The reading component of the assessment included questions that measure basic skills, such as print familiarity and letter recognition, as well as vocabulary knowledge and reading comprehension. The reading comprehension questions required children to define words, provide supporting detail, and make complex inferences. Like the mathematics assessment, we used IRT scores provided in the data file for analysis. Both the reading and math assessment scores are

highly reliable, with theta scores of 0.95 and 0.92 for reading and math, respectively (Tourangeau et al., 2015). Descriptive statistics on the dependent variables are shown in Table 1.

**Control variables.** Our models included a range of control variables to account for potential differences in district and school characteristics. At the district level, we included a variable indicating whether or not the school is located in an urban area and a measure of the percentage of school-aged children in the district that live in poverty. At the school level, we included variables that measure the percentage of students who received free or reduced lunch, are a member of a racial minority group, and are English Language Learners. Indicator variables capture whether or not the kindergarten program is full day or half day, whether the school met Annual Yearly Progress under No Child Left Behind, whether the school is public or private, received Title 1 funding, and has a preschool program located within the elementary school building. The latter of these, co-location of a preschool program, serves to address differential sorting patterns that could arise due to differences in information provided by early care providers that are more or less tightly coupled with the elementary setting. In addition to these control variables, we also control for school-wide variance in standardized mathematics and reading achievement scores. Because our analyses focus on school level indicators of the use of readiness assessments for placement decisions, our control variables focus on school level measures. Where we do control for lower level characteristics, we aggregate to the school level (such as percentage of students qualifying for FRPL rather than whether an individual student qualifies). Descriptive statistics on each of the covariates are included in Appendix Table A1.

## **Analysis**

We use several different strategies to assess the degree to which the use of readiness assessments for class placement is related to cross-class ability sorting. Our primary analytical approach relies on variance components generated from hierarchical linear models (HLM). HLM is form of regression analysis that accounts for the non-independence of observations within clusters, which in this case includes students nested within classrooms and classrooms nested within schools (Raudenbush & Bryk, 2002). Specifically, HLM allows the intercept within each nested cluster to vary randomly. Variance components, which are measures of the degree to which the intercept differs between groups in each level of the model are our key measures of sorting. Specifically, we use a three-level random intercept HLM model that partitions variance in the outcome into variance between students within classrooms, variance between classrooms within schools, and variance across schools. The model was specified as follows:

$$Y_{ijk} = \gamma_{000} + \sum_{p=1}^p \beta_{pijk} X_{pijk} + u_{00k} + u_{0jk} + r_{ijk}$$

where  $Y_{ijk}$  is the assessment score of student  $i$  in classroom  $j$  in school  $k$ ,  $\gamma_{000}$  is a constant,  $\beta_{pijk}$  are the coefficients for each predictor  $p$ , and  $u_{00k}$ ,  $u_{0jk}$ , and  $r_{ijk}$  are the partitioned variance components.

We estimate this model separately for three groups of schools: (1) Schools that report using readiness assessments for classroom placements (2) schools that report administering readiness assessments but not for placements, and (3) all schools that do not report using readiness assessments for classroom placements (both those that administer readiness assessments and those that do not). The key parameter in these models is  $u_{0jk}$ , which captures the

extent to which students cluster in kindergarten classrooms based on achievement as measured in the fall of their kindergarten year. The sorting hypothesis holds that this parameter should be greater in schools that use readiness assessments for classroom placement than in schools that do not administer assessments or administer them but do not use them for classroom placement. The null hypothesis, by contrast, holds that this parameter should not vary significantly with readiness assessment use.

In our primary analysis, we estimated three different HLM models for each outcome variable (math scores and reading scores). The first was an unconditional model that included no predictors. The second added a series of control variables (described in the data section). The third model added school-wide variance in achievement as a control. These models were run separately for (1) schools that used assessments for classroom placements, (2) schools that gave readiness assessments but did not use the information for class placements, and (3) all schools that did not use assessments for classroom placements. To determine whether or not the variance estimates were statistically different across schools using assessment data for class placement and those not doing so, we compared variance estimates using Welch's t-tests. These models were estimated for standardized achievement scores in both math and reading in schools that had at least two classrooms with at least three sampled students per classroom.

### **Robustness and Sensitivity Analyses**

We conducted a series of subsequent analyses with different estimation approaches in order to ascertain the robustness of the findings from the HLM models. First, we reran our primary HLM variance components models on a restricted sample of schools. Each school in the restricted sample contained at least three classrooms with at least five sampled students in each classroom. Given the ECLS-K's sampling strategy, some classrooms contain very few students

which complicates estimating differences in average test scores across classrooms. These “large-class” models partially address this limitation of the data by restricting estimates to schools and classrooms with a larger number of sampled students.

