HOMOGENEOUS GROUPING IN EARLY ELEMENTARY READING INSTRUCTION

The Challenge of Identifying Appropriate Comparisons and Examining Differential Associations between Grouping and Reading Growth

ABSTRACT

Separating students into homogeneous groups is a common instructional practice used by elementary teachers during reading instruction. Although researchers have been studying the effects of homogeneous grouping for many decades, there is little consensus on whether grouping is an effective or equitable instructional practice. The central challenge in estimating the effects of grouping is determining an appropriate comparison case for homogeneously grouped students. Using data from the 2010 cohort of the Early Childhood Longitudinal Study, this study uses student fixed effects to compare the reading growth of students who have different grouping experiences during reading instruction over their first 3 years of schooling. This study finds that, on average, homogeneously grouped students have slightly higher reading growth than students who are not grouped but also finds that this association between homogeneous grouping and reading growth depends on students' reading group placement and initial reading skills.

Susan Kemper Patrick
VANDERBILT UNIVERSITY

THE ELEMENTARY SCHOOL JOURNAL
Volume 120, Number 4. Published online May 28, 2020
 0013-5984/2020/12004-0003$10.00

SEPARATING students into instructional groups within the same classroom based on past performance or perceived ability is a common educational practice used by elementary teachers during reading instruction. Proponents of homogeneous grouping, also called within-class ability grouping,[1] argue that it is an effective strategy that allows teachers to better tailor their instruction and materials to the needs of their students (Kulik & Kulik, 1992; Lou et al., 1996; Slavin, 1987; Taylor et al., 2000). Others, however, have argued that grouping practices exacerbate achievement disparities if children in certain groups benefit more from these grouping arrangements (Condron, 2008; Tach & Farkas, 2006). Research on grouping suggests that students placed in the higher achievement groups receive more attention from the teacher and are exposed to more rigorous instruction than their peers in lower achievement groups (Chorzempa & Graham, 2006; Oakes, 2005). Thus, homogeneous grouping may be a school-based mechanism that contributes to educational inequality (Entwisle & Alexander, 1993; Gamoran, 1992, 2011; Pallas et al., 1994).

Homogeneous grouping appears to be an increasingly common practice among elementary teachers. One recent analysis using teacher survey data from the National Assessment of Educational Progress found that the percentage of fourth-grade teachers who report creating reading instructional groups based on ability has steadily increased between 1998 and 2009, from 28% to 71% (Loveless, 2013). My own calculations using the two waves of the Early Childhood Longitudinal Study (ECLS-K), which follow two nationally representative cohorts of incoming kindergarten students throughout elementary school, found that the proportion of kindergarten students whose teachers report using "achievement groups" in reading increased from 41% in 1998 to 79% in 2010. This increase parallels a growing emphasis on differentiated instruction, both as an instructional best practice (Tomlinson, 2001) and a policy mandate associated with high-stakes accountability and use of student-level data (Gamoran, 2011; Park & Datnow, 2016; Valli & Buese, 2007).

Although researchers have been studying the effects of homogeneous grouping for over a century (Steenbergen-Hu et al., 2016), there is little consensus on whether it is an effective or equitable instructional practice. The central challenge of this research is determining an appropriate comparison case for homogeneously grouped students. Much of the earliest work on grouping, which compared the outcomes for students based on a single year of data, has been criticized for not appropriately controlling for differences between grouped and ungrouped classrooms or between students placed in higher achievement groups and those in lower achievement groups. More recent empirical work on grouping in early elementary reading instruction, much of which uses data from the 1998 kindergarten cohort of the ECLS-K, generally reports that achievement grouping is positively associated with students' reading growth (Adelson & Carpenter, 2011; Hong & Hong, 2009; McCoach et al., 2006) but that these associations vary based on initial reading skills or group placement (Condron, 2008; Hong et al., 2012; Lleras & Rangel, 2009; Nomi, 2009; Tach & Farkas, 2006). These studies—which use either covariate-adjusted regression or propensity score matching—rely on the assumption that researchers can account for student, teacher, and/or school characteristics to create appropriate comparison groups.

In this analysis, I both update these prior findings with newer data and introduce a different strategy to estimate the relationship between homogeneous grouping and reading growth. This study updates prior findings by using data from a more recent wave of the ECLS-K, which follows students starting kindergarten in 2010.

Updating these findings is particularly important given that comparisons between the 1998 and 2010 kindergarten cohorts of the ECLS-K have reported an increased focus on academic skills and use of homogeneous grouping in the early grades (Bassok et al., 2016; Wrabel et al., 2015). Furthermore, by leveraging the longitudinal nature of the data, I use a different analytic approach than prior work. This approach uses student fixed effects to compare the outcomes of students who have different grouping experiences during reading instruction over their first 3 years of schooling. In models estimating the overall effect of grouping, a student's ungrouped year becomes the comparison case for their grouped year. I argue that this approach offers a more appropriate comparison group to assess grouping than has been used in earlier studies.

To better situate this analysis, I begin by exploring different conceptualizations of homogeneous grouping as either a mechanism for instructional differentiation or within-class stratification. I then briefly review the methodological approaches and results of prior research that examines the relationship between grouping and student reading growth. Next, I explain the data and methods used in this analysis, and then describe the findings. I conclude with a discussion of this study's limitations, implications for practice, and directions for future research.

Conceptualizing Grouping: Differentiation or Stratification?

Schools are social organizations that sort students in many ways as they progress through their educational experiences. Within-class grouping for reading instruction is just one way children are grouped at school (Entwisle & Alexander, 1993; Oakes, 2005). Teachers use many different forms of grouping for reading instruction, from temporarily grouping students for literacy stations to placing students into homogeneous groups for reading instruction that last most of the year. In this analysis, homogeneous or achievement grouping is defined as any form of grouping that occurs within the classroom that the teacher uses to create more homogeneous learning environments for students (Steenbergen-Hu et al., 2016). By sorting students into groups based on measures of student performance in reading (through the use of formal diagnostics, benchmark testing, and informal reading inventories) or their perceptions of students' literacy skills, teachers are then able to target the instruction and materials to each group's specific needs. Although such instructional differentiation is often considered best practice, many question whether it perpetuates inequality through an unequal distribution of instructional resources within classrooms (Oakes, 2005; Pallas et al., 1994). This enduring debate is reflected in how generations of educational researchers have conceptualized homogeneous grouping (Gamoran, 2011). In the following sections, I briefly review how homogeneous grouping has been conceptualized as both an organizational system supporting instructional differentiation and a within-school mechanism of educational stratification.

Grouping as Differentiation

Many conceptualize grouping as an instructional practice that allows teachers to better target instruction based on students' needs and skills. Many theories of learning hypothesize that students with varying skills respond differently to the same

instruction and that instruction is most effective when it builds on students' existing skills (Cronbach & Snow, 1977; Palinscar & Brown, 1984; Vygotsky, 1980). Within-class grouping allows teachers to target their instructional techniques and to use different materials for smaller groups of students. In contrast to whole-class instruction, instruction in small groups facilitates more remediation or enrichment, allows for more flexibility to adjust the pace of instruction, and capitalizes on the many benefits of collaborative learning among peers (Lou et al., 1996). When using homogeneous grouping as their grouping strategy, teachers can specifically plan instruction and small-group learning experiences for students based on their reading skill level.

When considering how homogeneous grouping can best support student learning, proponents have long argued that homogeneous groups should be small and that group placement should be flexible and regularly reevaluated by teachers based on students' current needs (Slavin, 1987; Taylor et al., 2000). Sociological research on grouping practices conducted in the 1980s found that organizational constraints—such as a necessity for equally sized groups or classroom configuration—result in a set number of groups and placements that are rarely changed throughout the school year (Eder, 1983; Hallinan & Sorensen, 1983). More recent analyses of grouping practices, in which teachers report having access to numerous data sources on students' skills and specific needs, suggest that teachers are now more likely to reassess group placement throughout the year (Hoover & Abrams, 2013; Park & Datnow, 2016). Thus, greater access to formative data may allow teachers to better follow best practices around homogeneous grouping (Missett et al., 2014).

