Received: 11 October 2016 | Revised: 17 March 2017 | Accepted: 27 March 2017

DOI: 10.1002/icd.2036

WILEY

RESEARCH ARTICLE

Multigrade kindergarten classrooms and children's academic achievement, executive function, and socioemotional development

Arya Ansari

University of Virginia, Charlottesville, VA, USA

Correspondence
Arya Ansari, Center for Advanced Study of Teaching and Learning, University of Virginia, PO Box 800784, Charlottesville, VA 22908-0784, USA.
Email: aa2zz@eservices.virginia.edu

Funding information
National Institute of Child Health and Human Development, Grant/Award Number: R24 HD42849, T32 HD007081-35, and R01 HD069564; Institute of Education Sciences, U.S. Department of Education, Grant/Award Number: R305B130013

Abstract

Using data from the Early Childhood Longitudinal Study Kindergarten Class of 2010–2011 (n = 11,000), this study examined the developmental outcomes of 5-year-old children in multigrade classrooms (combined prekindergarten and kindergarten classrooms serving 3-, 4-, and 5-year-olds) compared with those of 5-year-olds attending kindergarten-only classrooms serving primarily 5-year-olds. Results from regression and propensity score analyses revealed that 5-year-old children who attended multigrade classrooms with prekindergarteners made smaller gains in math and literacy skills and demonstrated less optimal executive function at the end of the school year than children who attended kindergarten-only classrooms. Classroom-level factors largely explained the differences in children's academic achievement but did not consistently explain differences in their executive functioning. No consistent differences emerged for children's social-behavioral development.

Highlights

- This paper examines the implications of multigrade kindergarten classrooms for 5-year-olds' school success and the mechanisms through which these classrooms affect children.
- Five-year-olds who attended multigrade classrooms demonstrate smaller gains in academics and exhibit less optimal executive function, which was partially explained by classroom-level factors.
- When possible, the implementation of multigrade kindergarten classrooms may need to be re-evaluated.

KEYWORDS

academic achievement, ECLS-K: 2011, executive function, multigrade kindergarten classrooms, socioemotional skills

Copyright © 2017 John Wiley & Sons, Ltd.

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1 | INTRODUCTION

The multigrade education philosophy—defined as grouping children of two or more grades into a single classroom—has long been a part of the U.S. educational system (Little, 2001; Veenman, 1995; Vincent, 1999). During the early elementary school years, these classrooms serve up to 5% of children across the country (Thomas, 2012), and although they are often encouraged from an educational standpoint, they typically arise because of economic and logistical constraints (Burns & Mason, 2002; Little, 2001; Veenman, 1995). Whether these classrooms help (or hinder) children's school success, however, has been a subject of great debate. Grounded in developmental theories of cognitive development (Vygotsky, 1978) and social learning (Bandura, 1986), some scholars argue that multigrade classrooms can be more effective than the single-grade model because children of various skills can stimulate each other's learning through modeling and scaffolding (Veenman, 1995). In contrast, other scholars posit that children in multigrade classrooms may exhibit fewer gains in early learning and development because teachers may provide children with less challenging content as a means of accommodating a wide range of skill levels, resulting in disengagement among older and more skilled children (Mason & Burns, 1996; Urberg & Kaplan, 1986).

Given the conflicting empirical evidence and the long-standing model of single-grade education, multigrade classrooms require continued attention. Such inquiry is especially necessary during the early elementary school years when children's skills are more malleable (Bornstein, Hahn, Putnick, & Suwalsky, 2014), and kindergarten, in particular, which has grown exponentially over the last several decades (Davis & Bauman, 2013) and constitutes a critical developmental period for shaping children's long-term life success (Duncan et al., 2007; Entwisle & Alexander, 1989; Jones, Greenberg, & Crowley, 2015). Thus, with the rise in the number of children attending kindergarten, coupled with the heightened emphasis on children's early learning (Bassok, Latham, & Rorem, 2016), we need to know whether the multigrade model is more or less effective than the single-grade model when implemented during the transition to school. The current study addresses these gaps in the literature by using data from the Early Childhood Longitudinal Study Kindergarten Class of 2010–2011 and contrasting the early learning of 5-year-olds attending multigrade classrooms that serve both preschoolers and kindergarteners and kindergarten-only classrooms. It should be noted this study cannot speak to the benefits or drawbacks of these two types programs for (a) the younger children (e.g., 3- and 4-year-olds) and (b) for 5-year-olds enrolled in classrooms with older children, as these types of questions are beyond the scope of this study and the data available.