Our next robustness check replaced the HLM models with a simpler Ordinary Least Squares (OLS) model. In this specification, the outcome variables were the calculated variance of the average classroom achievement scores within a school. In the fully specified versions of these models, we included the full set of covariates described in Appendix Table A1.

Our final approach focused on the range of average classroom achievement within schools. Modeled within an OLS framework, this approach used the difference between the average scores of the highest performing and lowest performing class within each school as the outcome of interest. In the fully specified versions of these models, we included the full set of covariates described in Appendix Table A1.

## **Results**

In this section, we present the results of our primary analytic approach as well as each of the robustness analyses. Our findings suggest that the use of kindergarten readiness assessments for the purpose of class placement decisions is predictive of slightly higher cross-class ability sorting by incoming reading achievement in kindergarten. The relationship with cross-class ability sorting in mathematics is less pronounced. This finding is generally robust to various sensitivity analyses, though we make particular note of data limitations that temper our results.

Descriptive statistics point to a possible relationship between the reported use of readiness assessments for classroom placement decisions and cross-class ability sorting. As shown in Table 1, the variance in the classroom average of reading and mathematics scores within a school is larger in schools that report using assessments for placement (column 2) than

in those that report using assessments but not for placement (column 3). This difference is statistically significant at the traditional ( $p < 0.05$ ) level for reading and marginally significant ( $p < 0.10$ ) for mathematics. In addition to significant differences in variance, schools differed significantly in the range of average classroom math and reading achievement scores. In particular, schools that reported using assessments for placement had, on average, about a one to one and a half point larger difference in the range between the maximum and minimum classroom averages of student achievement in both mathematics and reading as compared to schools that did not report using assessments for this purpose. This difference equates to an approximately 0.10 standard deviation larger range of test scores in schools that reported using assessments for placement than in those that did not.

Moving beyond the conditional mean comparisons, we turn to results of the HLM variance decomposition models. Table 2 shows the estimated variance components for each of the three groups of schools. In particular, we show results for schools that report using readiness assessments for class placement, those that do not but give assessments, and the full group of schools that do not report using readiness assessments for placement. We show variance components for the within classroom variance in achievement scores and the across classroom variance in achievement scores for both math and reading. Our interest is in comparing the variance components for the first group of schools (first column) to the other two groups of schools (second two columns). Greater across classroom variance in schools that use readiness assessment data for class placement would indicate greater cross-class ability sorting. Likewise, lower levels of within class variance (representing more homogenous classrooms) may also indicate greater cross-class ability sorting.

We compare the variance components across school groups using t-tests. The column labeled “a” indicates the statistical significance of a t-test of the variance component estimates between schools that report using assessments for placement and those that do not but do give assessments. The column labeled “b” indicates the statistical significance of a t-test of the variance component estimates between schools that report using assessments for placement and the full group of schools that do not (both those that report giving assessments but do not use them for placements and those that report not giving assessments at all). As shown, the cross-class variance component estimates in reading are consistently larger in schools that report using readiness assessments for placement decisions. This relationship is robust to the inclusion of the covariates and the school-wide variance in achievement scores. In mathematics, the differences are less pronounced with a few marginally significant differences emerging in models including controls. The within class variance estimates were statistically indistinguishable across school types.

Given the relatively small number of kindergarten classrooms and number of sampled students per classroom in some of the schools in the primary tables, we estimated similar equations with a restricted sample in which schools had at least three kindergarten classrooms with at least five sampled students per classroom. Table 3 shows results from the variance decomposition models with this sample. The organization of the table follows that described in Table 2. Notably, the results of this model displayed no statistically significant differences between the variance component estimates in schools that use readiness assessments for placement decisions and those that do not. Nevertheless, the point estimates on the variance components are very similar to those in the full sample and do indicate larger cross-class variance in schools that use assessments for placement decisions. For instance, the cross-class

variance component estimate in reading is nearly twice the size in schools that use readiness assessment for placement decisions than in those that do not. These models, while potentially reducing measurement error in the average classroom composition, result in reduced statistical power which may potentially contribute to the lack of statistical significance of the findings.