Grouping as Stratification

In contrast, these sorting procedures have also been conceptualized as a within-school mechanism that creates unequal access to educational opportunities (Pallas et al., 1994). Homogeneous grouping organizes students into groups that receive differentiated instruction and varying level of status within a classroom (Gamoran, 1986). Sociological research suggests that schools privilege certain cultural, social, and linguistic habits and conflate the cultural capital exhibited by dominant and powerful groups with academic ability (Bourdieu, 1977). Teachers, then, may consciously or unconsciously make grouping decisions based on cultural indicators, such as certain manners of speaking or student habits, which can act as proxies for ability. Such sorting may appear meritocratic but actually reproduces the differing cultural capital that students bring to school (Condron, 2007; Oakes, 2005). A child's group placement (e.g., whether they are placed in lower or higher achievement groups) within a classroom affects the "quantity, quality, and pace of instruction" and acts as a symbolic marker that can influence the perceptions and expectations of teachers, parents, and children themselves (Pallas et al., 1994, p. 27).

An enduring finding of research on homogeneous reading grouping is that certain students—particularly boys, low-income students, and Black children—are more likely to be placed in the lower achievement groups (Condron, 2007; Dreeben & Gamoran, 1986; Haller, 1985; Oakes, 2005; Pallas et al., 1994; Tach & Farkas, 2006). However, it is unclear whether this pattern simply mirrors the well-documented racial and socioeconomic disparities in the reading skills and academic preparation exhibited by incoming elementary students (Duncan & Murnane, 2014; Quinn, 2015).

Although some studies report that differential placement rates can be explained by differences in academic performance (Dreeben & Gamoran, 1986; Haller, 1985; Pallas et al., 1994), others found that boys and lower-income students are more likely to be placed in lower achievement groups even when controlling for a reading achievement test score (Condron, 2007; Tach & Farkas, 2006). Regardless of whether these differential placement rates can be explained by academic disparities, grouping could still contribute to inequality if students in higher achievement groups are more likely to academically benefit from homogeneous grouping. I explore the issue of differential effects in the following section.

Prior Research on Homogeneous Grouping

If and how students are grouped during instruction could potentially have a large influence on their instructional experience. Grouping influences how students interact with their teacher and their peers as well as the amount, type, and rigor of instruction that they experience. Researchers have been trying to better understand the effects of homogeneous grouping for over a century (Steenbergen-Hu et al., 2016). In this section, I highlight different methodological approaches used to study the relationship between within-class grouping during reading instruction and findings from relevant research using the ECLS-K.

Methodological Approaches in Grouping Research

Much of the research on grouping occurred in the 1970s, 1980s, and 1990s (see Gamoran [1992], Kulik & Kulik [1992], Lou et al. [1996], and Slavin [1987] for a thorough review of earlier work). In most of these studies, the reading scores of students in homogeneously grouped classrooms were compared with students in ungrouped or heterogeneously grouped classes, and researchers typically presented information about students, teachers, or classrooms to assert equivalence between these groups. The methodology of earlier work has been called into question for not appropriately isolating the effect of grouping. Citing numerous studies on grouping, Ferguson (1998, p. 366) concluded that there was "no methodologically sound evidence of the effects of within-class ability grouping for reading."

In the past 2 decades, most research on homogeneous grouping has used covariate-adjusted regression analysis or propensity score matching to estimate the relationship between grouping and students' reading outcomes (Adelson & Carpenter, 2011; Condron, 2007, 2008; Hong et al., 2012; Hong & Hong, 2009; Lleras & Rangel, 2009; McCoach et al., 2006; Nomi, 2009; Tach & Farkas, 2006). These analyses improve on prior work by estimating the differences in reading growth between grouped and ungrouped students while controlling for a larger set of student, teacher, classroom, or school characteristics. Used in many of these analyses, propensity score matching is a statistical matching technique that first determines an observation's probability of treatment given a set of observed covariates and then estimates the difference in outcomes between treated and untreated observations that are matched based on that probability (Rosenbaum & Rubin, 1983). Although proponents have argued that a propensity score matched sample can mimic a randomized controlled trial if all variables

that affect treatment have been measured (Austin, 2011; Rubin, 2006), the effectiveness of this approach is heavily dependent on the availability and choice of covariates (Cook et al., 2009; Thoemmes & Kim, 2011). In the specific case of homogeneous grouping, it is unlikely that researchers could observe and measure all of the factors about teachers and their students that may influence both teachers' decisions about if and how to use grouping and the reading performance outcomes of the students in their class. For example, teachers may be influenced by parental involvement in the classroom or student interest in reading when deciding whether to use grouping and how to place students into homogeneous groups. It is reasonable to assume that these sorts of factors could also influence students' overall reading growth, and it is unlikely that a researcher would be able to observe all of these factors. Therefore, estimates of the effect of homogeneous grouping may be capturing the influence of those unobserved factors rather than the true effect of grouping.

Studies using random selection into homogeneous grouped or ungrouped classes could provide causal evidence about the potential effects of grouping during reading instruction. In their recent meta-analysis synthesizing more than 100 years of research on grouping, Steenbergen-Hu et al. (2016) did not identify any experimental studies that estimate the effects of within-class grouping during reading instruction. There is experimental evidence about specific reading interventions or instructional programs that encourage differentiated instruction. For example, Carol Connor and colleagues have conducted numerous randomized controlled trials to estimate the effects of a specific intervention used in early elementary classrooms to provide data-driven recommendations and plans for individualized instruction during reading instruction (Connor et al., 2011, 2013). This intensive intervention includes frequent student assessment, software providing assessment and instructional feedback, teacher training, and supports for the implementation of flexible homogeneous small-group instruction. These studies consistently find that individualized reading instruction has positive, moderately sized effects ($d = 0.25$–0.50) on reading performance among early elementary students (Al Otaiba et al., 2011; Connor et al., 2011, 2013). These findings provide additional support that flexible homogeneous grouping is an effective instructional strategy. However, given the intensive nature of this specific intervention, it is unlikely to represent the typical experience of students in homogeneously grouped classes.

Research Using the Early Childhood Longitudinal Study

Most recent research on grouping uses data from the 1998 kindergarten cohort of the ECLS-K to examine whether homogeneous grouping in early elementary grades improves reading outcomes (Adelson & Carpenter, 2011; Hong & Hong, 2009; McCoach et al., 2006) or whether grouping may lead to differential effects based on group placement (Condron, 2008; Lleras & Rangel, 2009; Tach & Farkas, 2006), student race (Lleras & Rangel, 2009; Tach & Farkas, 2006), and initial reading skills (Hong et al., 2012; Nomi, 2009). Given my interest in this analysis and the use of the ECLS-K data, I focus my review on these studies.

Those studies estimating the relationship between homogeneous grouping and reading growth in the early elementary grades typically report a small (0.06–0.10 standard deviations), positive association. Two of these studies (Adelson & Carpenter, 2011; McCoach et al., 2006) only include a few student, classroom, and school

characteristics as control variables to compare the outcomes of grouped and ungrouped students. Hong and Hong (2009)—who use propensity score matching to examine the effect of homogeneous grouping based on the instructional time spent in groups—similarly conclude that homogeneous grouping may improve kindergarten reading growth on average. Their analysis suggests that there is an interdependent relationship between instructional time and grouping such that students only receive increased benefits from more intense grouping when there is sufficient instructional time and that students only benefit from increased reading instructional time when teachers use grouping. Their findings reinforce the importance of accounting for classroom-level instructional characteristics when trying to understand the effects of grouping.