1.1 | Multigrade classrooms and the transition to school

Regardless of whether multigrade classrooms are put into practice for pedagogical or logistical reasons, the evidence behind multigrade education is methodologically limited and has focused on the experiences of older children in the later elementary school years (e.g., first or second grade and beyond) rather than of children during the transition to school (Veenman, 1995). Prior evaluations of multigrade educational programs have also been ambiguous. Studies of multigrade classrooms in Virginia (second to fifth grades; Pratt, 1986), New York (second to fifth grades; Way, 1981), and California (kindergarten to third grade; Purl & Curtis, 1970) have demonstrated academic effects that were positive, negative, and statistically indistinguishable, respectively, and that aggregate to zero (Pratt, 1986; Veenman, 1995). Similarly, analyses of elementary schools that have transitioned from a single-grade to multigrade model for children between the first and eighth grade years have demonstrated no significant changes in children's test scores (Proehl, Douglas, Elias, Johnson, & Westsmith, 2013). In other words, despite the strong and long-standing advocacy in support of multigrade educational programs (e.g., Katz, Evangelou, & Hartmann, 1990), the extant literature suggests that the multigrade education model produces similar academic outcomes as compared with the single-grade model.

Most of these studies, and the multigrade literature more generally, have been unable to adequately address issues of selection, which may also contribute to differential effects across studies and potentially bias findings in support for the multigrade philosophy (see Mason & Burns, 1996). In other words, children are not randomly assigned to

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

multi- and/or single-grade classrooms, but rather, parents select whether to send their children to these schools, which may also affect the child outcomes in question. Moreover, other economic and geographical reasons might also be at play (e.g., there are no other alternative programs available). For example, more experienced teachers are more likely to teach multigrade classrooms, and these classrooms generally serve more advantaged children (Mason & Burns, 1996); therefore, any observed benefits of multigrade classrooms might be due to the socioeconomic and educational advantages of parents and teachers rather than children's educational experiences in multigrade classrooms. Thus, despite the fact that multigrade education dates back well over a century, the research on this educational practice has remained in its early stages.

There are a few recent exceptions, however. In an analysis of the Early Childhood Longitudinal Study Kindergarten Class of 1998 (ECLS-K: 1998) cohort, Thomas (2012) found no differences in the educational outcomes of first graders who experienced multigrade (combination of kindergarten and first grade or first and second grades) versus single-grade education. Unlike much of the existing literature, this study was able to account for a rich set of child-, family-, and school-level characteristics as a means of ruling out confounding factors. In contrast to these national data, in a quasi-experimental analysis of the Los Angeles School District, Mariano and Kirby (2009) found that multigrade classrooms had a negative (roughly 10% of a standard deviation) effect on student achievement across the elementary school years, even when examining the average treatment effect on the treated through propensity score techniques. Similarly, analyses of elementary schools across the state of California have revealed that the implementation of multigrade classrooms for children in the second and third grade years resulted in lower test scores, with effects corresponding to 6–10% of a standard deviation.

Of note is that each of these studies has generally focused on children of older ages (first through fifth grades), which is critical as multigrade and mixed-age classrooms may be particularly harmful for younger children. For example, recent analyses of the Head Start program revealed that 4-year-old children who were enrolled in mixed-age classroom performed less well on assessments of math and literacy than 4-year-olds who attended classrooms with fewer younger children (Ansari, Purtell, & Gershoff, 2016). Similar patterns have emerged in other evaluations of preschool programs from urban cities in the northeastern United States (Moller, Forbes-Jones, & Hightower, 2008). Consequently, the costs and benefits of modern-day multigrade and mixed-age classrooms are unclear, especially during the transition to school, which has long-term implications for children's life success (Duncan et al., 2007).