In addition to the variance decomposition models, we also modeled cross-class ability sorting through ordinary least squares regression models predicting the variance in average classroom achievement scores within a school as well as the range of average classroom achievement scores within a school (difference in maximum and minimum average class scores in a school). Table 4 shows estimated coefficients from models predicting both the variance and range of classroom average achievement scores within schools. Unlike the variance decomposition models, these estimation approaches used schools as the unit of observation given that the estimated outcomes are school level characteristics. As shown, estimates of the variance in classroom average of reading achievement and mathematics achievement (columns 1-4) are consistently positive in magnitude; however, the estimates are statistically indistinguishable from zero. The estimated coefficients on the range of classroom average scores also indicate greater cross-class ability sorting and, for three of the four estimates, reach at least levels of marginal statistical significance. For reading, in particular, the fully specified model including control variables (column 8) suggests that the reported use of readiness assessment data for class placement purposes is predictive of an approximately one point increase in the range of cross-class average classroom achievement scores in reading, a result that is statistically significant at the traditional level ( $p < 0.05$ ).

In summary, the directions of the various specifications consistently point towards increased cross-class ability sorting in schools that report using readiness assessments for class

placement decisions as compared to those that do not, though the results do not always reach levels of statistical significance. Of those that do, the relationship is particularly pronounced for cross-class ability sorting by reading achievement. In particular, results of the variance decomposition model for the full sample and results of the OLS models predicting the range of classroom average achievement demonstrated statistically significant relationships between the reported use of readiness assessments for class placement and cross-class ability sorting. These results are consistent with the conditional means shown in the descriptive statistics.

### **Discussion**

Prior research has demonstrated that early education and experiences in kindergarten, in particular, can have significant impacts on students' academic trajectories (Chetty et al., 2011; Claessens & Engel, 2013; Claessens et al., 2009). One important aspect of the kindergarten experience is the distribution of students across classes because that distribution can affect both the quality of peers and the quality of resources (including teachers) allocated to classrooms (Kalogrides & Loeb, 2013).

Our findings suggest that the distribution of students in kindergarten may be predicted by the availability and use of readiness assessment data. Specifically, the results of our study provide some support for our *sorting hypothesis*: that the use of kindergarten readiness assessments for classroom placement may be related to somewhat higher levels of cross-class ability sorting, particularly by reading achievement, in kindergarten. We cannot claim a causal connection. Nonetheless, our findings are consistent with the hypothesis that by providing a source of information on incoming student ability kindergarten readiness assessments may allow school administrators to engage in some level of cross-class ability sorting that would be difficult in the absence of information on students' prior ability. In the remainder of this section, we

consider the practical and policy implications of our findings, discuss the study's limitations, and point to avenues for future research.

In terms of the implications of our findings, we interpret the relationship between the use of readiness assessments and cross-class ability sorting as neither inherently positive nor negative for students. As noted in our literature review, there are, to our knowledge, no studies that explore the effects of cross-class ability sorting for kindergartners. If theoretical arguments for the benefits of cross-class ability sorting through curricular differentiation hold then our findings suggest that policies and practices that encourage the assessment of kindergartners at school entry may lead to the construction of classrooms that better meet the academic needs of students. If, however, we consider the results of research on cross-class sorting in the upper grades—results showing that lower achieving students are often sorted into classrooms that have lower quality teachers, lower achieving peers, and less rigorous content (Gamoran & Mare, 1989; Kalogrides & Loeb, 2013; Oakes, 2005)—then these same policies and practices may instead lead to classroom assignment decisions that serve to exacerbate existing gaps in achievement.

In schools where readiness assessments are used to sort students by ability across classrooms in the kindergarten year, it is important that educators do so in a way that does not disadvantage lower achieving students and that does not permanently place students on a track of lower expectations. For example, school leaders might work to ensure that classes with lower achieving students are coupled with the strongest kindergarten teacher or with additional resources to support student learning. Similarly, school personnel in schools that sort students across classrooms in the kindergarten year should be sure to allow opportunities for reassessment as students progress to first grade, thereby recognizing the potential for lower achieving students

to make significant developmental gains in the kindergarten year and preventing a tracking structure that might disadvantage lower achieving students.

In schools where readiness assessments are not being used for cross-class ability sorting, school personnel should be mindful of the potential for more variation in achievement within classrooms. The use of within-class ability sorting, using groups that are flexible and that adjust to student growth, may allow these schools to achieve optimally differentiated instruction within a single classroom. To do so, however, school leaders may need to allocate additional resources to support teachers serving a more heterogeneous group of learners.