Other studies using the ECLS-K have focused on whether homogeneous grouping leads to differential outcomes based on placement in lower or higher achievement groups (Condron, 2008; Lleras & Rangel, 2009; Tach & Farkas, 2006) or initial reading skills (Hong et al., 2012; Nomi, 2009). These studies, which vary in their analytic samples and methodology, report somewhat contradictory findings about whether homogeneous grouping may be contributing to unequal reading gains across student groups. Despite methodological differences, studies examining group placement find that grouped students in the higher achievement groups perform, on average, better than comparable students in nongrouped classrooms and that grouped students in lower achievement groups perform worse, on average, than comparable students in nongrouped classrooms (Condron, 2008; Lleras & Rangel, 2009; Tach & Farkas, 2006). These authors report relatively large differences between the estimated growth of higher and lower grouped students (0.3–0.42 standard deviations) and argue that homogeneous grouping may contribute to unequal reading gains across student groups. Two studies (Hong et al., 2012; Nomi, 2009), both using propensity score matching to compare grouped students to similar ungrouped students, examine whether the effects of grouping vary based on initial reading skills. These studies present some evidence of differential effects, but the extent of the differences varies based on school context (Nomi, 2009) and instructional time (Hong et al., 2012). For example, Hong et al. (2012) find that kindergarten students with the lowest incoming reading skills experience more growth when they are homogeneously grouped, but only in classrooms with a substantial amount of time dedicated to literacy instruction. These studies suggest that homogeneous grouping—as reported by teachers surveyed as part of the ECLS-K's kindergarten cohort of 1998—may have differential effects, but there is little conclusive evidence about whether these differential effects may contribute to unequal reading gains within classrooms.

Research Questions

Equivalent comparison groups are integral to estimating the relationship between homogeneous grouping and reading growth because we cannot observe what happens to a student who experiences homogeneous grouping during their kindergarten reading instruction and what would have happened to that same student if their kindergarten classroom was not grouped. Similarly, we cannot easily observe how the same student responds to placement in different achievement groups (e.g., low, middle, high achievement groups). As grouping practices and group placement are not

randomly assigned within schools, we cannot reasonably assume that the only difference between grouped and ungrouped classrooms is the practice of grouping itself. The studies reviewed in the previous section rely on the rich set of student, teacher, and classroom characteristics collected as part of the ECLS-K to control for other, observable differences between grouped and ungrouped students in cross-sectional analyses. They are unable to control for unobserved differences between grouped and ungrouped students or for unobserved differences between students based on their group placement.

This analysis takes advantage of the longitudinal nature of the ECLS-K and the fact that not every student experiences grouping the same way in every year. Specifically, this analysis examines the following research questions:

1. To what extent is homogeneous grouping during early elementary reading instruction associated with reading growth?
2. Does this association vary based on reading group placement or initial reading skills?

By using student fixed effects, I estimate the relationship between homogeneous grouping and reading growth by comparing the growth that students make when they experience homogeneous grouping in either kindergarten, first grade, or second grade to the growth of the same student when they are ungrouped. In addition, I examine whether the estimated association between grouping and reading growth varies by reading group placement or initial reading skills.

Data and Method

Data

This study uses data from the second wave of the ECLS-K, a nationally representative study of children starting kindergarten during the 2010–2011 academic year that is conducted by the National Center for Education Statistics (NCES).[2] This longitudinal study follows the same children from kindergarten through fifth grade and collects annual waves of data, including direct cognitive assessments in reading and math, behavioral rating scales, and surveys from teachers, parents, and school administrators. The base year sample was created using a three-stage stratified sampling procedure that selected geographic areas as the primary sampling unit, then schools as the secondary sampling unit, and then kindergarten students as the tertiary sampling unit. Approximately 23 kindergarten students per school were sampled from 968 schools across the country (NCES, 2015).

Sample

To create the sample for this analysis, I first identified all students who are captured for the first 3 years of the study (54,522 student-by-year observations). I limit the sample in three ways. First, the sample is limited to students who do not switch schools during the 3 years of the study (67% of all observations). By restricting the sample in

this way, the student fixed effects also control for all time-invariant characteristics of schools. Next, the sample only includes students who completed the reading assessments conducted by NCES during the first 3 years of the study (80% of all observations). These reading assessments serve as my key dependent variables. Finally, the sample is limited to students whose teachers answered questions about their grouping practices during each wave of the study (75% of all observations), as these questions are used to create my key independent variables. Because of these restrictions, the analytic sample (32,930 observations) is no longer a nationally representative sample but still represents 14,430 students in 842 schools. To account for additional missing data in these observations for the control variables (rates of missingness ranged from 0–21%), I performed multiple imputation on the analytic sample (Rubin, 1996; Schafer, 1999). This procedure, conducted using Markov Chain Monte Carlo procedures in Stata, created 10 imputed data sets. When performing the subsequent regression analyses, all inferential statistics are estimated separately for each data set and then averaged into a single set of statistics that incorporate the uncertainty associated with the imputed values (StataCorp, 2013).

Table 1 shows descriptive statistics for the full analytic sample and distinctive groups based on whether students experienced grouping in at least 1 but not all 3 years (labeled "switchers"), grouping in all 3 years ("always grouped"), or no grouping ("never grouped"). The vast majority of early elementary students in the sample have teachers who report using achievement grouping during reading instruction. Notably, only 28% of the total sample are switchers (i.e., they experienced homogeneous grouping in kindergarten, first grade, or second grade but not in all 3 years). In the student fixed effects models, the coefficients representing the relationship between grouping and reading growth are estimated using only these students. As shown in Table 1, switchers are quite similar to the full sample on observable characteristics, but it is possible that these students are systematically different on unobserved characteristics.

Measures

Independent variables. This analysis presents two distinct sets of analyses that operationalize grouping differently: (*a*) a binary indicator for grouping and (*b*) a group placement measure. All measures are created using teacher surveys administered in the spring of each school year. Teachers are asked to report on class-wide grouping practices and the relative group placement of each individual student in the sample. Many studies using the ECLS-K create a binary variable in which students are categorized as being grouped or ungrouped (e.g., see Adelson & Carpenter, 2011; Condron, 2008; Lleras & Rangel, 2009; Nomi, 2009; Tach & Farkas, 2006). This study first estimates the effect of grouping as a binary indicator for whether the student's teacher reported that they used achievement grouping in their class during reading instruction (based on teacher responses to the survey question "How often do you divide your class or classes into achievement groups for reading activities or lessons?").

For the second set of analyses, I have created a measure of homogeneous grouping based on students' relative reading group placement using two questions from the child-specific survey completed by teachers. One survey question asks teachers to

Table 1. Descriptive Statistics by Grouping Status

		Grouping Status		
	Full Analytic Sample	Switchers	Always Grouped	Never Grouped
Student percentages:				
Ever grouped (%)	83	57	100	0
Female (%)	49	50	49	46
Race/ethnicity:				
Asian (%)	8	10	8	12
Black (%)	11	10	12	13
Hispanic (%)	24	26	23	22
Other (%)	5	6	5	5
White (%)	51	48	52	48
Language minority (%)	29	31	28	28
Disability reported (%)	16	15	16	18
Student averages (*SD*):				
Prior reading score	64.11	64.64	64.39	57.06
	(20.42)	(20.23)	(20.51)	(18.99)
Year-end reading score	80.60	81.38	80.84	72.77
	(19.92)	(19.63)	(19.87)	(20.68)
Std. socioeconomic status	.05	.06	.03	.25
	(.98)	(.98)	(.98)	(1.06)
Teacher/class averages (*SD*):				
Class size (number of students)	20.61	20.55	20.71	19.58
	(4.87)	(5.19)	(4.62)	(6.20)
Reading instruction hours per week	9.45	9.17	9.68	7.83
	(3.83)	(3.86)	(3.74)	(4.37)
Teacher's years of experience	15.04	15.72	14.68	16.13
	(9.99)	(10.30)	(9.80)	(10.55)
Class behavior (teacher-reported, 0–5 scale)	3.44	3.46	3.42	3.46
	(.87)	(.86)	(.88)	(.82)
Reading below grade level (teacher-reported, proportion of whole class)	.21	.21	.21	.17
	(.16)	(.16)	(.16)	(.15)
Total observations	32,930	9,230	22,110	1,580

Note.—Observations are pooled across 3 years of data. Percentages reported here are column percentages. Standard deviations (*SD*) are presented in parentheses for all averages. Std. = standardized.

report on their number of reading groups ("How many achievement groups in reading do you currently have in this child's class?") and another asks teachers to report on each student's individual placement in a reading group ("In which reading group is this child currently placed? Use '1' for the highest achievement group"). This information allows me to create three categories: (*a*) lowest group placement, (*b*) middle group(s) placement, and (*c*) highest group placement. Placement in one of these groups depends on the reported number of groups in each class. For example, in a class in which the teacher reports only using two groups, students may only be placed in the lowest and highest achievement groups. In a class reporting five groups, students may be placed in the lowest achievement group, highest achievement group, or in one of the three middle groups (collapsed into the middle achievement group[s] placement category). Table 2 illustrates the descriptive differences for the students in my analytic sample based on their reading group placement.