1.2 | Mechanisms underlying the differences across multi- and single-grade classrooms

The mechanisms through which multigrade classrooms operate are also relatively unknown, with one of the primary hypothesized pathways being peer learning. Social-learning theory suggests that, during these early years, learning often occurs through modeling, whereby younger children can observe older and more competent children in the classroom and mimic their behaviors and actions (Bandura, 1986). Additionally, older children can scaffold younger children who, in turn, develop more independence and build on their kindergarten skills by reinforcing each other's abilities, and as part of this scaffolding process, older children can cement their own skills and knowledge (Vygotsky, 1978). These principles of cognitive and social-learning theories are the bedrock for the multigrade educational philosophy and underlie the peer effects literature (Justice, Logan, Lin, & Kaderavek, 2014; Mashburn, Justice, Downer, & Pianta, 2009), which demonstrates that classmates can (and do) affect children's learning.

Advocates of multigrade education also argue that these classrooms can promote the school success of children by reducing the rigid expectations that result from single-grade classrooms. In other words, children can learn at their own paces and skill levels (Katz et al., 1990), in part because teachers focus their instruction and daily activities on children's individual needs as opposed to the grade-level expectations. In an ideal classroom, even the oldest children should receive the appropriate degree of novel instructional stimulation as compared with children in single-grade classrooms. Thus, multigrade classrooms can provide teachers with a unique means of meeting the developmental needs of children of varying ages and abilities; whether these theories work in practice, especially for older and more

advanced children within a classroom, has remained contested (Ansari et al., 2016; Bell, Greenfield, & Bulotsky-Shearer, 2013; Justice et al., 2014; Moller et al., 2008).

In contrast, opponents of multigrade education argue that these classrooms may be less conducive for children's learning and development for a number of reasons (Mason & Burns, 1996). For example, there is a growing body of literature suggesting that children's own abilities shape and influence classroom standards and processes (Nurmi & Kiuru, 2015) so that, in practice, multigrade teachers often end up providing less individualized attention and less challenging content to older children in the classroom (Mason & Burns, 1996). Teachers in multigrade classrooms may also have a more challenging time managing children who display a wide range of academic and behavioral skills. Consequently, these teachers may have less time to dedicate to whole group and small group instruction and, instead, may spend a greater amount of time on classroom management and discipline, which may slow children's academic growth (Mason & Burns, 1996). Regardless of whether these pathways are direct or indirect, or whether these effects are positive or negative, these classrooms present a unique opportunity to answer a basic developmental question that has received conflicting empirical support: whether the eldest children make greater academic gains when they are with same-or different-aged peers.

1.3 | The current study

This study attempts to bridge these gaps in knowledge by using data from the nationally representative ECLS-K class of 2010–2011 to address three research questions: (a) What is the current prevalence of multigrade and kindergarten-only classrooms and what characteristics distinguish them? (b) How do the educational outcomes of 5-year-old children in multigrade classrooms compare with those of 5-year-olds in kindergarten-only classrooms? (c) And what are the mechanisms through which multigrade classrooms affect 5-year-olds' school success? Given the mixed evidenced regarding the benefits of multigrade classrooms, I did not make directional hypotheses. In addressing these questions, however, this study considers 5-year-olds' academic outcomes as well as their executive function and socioemotional development, both of which shape children's short- and long-term school success (Masten et al., 2012; McClelland, Acock, Piccinin, Rhea, & Stallings, 2013). Finally, this study also considers classroom organization, management and discipline, instructional delivery, and self-directed and peer learning as possible mechanisms for differential effects across multigrade and kindergarten-only classrooms.

In sum, this study is poised to fill a critical gap in the existing literature that has provided mixed empirical support for the effectiveness and efficacy of multigrade education during the early elementary school years. This investigation builds on the work of Thomas (2012) by shifting the attention from the implementation of multigrade education during the first grade year to kindergarten, and using the most recent cohort of the ECLS-K. Thus, the findings reported herein have potential implications for both developmental theory and educational policy during a period of heightened accountability: On the one hand, if contemporary multigrade kindergarten classrooms have a deleterious effect for children's educational outcomes, then programs that use these models need to be re-evaluated; on the other, if these programs foster children's school success, then multigrade classrooms may present a unique opportunity to educate young children during the transition to formal schooling. Finally, not only can this study delve into the pros and cons of multigrade education, but it can also help determine why these educational programs influence children's developmental outcomes.