It is important to keep in mind that the observed relationship between the use of readiness assessments for classroom placement and cross-class ability sorting is relatively modest and largely confined to reading achievement. School administrators may balance their use of assessment information to cluster students by ability with institutional norms that value fairness and, by consequence, the distribution of student ability evenly across classrooms, a balance that may suppress the size of our coefficients (Osborne-Lampkin & Cohen-Vogel, 2014). Another explanation for the modest relationship relates to the timing of readiness assessment administration. The timing in the administration of readiness assessments varies across schools, with some administered prior to the start of the school year and others after the school year begins (Conn-Powers & Peters, 2007; Education Commission of the States, 2014; Msyzak & Conn-Powers, 2008). While readiness assessments administered prior to the start of the year could easily be used to inform classroom placement decisions, those administered after the start of school are much less likely to affect classroom placements for the kindergarten year insofar as students are already in classrooms. It is possible then that when principals report using readiness assessments for classroom placement on the ECLS-K survey, they are actually indicating the use

of kindergarten readiness assessments for making placements for the following school year. Given that many states do not mandate other formal assessments during the kindergarten year, school personnel may see scores from the readiness assessment as the best available standardized metric for consideration in later grade placements. Unfortunately, the ECLS-K did not include information about the timing of readiness assessments, but future work might consider the relationship between reported use of data for classroom placement and cross-class sorting in 1<sup>st</sup> grade.

### **Limitations**

While addressing an important and relatively unexplored area, our study is limited in several ways. The primary limitation relates to the sampling frame. While the ECLS-K:2011 has the advantage of offering nationally representative estimates, it samples only a small number of students within any given classroom and a subset of teachers within a school. As a result, there may be considerable noise in estimates of variation in achievement scores across classrooms. While we attempted to address this issue by presenting models restricted to schools and classrooms with larger numbers of sampled students in each classroom and a minimum number of classrooms within the school, we sacrificed statistical power and generalizability by doing so. While these efforts partially mitigate this limitation of the data, we recognize that estimates in this study are not as precise as would be expected if the analyses contained achievement scores for all students. Replication of this study with administrative data, such as that collected by many state data systems, would be a useful next step for research in this domain.

A second limitation of this study concerns the lack of detail available on the readiness assessments administered. We are limited to only a binary indicator of whether readiness assessments are used for classroom placement. As such, we lack information on the types of

readiness assessments given, even as we know from the literature that some assessments emphasize academic skills while others place a greater emphasis on socio-emotional skills (Conn-Powers & Peters, 2007; Education Commission of the States, 2014; Gewertz, 2014 October 7; Public Citizens for Children and Youth, 2015). Also missing is information that can help build a more nuanced understanding of the ways schools use assessment results for placement decisions. Future mixed methods work might involve the collection and coding of readiness assessments used in a smaller sample of schools and districts combined with administrative data on classroom composition.

Finally, though we find evidence of a modest relationship between the reported use of readiness assessments for classroom placement and cross-class ability sorting, particularly in reading, we are unable to rule out the possibility that this relationship is the result of other unobserved factors. Though our models do control for a number of observable characteristics of schools, the results should be interpreted only as adjusted correlations rather than causal estimates. In particular, we are unable to separate the extent to which the use of assessments increases the sorting across kindergarten classrooms from the possibility that a school-level propensity to sort students drives the adoption of student assessments. Consequently, future research may seek sources of more plausibly exogenous variation that could push closer to causal estimates of the impact of readiness assessment use for placement on student sorting.

### **Conclusion**

School readiness assessments are now the reality for the majority of our nation's youngest students. Coupled with the fact that principals' reported use of these assessments for classroom placement decisions has increased significantly over the last decade, understanding their implications for the distribution of students is important. To our knowledge, this study is

the first to systematically examine the relationship between the use of readiness assessment data for classroom placement and kindergarten sorting. Our findings suggest that data use from readiness assessments may be predictive of somewhat greater cross-class ability sorting, in reading, most notably. While our study is unable to make a causal connection, the findings nevertheless suggest the possibility that the use of readiness assessments for classroom placement decisions contributes to sorting. As states continue to mandate the administration of school readiness assessments, further examination is needed both of the relationship between readiness assessments and sorting in kindergarten as well as the effect of kindergarten sorting on various student outcomes. Until such a time, individual schools should be mindful of the ways they use readiness assessment data, particularly for classroom placement decisions.

## References

- Allan, S. D. (1991). Ability grouping research reviews: What do they say about grouping and the gifted. *Educational Leadership*, 48(6), 60-65.
- Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata* (Vol. 2). College Station, TX: Stata Press.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2010). How does your kindergarten classroom affect your earnings? Evidence from Project STAR (No. w16381). National Bureau of Economic Research.
- Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type an international comparison of students' mathematics self-concept. *American Educational Research Journal*, 50(5), 925-957.
- Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review*, 28(4), 415-427.
- Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, 115(6), 1-29.
- Claessens, A., Engel, M., & Curran, F. C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, 51(2), 403-434.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778-820.
- Cohen-Vogel, L., & Harrison, C. (2013). Leading with data: Evidence from the National Center on Scaling up Effective Schools. *Leadership and Policy in Schools*, 12(2), 122-145.