Table 2. Descriptive Statistics by Group Placement

	Group Placement			
	Ungrouped Students	Lowest Group	Middle Group(s)	Highest Group
Student averages:				
Prior reading score mean	60.48	53.77	62.25	75.60
Prior reading score *SD*	(19.86)	(16.90)	(18.36)	(20.66)
Prior reading score range	[25.73, 115.41]	[25.45, 115.25]	[25.52, 114.67]	[26.83, 116.59]
Year-end reading score mean	76.62	69.01	79.65	91.64
Year-end reading score *SD*	(20.40)	(18.79)	(18.19)	(17.35)
Year-end reading score range	[26.26, 116.21]	[26.45, 115.79]	[26.70, 115.98]	[26.83, 116.59]
Std. socioeconomic status mean	.12	−.25	0	.25
Std. socioeconomic status	(1.01)	(.93)	(.95)	(.99)
Std. socioeconomic status range	[−3.46, 3.06]	[−2.83, 3.00]	[−2.83, 3.26]	[−2.83, 3.06]
Externalizing problem behavior mean	1.67	1.84	1.68	1.57
Externalizing problem behavior *SD*	(.62)	(.68)	(.61)	(.57)
Externalizing problem behavior range	[1, 4]	[1, 4]	[1, 4]	[1, 4]
Internalizing problem behavior mean	1.52	1.72	1.54	1.43
Internalizing problem behavior *SD*	(.49)	(.57)	(.50)	(.43)
Internalizing problem behavior range	[1, 4]	[1, 4]	[1, 4]	[1, 4]
Student percentages:				
Female (%)	49	43	49	54
Race/ethnicity:				
Asian (%)	10	6	7	10
Black (%)	11	13	12	10
Hispanic (%)	25	28	24	21
Other (%)	5	5	5	6
White (%)	48	49	52	52
Language minority (%)	31	34	30	25
Disability reported (%)	16	27	15	11
Grouping experience:				
Ever ungrouped (%)	100	20	18	21
Ever placed in lowest group (%)	40	100	36	21
Ever placed in middle group(s) (%)	41	48	100	39
Ever placed in high group (%)	30	10	25	100
Total observations	5,590	5,150	13,600	8,590

Note.—Observations are pooled across 3 years of data. Percentages reported here are column percentages. Standard deviations (*SD*) are presented in parentheses and ranges presented in brackets for all averages. Std. = standardized.

For both measures of grouping, it is important to note that I cannot disentangle static grouping (in which teachers assign reading groups that do not change throughout the year) and flexible grouping (in which teachers regularly reassess and reassign reading groups). The teacher surveys only ask about grouping practices and placement at one point in time throughout the year. This limitation is examined further in the Discussion section.

Outcome variables. My key outcome variable is student performance on direct cognitive assessments that were administered at the end of each school year as part of the ECLS-K (these serve as proxies for year-end reading performance). The NCES-created scale scores are intended to measure students' reading skills, including basic skills (print familiarity, letter recognition, beginning and ending sounds, rhyming words, and word recognition), vocabulary knowledge, and reading comprehension (NCES, 2015). These scale scores have high levels of reliability (weighted reliabilities

of .91–.95), and NCES provides longitudinal scale scores specifically meant to examine student reading progress over time (Najarian et al., 2018). These assessments were only used for research purposes, and the scores were not shared with teachers. Because these assessments were given across 3 academic years and were not all given at the exact same time during each year, all models will include year fixed effects and a measure of the time between the year-end score and the assessment that serves as the prior score (described in the next section). For ease of interpretation in the models including interactions, these scores have been mean-centered for each test administration.

Student, teacher, and classroom characteristics. Student, teacher, and classroom characteristics serve as important covariates. Even in models with student fixed effects, characteristics that vary across time have also been included in some models to account for changes that could affect reading performance. Student-level variables that do not change over the course of the 3-year time period include sex, race/ethnicity, and prekindergarten status (an indicator variable for whether a student attended a center-based prekindergarten program). Student-level variables that are measured annually include age, socioeconomic status, language-minority status, parent-reported disability status, teacher-reported measures of problem behavior, and teacher-estimated absences per school year. Socioeconomic status is a continuous composite variable created by the ECLS-K that incorporates information about family income, parental education, and parental occupation. The disability status is an indicator created from an annual survey in which parents or guardians are asked whether students have a diagnosed disability. To proxy for student behavior, two scales capturing externalizing and internalizing problem behavior are included in this analysis. Both scales are created by NCES using teacher survey responses and have reliability coefficients above .75 (see Tourangeau et al., 2017, for more information). Finally, the absences variable is a teacher-reported measure in which teachers estimate the number of days that students missed that school year.

As noted above, all analyses include a measure of students' performance on direct cognitive assessments conducted by the ECLS-K that proxy for students' prior reaching performance. For kindergarten students, this score is from a reading assessment that was administered at the beginning of kindergarten. For first- and second-grade students, this score is from the reading assessment administered at the end of the prior year. All models include year fixed effects and a student-specific measure of months between the "prior" assessment and "year-end" assessment.

Teacher and classroom covariates are added in some models. These variables include teacher's years of teaching experience, whether the teacher has a graduate degree, overall class size, reading instructional hours per week, a teacher rating of the class's overall behavior, the number of reading achievement groups that the teacher reports using, and classroom composition variables. The classroom composition variables are based on teacher's responses to annual surveys and include proportion of female students, proportion of Black students, proportion of Hispanic students, proportion of White students, proportion of gifted students, proportion of students reading below grade level, proportion of students receiving services for English-language learners, and proportion of students with disabilities. Finally, I include two variables—the class average socioeconomic status and the class average prior reading score—that are constructed based on the average for each class for students enrolled in the ECLS-K. The students in the sample were selected at random from each school but were not

necessarily selected to be representative of their classroom. As such, these averages are not true classroom averages but the best approximation based on the available data.

Analytic Approach

This study uses covariate-adjusted regression with student fixed effects to estimate the relationship between homogeneous grouping and reading growth across 3 years of panel data for the kindergarten, first-grade, and second-grade years of the ECLS-K. I include two distinct analyses in this article, one of which includes a binary indicator capturing whether the students' teacher reports using homogeneous grouping and one of which captures students' achievement group placement (e.g., lowest, middle, highest achievement group). In each analysis, I use the same set of models. The base model includes a grouping measure, a prior reading score, time between testing periods, and year fixed effects. The next model adds student-level characteristics. The full model adds teacher/classroom-level characteristics. For each model, I include a set of results without student fixed effects and a set of results with student fixed effects. The set of results without student fixed effects are estimated using the full sample. The coefficient on the binary indicator of grouping estimates the difference in reading growth between grouped and ungrouped students while controlling for all other covariates in the models. In the student fixed effects models, this coefficient estimates the difference in reading growth between a student in their grouped year(s) and the same student in their ungrouped year(s) while controlling for the time-varying characteristics of students and classrooms included in the models. Importantly, the models that include student fixed effects estimate the association between grouping and reading performance only for switchers (i.e., students whose grouping experiences vary across the 3 years of data). Although models with student fixed effects are estimated using a subset of the data, the benefit of this approach is that the student fixed effects capture all unobserved, time-invariant factors about these students and their schools that may influence a student's reading performance. The entire analytic sample (even those whose grouping experiences do not change over this 3-year period) is included in all models to more precisely estimate the relationship between the covariates and student reading growth.