2 | METHOD

This study used a nationally representative sample of kindergarteners drawn from the ECLS-K class of 2010–2011 (see Tourangeau et al., 2014). The ECLS-K included approximately 18,760 children in both part- and full-day kindergarten programs across 970 schools in the United States. The first stage of sampling was based on 90 primary sampling units, which were selected from a national sample that represented different counties and geographic areas.

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/rcd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cad.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

At the second stage, 1,310 schools (80% public) were selected within the primary sampling units, with 970 eventually participating. Finally, in the third stage, up to 23 kindergarteners were selected per school. It is important to note that the ECLS-K sampling frame was based on kindergarten enrollment, not child age; thus, the results of this study reflect families with kindergarteners during the 2010–2011 school year (for more information on sampling, see Tourangeau et al., 2014). To date, these children have been followed up through the end of the fourth grade year, and data were collected from multiple informants including direct child assessments as well as parent, caregiver, and teacher reports of the home, community, and school context.

For the purposes of the current study, children were required to have (a) a valid longitudinal weight, ensuring that the data are nationally representative of the larger population of children and families (and correcting for nonrandom attrition); (b) been enrolled in a classroom that served only kindergarteners or served both prekindergarteners and kindergarteners (more details provided below); and (c) been enrolled in a full-day kindergarten classroom. The first exclusion criterion was in place because children without a longitudinal weight were those who left the study sample by the end of kindergarten and, thus, did not experience a full year in multi- or single-grade classrooms. The second exclusion criterion meant that teachers who reported that they taught a classroom that included any grade beyond prekindergarten or kindergarten were excluded from these analyses. These children were excluded because the purpose of the study was to examine the effects of being a 5-year-old in a classroom with prekindergarteners. Finally, the last exclusion factor was because a large number of multigrade kindergarten classrooms were full-day programs. These three exclusion factors resulted in a final sample of 11,000 children (per IES/NCES guidelines, all sample sizes have been rounded to the nearest 10). See Table 1 for sample descriptives.

2.1 | Measures

Weighted descriptive statistics for the focal variables are presented in Tables 1 and 2, separated by type of kindergarten class (i.e., kindergarten only vs. multigrade classrooms). All reported reliabilities are from the ECLS-K: 2011 user's manual (Tourangeau et al., 2014).

2.1.1 | Multigrade classrooms

During the start of the school year, teachers reported whether they taught a multigrade classroom and, if so, what grade levels were included. Options included prekindergarten, transitional kindergarten, regular kindergarten, first grade, second grade, and third grade or higher. Teachers who said that they did not teach a multigrade classroom and/or reported that their class only served kindergartens were classified as teaching a kindergarten-only classroom. In contrast, teachers who answered "yes" to teaching a multigrade classroom and reported that one of the grade levels included prekindergarten and a regular kindergarten were classified as multigrade. Children were also classified as being enrolled in a multigrade classroom if teachers reported that they did not teach a multigrade class (or they did not respond to this question) but noted that their class served both prekindergarteners and kindergarteners. Thus, multigrade kindergarten classrooms were those that served both prekindergarteners and kindergartners, whereas kindergarten-only classrooms were those that served strictly kindergarteners.

2.1.2 | Academic achievement

Children's academic achievement was assessed at the beginning and end of the school year using two measures developed specifically for the ECLS-K and were based on national and state standards. The reading assessment (T1 and T2 αs = .95) measured children's print familiarity, letter recognition, decoding, and vocabulary knowledge, whereas the math assessment (T1 α = .92 and T2 α = .94) captured children's problem solving, geometry and spatial sense, and measurement skills. For the purposes of this study, the item response theory scores were used, which allows for the estimation of growth over time. These scores were generated on the basis of the pattern of right and wrong responses to the questions administered in an assessment and adjusted for the possibility of children guessing several difficult items correctly.