- Cohen-Vogel, L., & Rutledge, S. (2009). The pushes and pulls of new localism: School-level instructional arrangements, instructional resources, and family-community partnerships. *Yearbook of the National Society for the Study of Education, 108*(1), 70-103.
- Collins, C. A., & Gan, L. (2013). *Does sorting students improve scores? An analysis of class composition* (No. w18848). National Bureau of Economic Research.
- Condran, D.J. (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems, 54* (1).
- Conn-Powers, M. & Peters, J. (2007). Kindergarten readiness assessment practices in Indiana. Early Childhood Center. Indiana University. Retrieved from:  
<https://www.iidc.indiana.edu/styles/iidc/defiles/ECC/SRSA-KRAssmtResults.pdf>
- Council of Chief State School Officers. (2010). State early childhood standards and assessments: Five years of development. Report of 2008 Survey by the Early Childhood State Collaborative on Assessment and Student Standards. Washington, DC: Author.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2010). How does your kindergarten classroom affect your earnings? Evidence from Project STAR (No. w16381). National Bureau of Economic Research.
- Dieterle, S., Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management, 34*(1), 32-58.
- Education Commission of the States. (2014). Kindergarten Entrance Assessments. Retrieved from: <http://ecs.force.com/mbdata/mbquestRT?rep=Kq1407>

- Epstein, A. S., Schweinhart, L. J., DeBruin-Parecki, A., & Robin, K. B. (2004). Preschool assessment: A guide to developing a balanced approach. *Preschool Policy Matters*, 7, 1-2.
- Gallagher, L., Means, B., & Padilla, C. (2008). *Teachers' Use of Student Data Systems to Improve Instruction: 2005 to 2007*. Washington, DC: US Department of Education.
- Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality?. *American Journal of Sociology*, 1146-1183.
- Garrett, R., & Hong, G. (2015). Impacts of grouping and time on the math learning of language minority kindergartners. *Educational Evaluation and Policy Analysis*, 38(2), 222-244.
- Gewertz, C. (2014, October 7). Kindergarten-readiness tests gain ground: Aim is to support instruction, but concerns persist. *Education Week*.
- Gottfried, M. A. (2014). Peer effects in urban schools: Assessing the impact of classroom composition on student achievement. *Educational Policy*, 28(5), 607-647.
- Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement?. *Journal of Applied Econometrics*, 18(5), 527-544.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7), 798-812.
- Hong, G., Corter, C., Hong, Y., & Pelletier, J. (2012). Differential effects of literacy instruction time and homogeneous ability grouping in kindergarten classrooms who will benefit? Who will suffer?. *Educational Evaluation and Policy Analysis*, 34(1), 69-88.
- Horvath, H. (2015). Classroom assignment policies and implications for teacher value-added estimation. Working Paper. Retrieved from:  
<https://sites.google.com/site/hedvigh/research>

- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers the magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304-316.
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6), 615-631.
- Kulik, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives*. National Research Center on the Gifted and Talented, University of Connecticut.
- Ladd, H. F., & Sorensen, L. C. (2014). Returns to teacher experience: Student achievement and motivation in middle school. *Center for Analysis of Longitudinal Data in Education Research Working Paper*, 112.
- Little, M. H. & Cohen-Vogel, L. (2017). Ready for school? Assessing America's kindergarteners. *Teachers College Record*. Retrieved from:  
<http://www.tcrecord.org/Content.asp?ContentID=21959>
- Little, M. H., Cohen-Vogel, L., & Curran, F. C. (2016). Facilitating the transition to kindergarten: What ECLS-K data tell us about school practices then and now. *AERA Open*, 2(3), doi: 2332858416655766.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of educational research*, 66(4), 423-458.
- Loveless, T. (1998). The tracking and ability grouping debate. The Fordham Report. Washington, DC: The Fordham Foundation.

- Loveless, T. (2013). The Resurgence of Ability Grouping and Persistence of Tracking: Part II of the 2013 Brown Center Report on American Education. *Brookings Institution Report, Brown Center on Education Policy*.
- Lucas, S. R. (1999). *Tracking Inequality: Stratification and Mobility in American High Schools. Sociology of Education Series*. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027.
- March, J.G. & Olsen, J.P. (1984). The new institutionalism: Organizational factors in political life. *The American Political Science Review*, 78(3), 734-749.
- Matthews, M. S., Ritchotte, J. A., & McBee, M. T. (2013). Effects of schoolwide cluster grouping and within-class ability grouping on elementary school students' academic achievement growth. *High Ability Studies*, 24(2), 81-97.
- McCoach, D.B, O'Connell, A.A. & Levitt, H. (2006) Ability grouping across Kindergarten using an early childhood longitudinal study. *The Journal of Educational Research*, 99(6), 339-346,
- McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, 22(2), 131-141.
- Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing Data-Informed Decision Making in Schools--Teacher Access, Supports and Use*. Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Prepared by SRI International, Menlo Park, CA
- Merrill, B., Little, M.H., Sadler, J., Cohen-Vogel, L., Lee, K. A portrait of pre-k quality in the United States. Working Paper. University of North Carolina.