The full model with a student fixed effect term is given as

$$\begin{aligned} \text{read}_{it} &= \beta_0 + \beta_1 \text{grp}_{it} + \beta_2 \text{prior}_{it} + \beta_3 \text{testpd}_{it} + \beta_k X_{it} + \beta_m W_{it} + \theta_i \\ &\quad + \theta_t + \varepsilon_{it}, \quad t = 1, 2, 3. \end{aligned}$$

The outcome variable, read_{it}, represents the year-end reading score for student i in year t. The time periods represent the 3 years of data: kindergarten = 1, first grade = 2, and second grade = 3. The first coefficient (β_1) estimates the effect of grp_{it}, which measures student i's experience with homogeneous grouping in year t. For each set of analyses, grouping will be represented by one of the two grouping variables described previously: (*a*) the grouping binary indicator or (*b*) the group placement variable. The full model includes prior-year reading score (prior_{it}), time between testing periods (testpd_{it}), a vector of student-level characteristics (X_{it}), a vector of teacher/classroom-level characteristics (W_{it}), student fixed effects (θ_i), year fixed effects (θ_t), and an idiosyncratic error term (ε_{it}) that captures any unobserved, time-varying factors that influence the outcome variable. All standard errors are clustered at the student level.

Results

To answer the first research question, I present a set of regression models in Table 3 that estimate differences in reading growth between homogeneously grouped and ungrouped students. In Column 1, the model only includes a binary indicator for grouping, the student's prior reading score, a variable measuring months between testing administrations, and year fixed effects. In this case, homogeneously grouped students are predicted to have reading growth that is 0.64 points higher than students who do not experience grouping, all else equal (when using a standardized outcome variable, this is the equivalent of a 0.03 standard deviation difference). As discussed previously, students who experience grouping are likely different than those who are ungrouped across a number of dimensions. Although this model accounts for differences in prior reading score, it does little to ensure that other differences between grouped and ungrouped students are being properly controlled. Column 2 introduces student fixed effects and controls for all time-invariant student characteristics that could influence reading growth. As a result, the interpretation of the grouping indicator changes to estimate the difference in reading growth between a given student in a year when they experience homogeneous grouping and that same student in a year when they do not experience grouping (only for students labeled switchers in Table 1). As Column 2 illustrates, these students are predicted to have reading growth that is 0.76 points higher in the year that they experienced grouping than in the year they are ungrouped (equivalent of 0.04 standard deviation difference).

Additional covariates are added in subsequent models. The full model, shown in Column 6, includes student, teacher, and classroom covariates and student fixed effects. This model estimates that, on average, students are predicted to experience slightly more reading growth (0.66 points) in a year that they are homogeneously grouped during reading instruction than in a year in which they are not homogeneously grouped in reading, all else equal (equivalent of 0.03 standard deviation difference). In all models, the grouping indicator is a statistically significant and positive predictor of reading growth, suggesting that these results are robust to multiple specifications. This estimated difference is relatively small when compared with the typical growth that these students experience in reading skills, as measured by the cognitive assessments administered by the ECLS-K. On average, students in the sample demonstrate 16.5 points of reading growth between the two test administrations. This estimated difference of 0.66 points between grouped and ungrouped students represents about 4% of this overall growth.

This overall estimate may conceal variation in the relationship between grouping and reading growth. To answer the second research question, the remaining analyses

Table 3. Regression Results for the Relationship Between Homogeneous Grouping (Binary Indicator) and Reading Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Homogeneous grouping	.64***	.76***	.71***	.77***	.76***	.66***
	(.11)	(.14)	(.11)	(.14)	(.14)	(.17)
Student characteristics			X	X	X	X
Teacher/class characteristics					X	X
Student fixed effects		X		X		X
Observations	32,930	32,930	32,930	32,930	32,930	32,930

Note.—All models include students' prior reading score, time between testing administrations, and year fixed effects. Standard errors are reported in parentheses.

*** $p < .001$.

explore whether this relationship varies based on group placement and initial reading skills. Table 4 presents results from the same set of models with a categorical variable for group placement. The reference group is still ungrouped students, so these estimates represent the average difference in reading growth between each reading group category (lowest group, middle group[s], and highest group) and ungrouped students, while controlling for the same covariates as before. Reflecting the trends found in earlier research (Condron, 2008; Lleras & Rangel, 2009), the models without student fixed effects (Columns 1, 3, and 5) find that the lowest group students, on average, experience significantly less reading growth than ungrouped students (−3.28 points in the full model presented in Column 5) and the highest group students grow significantly more than ungrouped students (+3.40 points in the full model presented in Column 5). In all three models without student fixed effects, these differences are both statistically significant and practically meaningful. For example, the estimated difference in reading growth between the lowest grouped students and highest grouped students (as illustrated in the full model in Column 5) is equivalent to 40% of the average reading growth exhibited by the students in this sample (equivalent of 0.34 standard deviations). These initial results support the hypothesis that homogeneous grouping can substantially widen reading achievement gaps between the students placed in the lowest reading groups and the students placed in the highest reading groups.

Once student fixed effects are included in Columns 2, 4, and 6, the overall pattern of results is similar but the estimated difference in reading growth between the lowest grouped and highest grouped students shrinks considerably (to about one third of the estimated difference in the models without fixed effects). These models leverage changes in group placement (lowest group, middle group[s], highest group, or ungrouped) for the same student across the 3 years of data to estimate differences in reading growth. As Table 2 illustrates, most students in each group placement (i.e., lowest, middle, or highest groups) have either been ungrouped or been placed in other placements during other years. This difference in magnitude suggests that covariate-adjusted regression may not be adequately controlling for selection mechanisms that influence both group placement and reading growth. For example, factors such as student motivation

Table 4. Regression Results for the Relationship Between Reading Group Placement and Reading Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Group placement:						
Lowest group	−3.45***	−.47*	−3.21***	−.46**	−3.28***	−.57**
	(.17)	(.20)	(.16)	(.20)	(.17)	(.21)
Middle group(s)	.87***	.74***	.83***	.75***	1.01***	.66***
	(.13)	(.16)	(.12)	(.16)	(.15)	(.19)
Highest group	2.82***	1.50***	2.95***	1.51***	3.40***	1.47***
	(.14)	(.17)	(.14)	(.17)	(.15)	(.19)
Student characteristics			X	X	X	X
Teacher/class characteristics					X	X
Student fixed effects		X		X		X
Observations	32,930	32,930	32,930	32,930	32,930	32,930

Note.—All models include students' prior reading score, time between testing administrations, and year fixed effects. Standard errors are reported in parentheses.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

or family reading habits may influence teachers' placement of students into reading groups and students' ultimate reading growth but cannot be observed by the researcher.

Although the differences across models provide evidence that covariate-adjusted regression may not appropriately isolate the relationship between group placement and reading growth, the patterns of results still support the hypothesis that there are differential relationships between grouping and reading growth by group placement. Reading growth is predicted to be significantly higher (1.47 points in the full model in Column 6) during the years that a student is placed in the highest reading groups in their classroom when compared with years in which that same student is ungrouped. Similarly, the reading growth for students when they are the lowest reading group in their class is predicted to be 0.57 points lower than when that student is ungrouped. Thus, students in the highest reading groups appear to benefit more from being grouped. This estimated difference between the growth of the highest grouped and lowest group students (2.04 points) is about 12% of the average reading growth exhibited by students in this study (equivalent of a 0.10 standard deviation difference).