TABLE 1 Sample descriptives across kindergarten-only and multigrade classrooms

	Kindergarten only	Multigrade	Significant group difference
Child/household characteristics			
Age at kindergarten entry	66.26 (4.64)	63.79 (5.70)	***
Child male	0.52	0.57	
Child White	0.51	0.54	
Child Black	0.16	0.14	
Child Latino	0.24	0.20	
Child Asian/other	0.09	0.13	†
English home language	0.84	0.93	***
First time kindergartener	0.94	0.92	
Child diagnosed with disability	0.21	0.23	
Beginning of year skills			
Math	29.87 (10.71)	31.72 (10.05)	*
Reading	36.98 (9.19)	39.79 (11.15)	***
Working memory	432.00 (29.97)	437.22 (29.71)	*
Cognitive flexibility	14.24 (3.24)	14.18 (3.27)	
Attentional control	3.87 (0.89)	3.98 (0.78)	
Externalizing	1.63 (0.64)	1.62 (0.60)	
Internalizing	1.47 (0.50)	1.44 (0.49)	
Social skills	2.97 (0.64)	3.03 (0.59)	
Parent education	13.55 (2.61)	15.62 (2.40)	***
Parent age	33.51 (6.97)	36.51 (6.29)	***
Parents' married	0.64	0.76	***
Children in house	2.51 (1.15)	2.22 (0.96)	***
Mom employed full time	0.44	0.56	***
Mom employed part time	0.20	0.21	
Mom unemployed	0.37	0.24	***
Household income	9.57 (5.58)	13.56 (5.24)	***
School/class characteristics			
Class size	20.41 (4.95)	16.19 (9.35)	***
Teacher experience	14.19 (9.64)	16.37 (10.76)	***
Teacher years of education	16.95 (1.12)	15.92 (1.34)	***
Teacher change	0.07	0.08	
Public school	0.89	0.30	***
Percent minority	0.47	0.48	
Percent eligible for free/reduced lunch	0.48	0.17	***
School size	513.81 (236.13)	267.52 (238.53)	***
Northeast	0.13	0.19	*
Midwest	0.24	0.07	***
South	0.44	0.57	***
West	0.19	0.18	
City	0.33	0.45	***

(*Continues*)

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

	Kindergarten only	Multigrade	Significant group difference
Suburb	0.29	0.40	***
Town	0.12	0.03	***
Rural	0.26	0.12	***
End of kindergarten outcomes			
Math	43.12 (11.36)	43.21 (11.31)	
Reading	49.53 (11.33)	49.89 (13.25)	
Working memory	449.65 (30.21)	451.75 (28.59)	
Cognitive flexibility	15.18 (2.71)	14.93 (2.98)	
Attentional control	4.03 (0.89)	4.05 (0.79)	
Externalizing	1.66 (0.65)	1.78 (0.63)	*
Internalizing	1.53 (0.50)	1.57 (0.50)	
Social skills	3.12 (0.65)	3.14 (0.65)	
Child level sample size	10,800	200	
Classroom-level sample size	2,360	50	
School level sample size	680	30	
Population estimate	2,760,200	57,000	

Note. Proportions might not sum to 1.00 because of rounding.

***$p < .01$.

**$p < .01$.

*$p < .05$.

†$p < .10$.

2.1.3 | Executive function

Two facets of children's executive function were directly assessed: (a) cognitive flexibility, which was assessed using the Dimensional Change Card Sort (DCCS; Zelazo, 2006), and (b) working memory, which was assessed using the Numbers Reversed subtest of the Woodcock-Johnson (NR-WJ; Mather & Woodcock, 2001). During the DCCS assessment, children were asked to sort cards into different trays on the basis of rules (e.g., by shape or by color) that changed periodically. Children's scores on the DCCS were based on a combination of accuracy and reaction time. In contrast, during the NR-WJ, children were asked to repeat a series of numbers that were dictated to the child, backwards. If children responded incorrectly to a series of numbers, then the task ended; if, however, children respond correctly, then the number span increased by one digit at a time. For the purposes of this study, the W score was used for the NR-WJ assessment and the combined score was used for the DCCS. The W scores were generated using norming data provided by the publishers, which are a function of a child's raw score, age, and the language of assessment (Spanish or English). It should be noted that unlike the item response theory scores for math and reading, the W scores only account for the total number of administered sequences answered correctly and do not reflect the pattern of response.