- Mickelson, R., & Everett, B. (2008). Neotracking in North Carolina: How high school courses of study reproduce race and class-based stratification. *The Teachers College Record*, 110(3), 535-570.
- Msyzak, J.P. & Conn-Powers, M. (2008). Kindergarten readiness assessment: An analysis of existing and recommended practices. Early Childhood Center. Indiana University. Retrieved from: <https://www.iidc.indiana.edu/styles/iidc/defiles/ecc/srsa-krassmntanalysis.pdf>
- Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using student-assessment results to improve instruction: lessons from a workshop. *Journal of Education for Students Placed at Risk*, 10(3), 269–280.
- Niemeyer, J., & Scott-Little, C. (2001). Assessing Kindergarten Children: A Compendium of Assessment Instruments. Office of Educational Research and Improvement (ED), Washington, DC.
- Nomi, T., & Allensworth, E. (2014). *Skill-Based Sorting in the Era of College Prep for All: Costs and Benefits. Research Brief*. University of Chicago Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.
- Oakes, J. (2005) Keeping track: How schools structure inequality. New Haven, CT: Yale University Press.
- Oakes, J. & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal*, 32(1), 3-33.
- Osborne-Lampkin, L. T., & Cohen-Vogel, L. (2014). “Spreading the wealth”: How principals use performance data to populate classrooms. *Leadership and Policy in Schools*, 13(2), 188-208.

- Park, V., & Datnow, A. (2017). Ability grouping and differentiated instruction in an era of data-driven decision making. *American Journal of Education, 123*(2).
- Passow, A. H. (1962). The maze of research on ability grouping. *Educational Forum, 26*, 281-288.
- Powell, W. & DiMaggio, P. (1991). *The new institutionalism in organizational analysis*. University of Chicago Press.
- Public Citizens for Children and Youth (2015). Assessing kindergarten readiness in Pennsylvania: How can the kindergarten entry inventory boost skills of young students?
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247-252.
- Royston, P., & White, I. R. (2011). Multiple imputation by chained equations (MICE): implementation in Stata. *Journal of Statistical Software, 45*(4), 1-20.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association, 91*(434), 473-489.
- Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Research progress report: University of Tennessee Value-Added Research and Assessment Center.
- Sharkey, N., & Murnane, R. (2003, April). Helping K-12 educations learn from assessment data. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*, 57, 293–336.
- Smylie, M. A. (1995). New perspectives on teacher leadership. *The Elementary School Journal*, 3-7.
- Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. National Academies Press.
- Stedron, J. M., & Berger, A. (2010). NCSL technical report: State approaches to school readiness assessment. In *National Conference of State Legislatures*. <http://www.ncsl.org/documents/Educ/KindergartenAssessment.pdf>.
- Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on k–12 students' academic achievement findings of two second-order meta-analyses. *Review of Educational Research*, 86(4), 849–899.
- Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research*, 35(4), 1048-1079.
- Tomlinson, C. (1995). *How to differentiate in mixed ability classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tomlinson, C. (1999). *The differentiated classroom. Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., Daly, P., & Najarian, M. (2015). Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011). User's Manual for the ECLS-K: 2011 Kindergarten Data File and Electronic Codebook, Public Version. NCES 2015-074. *National Center for Education Statistics*.

Wachen, J., C. Harrison & Cohen-Vogel, L. (2015). Data use and classroom instruction: Have we hit a wall? Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL.