Table 5 presents analyses that examine whether the relationship between homogeneous grouping and reading growth varies based on students' initial reading skills.[3] The first analysis, shown in Panel A of Table 5, interacts the mean-centered prior

Table 5. Differential Relationships Between Grouping and Reading Performance, by Initial Reading Skills

	Panel A: Grouping Indicator		Panel B: Group Placement	
	(5)	(6)	(5)	(6)
Prior reading score	.78***	.02	.77**	.03*
	(.01)	(.01)	(.01)	(.01)
Grouping indicator	.79***	.68***		
	(.14)	(.17)		
Prior reading score × grouping indicator	−.04***	−.02*		
	(.01)	(.01)		
Group placement:				
Lowest group			−2.38***	−.25
			(.18)	(.22)
Middle group(s)			.89***	.66***
			(.15)	(.19)
Highest group			4.38***	1.66***
			(.17)	(.20)
Prior reading score × group placement:				
Score × lowest group			−.01	−.01
			(.01)	(.02)
Score × middle group(s)			−.14***	−.07***
			(.01)	(.01)
Score × highest group			−.21***	−.06***
			(.01)	(.02)
Student characteristics	X	X	X	X
Teacher/class characteristics	X	X	X	X
Student fixed effects		X		X
Observations	32,930	32,930	32,930	32,930

Note.—All models include students' prior reading score, time between testing administrations, and year fixed effects. Standard errors are reported in parentheses.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

reading score and the grouping indicator variable to examine the extent to which the estimated relationship between grouping and reading growth may vary over different initial skill levels (only the full models are shown in Table 5, the equivalent of Columns 5 and 6 in Table 3). In this model, the coefficient for grouping estimates the difference in the intercept between grouped and ungrouped students when the prior reading score is at its mean. In the student fixed effects model (shown in Panel A, Column 6), grouped students with average prior reading scores are estimated to experience slightly more reading growth than ungrouped students (0.68 points). The interaction term between prior reading score and the grouping indicator estimates the difference in slope for prior reading score between ungrouped and grouped students. The statistically significant and negative coefficient (–0.02) for this interaction signals that the estimated difference in reading growth between grouped and ungrouped students decreases as prior reading score increases.

The analysis presented in Panel B of Table 5 and illustrated in Figure 1 examines whether there is a differential relationship between reading group placement and reading growth based on prior reading score. Figure 1 displays differences in the predicted mean-centered reading growth by group placement across different levels of initial reading skills (as shown in Column 6 in Panel B of Table 5) estimated using Stata's margins command and holding all other covariates at their mean. Across all levels of initial reading skills, the predicted reading growth of students grouped in the lowest reading group is never statistically different from ungrouped students. The estimated differences in reading achievement between students grouped in the middle/highest groups and ungrouped students are higher among students who start with initial reading skills that are below the sample average. For students whose initial

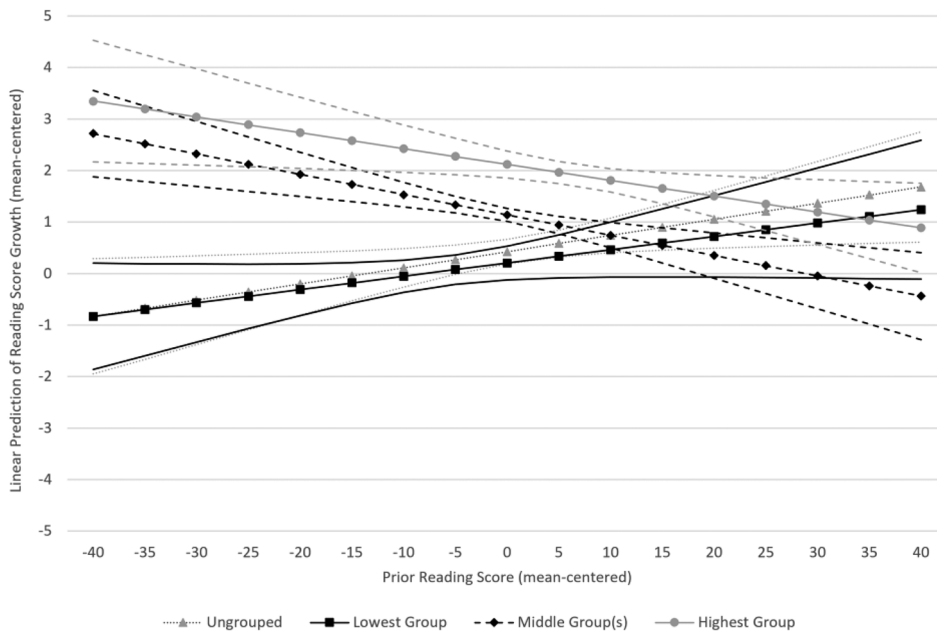

Figure 1. Differential relationship between reading group placement and reading performance, by initial reading skills. Linear predictions estimated using Column 6 from Panel B in Table 5, with all other covariates held at their mean.

reading skills are more than 20 points below average (one standard deviation), the predicted reading growth for ungrouped students and students placed in the lowest reading group are virtually identical (and not statistically different) whereas students placed in the middle reading group(s) and highest reading group are predicted to experience 2–4 more points of reading growth than ungrouped students. For students whose initial reading skills are more than 20 points above average (one standard deviation), there are not statistically significant differences in reading growth between ungrouped students and students placed in any of the reading groups (lowest, middle, or highest).

To better account for these differential outcomes, I descriptively examined the classroom composition and environment of students whose initial reading skills are below the sample average but who are placed in different reading group placements. Although most classroom characteristics appear descriptively similar across students in the three group placements, students with lower than average initial reading skills but who are placed in higher reading groups tend to be found in classrooms with lower average initial reading skills and less variation (as proxied by the class mean and standard deviation of prior reading scores of students sampled in the ECLS-K study). This may suggest that students' initial reading skills relative to their class may be more important than their skills relative to a national sample.[4] Taken together, the findings from the second research question indicate that homogeneous grouping is associated with reading growth for students with lower initial reading skills, but only for students placed in the middle/highest reading groups in their class.

Discussion

Building on prior work on homogeneous grouping during elementary reading instruction, my analysis finds that the practice of homogeneous grouping has a small, positive association with reading growth. This result mirrors past research using the 1998 ECLS-K (Adelson & Carpenter, 2011; Hong & Hong, 2009; McCoach et al., 2006), but the magnitude of the estimated difference in reading growth between grouped and ungrouped students is slightly smaller in my analysis. However, this overall estimate masks a differential relationship by group placement and initial reading skills. My analyses interacting initial reading skills with grouping or group placement suggest that homogeneous grouping is only associated with reading growth for students with initial reading skills that are lower than the sample average and, among those students, students only appear to benefit from grouping if they are placed in the middle or highest achievement groups. This finding clarifies past research on grouping, which has separately found differences by group placement (Condron, 2008; Lleras & Rangel, 2009; Tach & Farkas, 2006) and by initial reading skills (Hong et al., 2012).

In this concluding section, I discuss the methodological implications and limitations of this study, implications for practice, and directions for future research.

Methodological Implications

Numerous factors may influence teachers' decisions to use grouping and to place specific students into groups, including the composition and needs of their particular

class. Given that observable characteristics vary between grouped and ungrouped students and between students placed in the lowest, middle, and highest achievement groups, we can likely assume that these groups vary in unobservable ways that may affect students' reading growth and teachers' decision making about grouping. As a result, these groups are not equal in expectation, a condition for estimating the causal effect of grouping (Murnane & Willett, 2010). By using student fixed effects, this analysis creates a different comparison in which a student's reading growth during a year that they experience homogeneous grouping is compared with their reading growth when in an ungrouped class. This estimation strategy accounts for the effect of all observed and unobserved time-invariant differences among students and, because the sample is limited to students who do not switch schools, all time-invariant differences among schools. Although I do not present these results as causal, this approach offers an alternative to isolating the relationship between homogeneous grouping and reading growth that addresses some of the criticism about appropriate comparison groups in past research on grouping (Ferguson, 1998).