2.1.4 | Social–emotional skills

During the fall and spring of the kindergarten year, teachers reported on four dimensions of children's social–emotional development using subscales from the Social Skills Rating System (Gresham & Elliott, 1990): attentional control, internalizing and externalizing behavior problems, and social skills. The Social Skills Rating System is based on a 4-point Likert scale (0 = *never* to 3 = *very often*) and has demonstrated adequate reliability in the ECLS-K (αs across waves and subscales = .78–.89).

TABLE 2 Weighted bivariate and multivariate differences in classroom-level processes

	Bivariate differences, *M* (*SD*)					Multivariate differences[a]		
	Kindergarten only	Multigrade	Diff.	ES		*b* [95% CI]	ES	
Classroom age composition								
% 3 years of age	0.00	0.26	0.26	4.72	***	—	—	—
% 4 years of age	0.03	0.43	0.40	3.77	***	—	—	—
% 5 years of age	0.82	0.28	−0.54	−3.03	***	—	—	—
% 6 years of age or greater	0.15	0.02	−0.13	−0.84	***	—	—	—
Hours per week spent in …								
Whole group activities	11.32 (4.73)	6.60 (4.88)	−4.72	−0.99	***	−4.91 [−6.62, −3.19]	−1.03	***
Small group activities	7.53 (4.02)	6.76 (4.72)	−0.77	−0.19	**	−0.07 [−1.47, 1.33]	−0.02	
Individual activities	4.65 (2.98)	7.52 (5.81)	2.87	0.93	***	2.87 [0.81, 4.93]	0.93	**
Child-selected activities	4.09 (2.53)	10.21 (6.04)	6.11	2.19	***	5.60 [3.52, 7.68]	2.01	***
Language arts activities	10.25 (3.74)	7.32 (4.16)	−2.93	−0.77	***	−1.30 [−2.67, 0.06]	−0.34	†
Math activities	6.35 (2.74)	5.90 (3.70)	−0.44	−0.16	*	0.12 [−1.08, 1.31]	0.04	
Dealing with misbehavior	3.83 (3.32)	4.04 (3.27)	0.21	0.06		0.47 [−0.70, 1.64]	0.14	
Overall class behavior	3.41 (0.88)	3.46 (0.75)	0.05	0.06		−0.03 [−0.30, 0.24]	−0.03	
% of children below grade level	0.17	0.36	0.19	0.76	***	0.20 [−0.01, 0.40]	0.80	†

Note. ES, effect size in standard deviation units.

[a]The multivariate models included clustered standard errors at the school level and adjusted for the baseline factors listed in Table 1.

***$p < .01$.

**$p < .01$.

*$p < .05$.

†$p < .10$.

2.1.5 | Classroom mediators

During the spring of the kindergarten year, teachers reported (a) how many days per week they taught reading/language arts and mathematics (0 = *never* to 7 = *5 days week*) and (b) on the days they taught these subjects, how many hours children spent working on lessons or projects within these subject areas (0 = *never* to 8 = *3 hr or more*). Teachers' responses to these two questions were cross-classified to generate an estimate for the hours per week children spent working on math and literacy activities. Teachers also reported how much time each day children in their classroom spent in teacher-directed whole class activities, teacher-directed small group activities, teacher-directed individual activities, and child-selected activities with responses ranging from 0 (*never*) to 6 (*4 hr or more*). Because teachers did not report how many days per week children spent in the aforementioned activities, it was assumed that these activities were done each day so that all classroom-level variables would be on the same metric (i.e., hours per week). Thus, teachers' responses were multiplied by five to generate an estimate for approximately how long children spent in whole class, small group, child-selected, and individualized activities each week. Teachers also reported how much time each day was spent on classroom discipline or handling disruptive behavior. Responses ranged from 0 (*less than 30 min a day*) to 6 (*3 hr or more a day*). To translate these estimates to hours per week, these responses were also multiplied by 5 days. Next, during the end of the school year, teachers reported how many children in their classroom (including those not sampled in the ECLS-K) were below grade level in reading and math. These reports were divided by the number of children in the classroom to generate an estimate for the percent of children in the classroom scoring below grade level. Owing to high overlap between the percent of children who scored below grade level in math and reading, the average of the two was taken. Finally, teachers rated the classroom-level behavior on a scale of 1 (*group misbehaves frequently*) to 5 (*group behaves exceptionally well*).