Table 1. Means and standard errors of measures of key dependent variables and covariates for the full sample and by use of kindergarten readiness assessments for classroom placement decisions

	Full Sample	Assessments used for Placement	Assessments Not Used for Placement but Given	Assessments Not Used for Placement	Significance of t-test of columns	
					2v3	2v4
	(1)	(2)	(3)	(4)	(5)	(6)
Readiness assessment used for placement (0/1)	0.357 (0.028)	1.000 .	0.000 .	0.000 .		
Readiness assessment given (0/1)	0.755 (0.025)	1.000 .	1.000 .	0.619 (0.030)		
Math achievement score	31.850 (0.293)	32.026 (0.441)	31.741 (0.357)	31.753 (0.331)		
Reading achievement score	47.177 (0.262)	47.343 (0.410)	46.967 (0.366)	47.085 (0.329)		
Math achievement score (std.)	-0.010 (0.026)	0.005 (0.039)	-0.020 (0.031)	-0.019 (0.029)		
Reading achievement score (std.)	0.000 (0.023)	0.014 (0.036)	-0.019 (0.032)	-0.008 (0.029)		
Variance in classroom average of math scores in school	24.341 (1.112)	27.384 (2.876)	21.684 (1.278)	22.654 (1.019)	†	
Variance in classroom average of reading scores in school	25.828 (1.310)	28.037 (2.444)	22.039 (1.832)	24.602 (1.495)	*	
Max average classroom math score in school	36.183 (0.340)	36.741 (0.483)	35.767 (0.440)	35.873 (0.395)		
Min average classroom math score in school	27.844 (0.308)	27.713 (0.499)	27.958 (0.354)	27.916 (0.341)		
Max average classroom reading score in school	51.515 (0.327)	51.972 (0.497)	50.740 (0.482)	51.262 (0.416)	†	
Min average classroom reading score in school	43.127 (0.263)	43.006 (0.410)	43.263 (0.330)	43.194 (0.311)		
Diff in max math and min math scores in school	8.339 (0.210)	9.028 (0.418)	7.809 (0.305)	7.957 (0.230)	*	*
Diff in max reading and min reading scores in school	8.389 (0.258)	8.966 (0.435)	7.478 (0.355)	8.068 (0.293)	*	†
<i>n</i>	10440	3470	4320	6970		

Note. Standard errors in parentheses. All models are weighted using school level weights and accounting for the complex survey design of the ECLS-K. Unit of observations are students. \*\* p<0.01, \* p<0.05, + p<0.1

Table 2. Variance component estimates and standard errors from HLM variance decomposition models

	Variance Level		Assessment data is used for class placement		Assessment data is not used for classroom placement (but assessment is used)		Assessment data is not used (full sample denominator)		a	b
			Variance	SE	Variance	SE	Variance	SE		
Null Model	Within Classroom Variance	Math Assessment	0.655	0.027	0.617	0.017	0.626	0.015		
		Reading Assessment	0.644	0.037	0.637	0.030	0.668	0.023		
	Across Classroom Variance	Math Assessment	0.175	0.018	0.141	0.012	0.145	0.010		
		Reading Assessment	0.190	0.019	0.129	0.014	0.142	0.011	*	*
School Controls	Within Classroom Variance	Math Assessment	0.654	0.027	0.616	0.017	0.626	0.015		
		Reading Assessment	0.643	0.037	0.636	0.030	0.668	0.023		
	Across Classroom Variance	Math Assessment	0.176	0.017	0.135	0.012	0.141	0.010	*	†
		Reading Assessment	0.191	0.022	0.125	0.014	0.139	0.012	*	*
School Controls + School Variance	Within Classroom Variance	Math Assessment	0.654	0.027	0.616	0.017	0.626	0.015		
		Reading Assessment	0.643	0.037	0.636	0.030	0.668	0.023		
	Across Classroom Variance	Math Assessment	0.176	0.018	0.135	0.012	0.140	0.010	†	†
		Reading Assessment	0.193	0.020	0.124	0.013	0.138	0.011	*	*
n			3470		4320		6970			

Note. a column represents statistical significance of a t-test comparing the variance estimate for the assessment data is used for class placement group to the assessment data is not used for classroom placement but readiness assessments are given group.

b column represents statistical significance of a t-test comparing the variance estimate for the assessment data is used for class placement group to the assessment data is not used full sample (those not giving readiness assessments and those giving assessments but not using them for placement).

\*\* p<0.01, \* p<0.05, † p<0.1

Table 3. Variance component estimates and standard errors from HLM variance decomposition models with the large class sample