Prior research mainly relies on covariate-adjusted regression and propensity score matching to compare the reading growth of grouped students to that of ungrouped students who appear the most similar on observable characteristics. When I estimate the overall association between homogeneous grouping and reading growth, I find similar results using both covariate-adjusted regression and student fixed effects models. However, in the analyses that examine differential association by group placement, the estimated difference in reading growth between students placed in the highest reading groups compared with students placed in the lowest reading groups is three times higher in the covariate-adjusted regression model than in the student fixed effects model. The difference in reading growth estimated in the covariate-adjusted regression model is equivalent to 40% of the average reading growth exhibited by students in the analytic sample, and this estimated magnitude is very similar to those reported in past studies that examine differences in reading growth by group placement (Condron, 2008; Lleras & Rangel, 2009). Although there is still a significant association between group placement and reading growth in the student fixed effects models, the magnitude of the estimated difference in reading growth between lowest grouped students and highest grouped students shrinks to the equivalent of 12% of the average reading growth exhibited by students in this sample. Covariate-adjusted regression is likely not adequately accounting for student characteristics influencing group placement. This suggests that estimated differential effects based on group placement may be overstated in prior research.

Although the use of student fixed effects addresses some limitations of past work, it is still an imperfect method for estimating whether students benefit from homogeneous grouping. First, the grouping coefficients estimated in models with student fixed effects only estimate the relationship between grouping and reading growth for those students who have different experiences in grouping over their first 3 years of elementary school. The results of these analyses do not generalize to students whose experience with grouping (e.g., always homogeneously grouped) or group placement (e.g., always placed in the lowest group) do not vary over these 3 years. This restriction is more problematic for the analysis of grouping as a binary indicator because only 28% of the sample switches from grouped to ungrouped classrooms over the 3 years of the study. In the analysis of group placement, more than half of the sample switches

group placement across the 3 years of the data, so the restriction to switchers is not as limiting. This choice represents a trade-off between the internal validity of the estimated relationship between grouping and reading growth and the external validity of these findings.

Furthermore, the fixed effect models do not control for all time-varying characteristics about students and their classrooms that could influence both grouping practice and reading growth. The set of analyses estimating the relationship between homogeneous grouping and reading growth could be biased if they omit teacher or classroom characteristics that influence how teachers approach grouping and students' ultimate reading growth. If, for example, more effective teachers systematically chose to homogeneously group their students, then the estimated association between grouping and reading growth would be capturing the effect of both grouping and teacher effectiveness. Given the small magnitude of the association between the practice of grouping and reading growth, such a scenario is unlikely. However, this analysis cannot rule out whether unobserved teacher or classroom characteristics may be influencing the results. In particular, a teacher's decision about whether and how to use grouping may be driven by their specific pedagogical beliefs, certain aspects of their classroom composition and relationships among students, and nuanced understandings of students' learning needs. It is difficult to capture all of these factors in this sort of analysis. Future analyses that follow teachers and students longitudinally may be better positioned to eliminate such confounding variables. Although this analysis controls for time-invariant characteristics of students as well as a set of time-varying measures about students, there may be unobserved time-varying student traits—such as changes in a student's family stability, interest in reading, or orientation toward school—that could bias these results.

Despite these limitations, this approach does offer a valuable alternative for estimating the relationship between homogeneous grouping and student performance. By controlling for all time-invariant characteristics of students and schools (observed and unobserved), the student fixed effects models improve on past research that only includes observed characteristics.

Implications for Practice

These findings also have practical implications for school leaders and teachers considering whether to use homogeneous grouping in their classrooms. The results indicate that, on average, students benefit from being homogeneously grouped during early reading instruction. The difference in reading growth between grouped and ungrouped students estimated here is quite modest, representing just 4% of the average annual growth observed in this sample. This supports the hypothesis that homogeneous grouping, on average, may be a slightly more effective instructional practice than whole-group instruction.

More importantly, these findings provide evidence that the relationship between homogeneous grouping and reading growth depend on reading group placement and initial reading skills. In particular, this study finds little evidence that homogeneous grouping is related to reading growth for students whose initial reading skills are above average. For students whose initial reading skills are below average, homogeneous grouping is positively associated with reading growth for students who are

placed in the middle or highest reading groups in their class. For students who are placed in the lowest reading group in their class, there is no discernable difference in their reading growth when compared with their growth in an ungrouped year. This suggests that the experience of homogeneous grouping may be qualitatively different for students placed in the lowest reading groups. Given that students placed in higher reading groups appear to benefit more from grouping, these results support the hypothesis that homogeneous grouping increases inequalities between students within the same class (Entwisle & Alexander, 1993; Gamoran, 1992, 2011; Pallas et al., 1994). Teachers and school leaders deciding whether and how to group their students for reading instruction should think carefully about the instructional experiences of students within different groups and consider ways to ensure that students within the lowest reading groups receive equal access to rigorous instruction and materials.

Directions for Future Research

How teachers create reading groups and how instruction is differentiated by group can vary substantially from teacher to teacher (Buttaro et al., 2010; Eder, 1983; Moody et al., 1997; Schumm et al., 2000). In their study of how 30 elementary school classrooms utilize grouping during literacy instruction, Schumm et al. (2000) concluded that the time spent in groups varied across and within classrooms, with some teachers starting and stopping the practice throughout the academic year. This study is limited in its ability to measure how teachers use homogeneous groups, as I rely on limited, self-reported information from teachers about their instructional practices and classroom composition. This analysis cannot capture the complexity of the many different ways that teachers use homogeneous groups and only offers limited guidance for educators seeking to make decisions about their reading instruction. The teacher survey data included in the ECLS-K only capture how teachers report on the number of reading groups, each child's relative placement within those groups, and the duration and frequency of those groups at one point in time.

Future research on grouping should consider alternative ways to capture teachers' grouping practices and potential effects on reading growth. Although focusing on a specific intervention in individualized reading instruction, the longitudinal, randomized controlled trials conducted by Carol Connor and colleagues offer a model for capturing both grouping practices and reading growth over time among early elementary students (Connor et al., 2013). Researchers could conduct observational studies that capture the specific ways that teachers create and use homogeneous groups or regularly survey teachers about their use of grouping and the relative group placement of their students. This research would be particularly powerful if it could be combined with independent assessment of students' reading growth over time and use student fixed effects models to further disentangle the relationship between grouping practices, students' placement in reading groups, initial reading skills, and reading growth.

Conclusion

Very few studies have utilized the newer kindergarten cohort of the ECLS-K to examine homogeneous grouping (see Wrabel et al., 2015), and none have specifically

looked at differential relationships based on achievement group placement or initial reading skills. Given the shifts in enrollment demographics, better prekindergarten preparation, and greater academic focus on literacy found in studies of more recent kindergarten cohorts (Bassok et al., 2016), it is important to conduct analyses using more recent data from early elementary classrooms. In fact, the continued shift within early elementary grades to focus more on literacy skills may have contributed to the increased prevalence of homogeneous grouping and indicates that this research focus is increasingly salient for early elementary students and their teachers.

Notes

Susan Kemper Patrick is a postdoctoral scholar at the Tennessee Education Research Alliance and the Department of Leadership, Policy, and Organizations at Vanderbilt University. Correspondence may be sent to Susan Kemper Patrick at susan.k.patrick@vanderbilt.edu.

1. Throughout this text, I use the terms "homogeneous grouping" and "achievement grouping." I prefer these terms to "ability grouping" because I believe most grouping decisions are based on measures of student performance or achievement rather than ability, which cannot be easily measured. These terms also better reflect the wording used in the teacher survey questions within the ECLS-K that are used to create the measures of grouping in this analysis.

2. This analysis uses data from the restricted-use version of the ECLS-K's kindergarten cohort of 2010. As required by NCES, all sample sizes have been rounded to the nearest 10.

3. These prior reading scores do not account for students' initial reading skills relative to their peers in their classroom or school, although the regression models do include a covariate that is an approximation of the classroom average prior reading score for all the students in the classroom enrolled in the ECLS-K.

4. The sampling strategy of ECLS-K (in which only some students from every kindergarten classroom are included in the initial sample) makes it difficult to empirically test this assertion.