Readiness Assessment Condition	Variance Level		Assessment data is						a	b
			used for class placement		not used for classroom placement (but assessment is used)		not used (full sample denominator)			
			Variance	SE	Variance	SE	Variance	SE		
Null Model	Within Classroom Variance	Math Assessment	0.640	0.068	0.665	0.045	0.642	0.036		
		Reading Assessment	0.584	0.060	0.701	0.077	0.694	0.057		
	Across Classroom Variance	Math Assessment	0.141	0.027	0.137	0.021	0.155	0.025		
		Reading Assessment	0.200	0.051	0.110	0.024	0.137	0.023		
School Controls	Within Classroom Variance	Math Assessment	0.635	0.067	0.663	0.045	0.641	0.036		
		Reading Assessment	0.576	0.060	0.698	0.077	0.692	0.057		
	Across Classroom Variance	Math Assessment	0.134	0.026	0.138	0.022	0.156	0.025		
		Reading Assessment	0.196	0.052	0.113	0.026	0.138	0.023		
School Controls + School Variance	Within Classroom Variance	Math Assessment	0.635	0.067	0.663	0.045	0.641	0.036		
		Reading Assessment	0.576	0.060	0.698	0.077	0.691	0.057		
	Across Classroom Variance	Math Assessment	0.135	0.027	0.136	0.021	0.156	0.025		
		Reading Assessment	0.197	0.052	0.112	0.025	0.138	0.023		
n			820		950		1590			

Note. a column represents statistical significance of a t-test comparing the variance estimate for the assessment data is used for class placement group to the assessment data is not used for classroom placement but readiness assessments are given group.

b column represents statistical significance of a t-test comparing the variance estimate for the assessment data is used for class placement group to the assessment data is not used full sample (those not giving readiness assessments and those giving assessments but not using them for placement).

\*\* p<0.01, \* p<0.05, † p<0.1

Table 4. Coefficients and standard errors from regression models predicting variance in average classroom math and reading scores as well as difference in the max and min average math and reading scores in a school

	Variance in classroom average of math scores in school		Variance in classroom average of reading scores in school		Diff in max math and min math scores in school		Diff in max math and min reading scores in school	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Assessment used for placement class placement	2.951 (3.033)	3.297 (2.841)	2.765 (2.705)	3.921 (2.764)	0.959+ (0.513)	0.687 (0.470)	0.956+ (0.541)	1.056* (0.505)
Constant	20.94** (1.326)	-22.49** (7.783)	22.28** (1.449)	-17.95* (7.203)	7.011** (0.251)	-2.943* (1.164)	7.087** (0.263)	-2.292+ (1.351)
Controls		Yes		Yes		Yes		Yes
n	620	620	620	620	620	620	620	620

Note. Standard errors in parentheses. All models are weighted using school level weights and accounting for the complex survey design of the ECLS-K. Controls include those listed in Table 1. Unit of observations are schools.

\*\* p<0.01, \* p<0.05, † p<0.1

Appendix Table A1. Means and standard errors of measures of covariates for the full sample and by use of kindergarten readiness assessments for classroom placement decisions

	Full Sample	Assessments used for Placement	Assessments Not Used for Placement		Significance of t-test of columns	
			Not Used for Placement but Given	Assessments Not Used for Placement	2v3	2v4
	(1)	(2)	(3)	(4)	(5)	(6)
Variance in math achievement scores in school	101.496 (2.147)	105.337 (3.993)	98.764 (2.793)	99.366 (2.365)		
Variance in reading achievement scores in school	108.465 (3.101)	108.520 (5.674)	103.165 (4.829)	108.434 (3.873)		
District poverty composite (2-54)	18.614 (0.915)	17.836 (1.117)	19.243 (0.994)	19.046 (1.020)		
Free lunch eligible students in school (%)	43.443 (1.860)	42.970 (2.514)	43.856 (2.479)	43.705 (2.275)		
Non-white students in school (%)	44.037 (2.464)	40.216 (2.627)	46.497 (3.202)	46.157 (2.890)		
English language learners in school (%)	14.618 (1.254)	15.031 (1.736)	15.111 (1.680)	14.389 (1.411)		
Full day kindergarten (0/1)	0.810 (0.028)	0.795 (0.035)	0.787 (0.038)	0.818 (0.032)		
Pre-k in school (0/1)	0.473 (0.029)	0.477 (0.045)	0.451 (0.038)	0.471 (0.032)		
Kindergarten enrollment (# students)	91.405 (2.797)	98.817 (3.851)	91.667 (4.618)	87.294 (3.213)		*
Met AYP (0/1)	0.707 (0.046)	0.675 (0.057)	0.710 (0.046)	0.725 (0.049)		
Public school (0/1)	0.921 (0.010)	0.931 (0.015)	0.935 (0.017)	0.916 (0.015)		
School received Title 1 funds (0/1)	0.682 (0.025)	0.734 (0.035)	0.663 (0.036)	0.652 (0.032)		†
Urban location (0/1)	0.335 (0.034)	0.305 (0.041)	0.357 (0.047)	0.352 (0.042)		
<i>n</i>	10440	3470	4320	6970		

Note. Standard errors in parentheses. All models are weighted using school level weights and accounting for the complex survey design of the ECLS-K. Unit of observations are students. \*\* p<0.01, \* p<0.05, + p<0.1