References

Adelson, J. L., & Carpenter, B. D. (2011). Grouping for achievement gains: For whom does achievement grouping increase kindergarten reading growth? *Gifted Child Quarterly*, **55**(4), 265–278. https://doi.org/10.1177/0016986211417306

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data–informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *Elementary School Journal*, **111**(4), 535–560. https://doi.org/10.1086/659031

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, **46**(3), 399–424.

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, **2**(1), 1–31. https://doi.org/10.1177/2332858415616358

Bourdieu, P. (1977). Cultural reproduction and social reproduction. In J. Karabel & A. H. Halsey (Eds.), *Power and ideology in education* (pp. 487–511). Oxford University Press.

Buttaro, A., Catsambis, S., Mulkey, L., & Steelman, L. (2010). An organizational perspective on the origins of instructional segregation: School composition and use of within-class ability grouping in American kindergartens. *Teachers College Record*, **112**(5), 1300–1337.

Chorzempa, B. F., & Graham, S. (2006). Primary-grade teachers' use of within-class ability grouping in reading. *Journal of Educational Psychology*, **98**(3), 529–541. https://doi.org/10.1037/0022-0663.98.3.529

Condron, D. J. (2007). Stratification and educational sorting: Explaining ascriptive inequalities in early childhood reading group placement. *Social Problems*, **54**(1), 139–160. https://doi.org/10.1525/sp.2007.54.1.139

Condron, D. J. (2008). An early start: Skill grouping and unequal reading gains in the elementary years. *Sociological Quarterly*, **49**(2), 363–394. https://doi.org/10.1111/j.1533-8525.2008.00119.x

Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science*, **24**(8), 1408–1419. https://doi.org/10.1177/0956797612472204

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J. R., Lundblom, E., Crowe, E. C., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristics by reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*, **4**(3), 173–207. https://doi.org/10.1080/19345747.2010.510179

Cook, T. D., Steiner, P. M., & Pohl, S. (2009). How bias reduction is affected by covariate choice, unreliability, and mode of data analysis: Results from two types of within-study comparisons. *Multivariate Behavioral Research*, **44**(6), 828–847. https://doi.org/10.1080/00273170903333673

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. Irvington.

Dreeben, R., & Gamoran, A. (1986). Race, instruction, and learning. *American Sociological Review*, **51**(5), 660–669. https://www.jstor.org/stable/2095491

Duncan, G. J., & Murnane, R. J. (2014). *Restoring opportunity: The crisis of inequality and the challenge for American education*. Harvard Education Press.

Eder, D. (1983). Organizational constraints and individual mobility: Ability group formation and maintenance. *Sociological Quarterly*, **24**(3), 405–420. https://doi.org/10.1111/j.1533-8525.1983.tb00710.x

Entwisle, D. R., & Alexander, K. L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology*, **19**(1), 401–423. https://doi.org/10.1146/annurev.so.19.080193.002153

Ferguson, R. F. (1998). Can schools narrow the Black–White test score gap? In C. Jencks & M. Phillips (Eds.), *The Black–White test score gap* (pp. 318–374). Brookings Institution.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, **59**(4), 185–198. https://doi.org/10.2307/2112346

Gamoran, A. (1992). Synthesis of research: Is ability grouping equitable? *Educational Leadership*, **50**, 11–17.

Gamoran, A. (2011). Designing instruction and grouping students to enhance the learning of all: New hope or false promise? In M. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 111–126). Springer. https://doi.org/10.1007/978-94-007-1576-9_6

Haller, E. J. (1985). Pupil race and elementary school ability grouping: Are teachers biased against Black children? *American Educational Research Journal*, **22**(4), 465–483. https://doi.org/10.2307/1163135

Hallinan, M. T., & Sorensen, A. B. (1983). The formation and stability of instructional groups. *American Sociological Review*, **48**(6), 838–851. https://www.jstor.org/stable/2095329

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. https://doi.org/10.3102/0162373711424206

Hong, G., & Hong, Y. (2009). Reading instruction time and homogeneous grouping in kindergarten: An application of marginal mean weighting through stratification. *Educational Evaluation and Policy Analysis*, **31**(1), 54–81. https://doi.org/10.3102/0162373708328259

Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education*, **26**(3), 219–231. https://doi.org/10.1080/08957347.2013.793187

Kulik, J. A., & Kulik, C.-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, **36**(2), 73–77. https://doi.org/10.1177/001698629203600204

Lleras, C., & Rangel, C. (2009). Ability grouping practices in elementary school and African American/Hispanic achievement. *American Journal of Education*, **115**(2), 279–304. https://doi.org/10.1086/595667

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. https://doi.org/10.2307/1170650

Loveless, T. (2013, March 18). *2013 Brown Center Report on American Education: How well are American students learning?* Brookings Institution. https://www.brookings.edu/research/2013-brown-center-report-on-american-education-how-well-are-american-students-learning/

McCoach, D. B., O'Connell, A. A., & Levitt, H. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *Journal of Educational Research*, **99**(6), 339–346. https://doi.org/10.3200/JOER.99.6.339-346

Missett, T. C., Brunner, M. M., Callahan, C. M., Moon, T. R., & Price Azano, A. (2014). Exploring teacher beliefs and use of acceleration, ability grouping, and formative assessment. *Journal for the Education of the Gifted*, **37**(3), 245–268. https://doi.org/10.1177/0162353214541326

Moody, S. W., Vaughn, S., & Schumm, J. S. (1997). Instructional grouping for reading: Teachers' views. *Remedial and Special Education*, **18**(6), 347–355. https://doi.org/10.1177/074193259701800604

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Najarian, M., Tourangeau, R., Nord, C., & Wallner-Allen, K. (2018). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–2011 (ECLS-K:2011), first- and second-grade psychometric report* (NCES 2018-183). National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

National Center for Education Statistics. (2015). *User's manual for the ECLS-K:2011 kindergarten–first grade data file and electronic codebook, public version*. US Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015078

Nomi, T. (2009). The effects of within-class ability grouping on academic achievement in early elementary years. *Journal of Research on Educational Effectiveness*, **3**(1), 56–92. https://doi.org/10.1080/19345740903277601

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). Yale University Press.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, **1**(2), 117–175. https://doi.org/10.1207/s1532690xci0102_1

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education*, **67**(1), 27–46. https://www.jstor.org/stable/2112748

Park, V., & Datnow, A. (2016). Ability grouping and differentiated instruction in an era of data-driven decision making. *American Journal of Education*, **123**(2), 281–306. https://doi.org/10.1086/689930

Quinn, D. M. (2015). Kindergarten Black–White test score gaps: Re-examining the roles of socioeconomic status and school quality with new data. *Sociology of Education*, **88**(2), 120–139. https://doi.org/10.1177/0038040715573027

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, **91**(434), 473–489.

Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge University Press.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, **8**(1), 3–15. https://doi.org/10.1177/096228029900800102

Schumm, J. S., Moody, S. W., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities*, **33**(5), 477–488. https://doi.org/10.1177/002221940003300508

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, **57**(3), 293–336. https://doi.org/10.3102/00346543057003293

StataCorp. (2013). *Stata multiple-imputation reference manual*. https://www.stata.com/manuals13/mi.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. https://doi.org/10.3102/0034654316675417

Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research*, **35**(4), 1048–1079. https://doi.org/10.1016/j.ssresearch.2005.08.001

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*, **101**(2), 121–165. https://doi.org/10.1086/499662

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, **46**(1), 90–118. https://doi.org/10.1080/00273171.2011.540475

Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms*. Association for Supervision & Curriculum Development.

Tourangeau, K., Nord, C., Le, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Najarian, M. (2017). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–2011 (ECLS-K:2011) user's manual for the ECLS-K:2011 kindergarten–second grade data file and electronic codebook*. National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, **44**(3), 519–558. https://doi.org/10.3102/0002831207306859

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wrabel, S. L., Gottfried, M., & Polikoff, M. S. (2015, February 27). *Instructional groupings in the inclusive kindergarten classroom: A cross-cohort analysis* [Paper presentation]. Association for Education Finance and Policy Annual Conference.