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.1.6 | Covariates

All models adjusted for a full set of child-, family-, household-, and school-level covariates that were drawn from the fall wave of data collection (see Table 1): child age, child gender, child race/ethnicity, home language, whether child was a first-time kindergartener, whether child was diagnosed with a disability, parents' years of education, parents' age, whether parents were married, parents' employment status, household income, number of children in the household, region, urbanicity, class size, teachers' years of experience, teachers' years of education, change in teachers across the school year, school sector, percent of minority children enrolled at the child's school, percent of children eligible for free/reduced lunch, and school size.

2.2 | Analytic strategy

All analyses were estimated in Stata (version 14; StataCorp, 2009) and Mplus (version 7; Muthén & Muthén, 1998–2013). Stata was primarily used to address missing data and generate sample descriptives and for propensity score matching (discussed in more depth below), whereas Mplus was used for the focal multivariate analyses. Missing data were minimal (mean = 5%, range = 0–20%) and did not systematically vary as a function of classroom type. The highest levels of missing data were for indicators of children's disability status, household income, and some of the school/classroom characteristics (e.g., percent minority and percent of children below grade level in academics). In order to maximize the sample size, we addressed missing data through 50 imputed datasets using the chained equations method in Stata (Azur, Stuart, Frangakis, & Leaf, 2011). The multilevel nature of data was accounted for by using clustered standard errors at the school level (see also Weiland & Yoshikawa, 2014), and all models included the longitudinal kindergarten weight to account for stratification and cross-wave attrition. For each outcome, two sets of models were estimated in Mplus. In Model 1, the outcomes of interest were individually regressed on the covariates and an indicator of whether children were enrolled in a kindergarten only (0) or a multigrade classroom (1). Then, in Model 2, the classroom-level mediators were included to determine how much of the aforementioned associations were due to children's experiences in the different types of classrooms. The indirect command was used to test for mediation in Mplus. When predicting child outcomes, all focal variables of interest were standardized to have a mean of 0 and standard deviation of 1; thus, all reported covariate-adjusted estimates of children's kindergarten performance correspond to effect sizes in terms of standard deviational units (adjusting for covariates).

Any evaluation of multigrade education must acknowledge selection bias (see Mason & Burns, 1996); in other words, children are not randomly assigned to multi- and/or single-grade classrooms, but rather, parents select whether to send their children to these neighborhoods, schools, or classrooms, and thus factors that influence parents to choose one over the other may also affect the child outcomes in question. To address selection bias, all analyses accounted for children's incoming skills and behaviors (i.e., lagged dependent variable models), which is one of the strongest adjustments for selection bias (National Institute of Child Health and Human Development Early Child Care and Youth Development & Duncan, 2003). In addition, propensity score matching methods (Rosenbaum & Rubin, 1983) were used to minimize selection bias, whereby children were matched across groups so that they displayed equal or similar characteristics at school entry (methodology discussed at greater length below).

3 | RESULTS

3.1 | Prevalence and characteristics of multigrade kindergarten classrooms

The first aim of this investigation was to document the prevalence and characteristics of multigrade kindergarten programs that combined prekindergarten and kindergarten classrooms versus kindergarten-only classrooms. Estimates from ECLS-K: 2010–2011 cohort revealed that 200 children were enrolled in full-day multigrade

classrooms ($n_{classroom}$ = 50; n_{school} = 30), which when weighted to be nationally representative extrapolates to a little over 57,000 five-year-old children across the nation (population estimates are generated on the basis of the ECLS-K weights). In contrast, roughly 2.76 million children attended full-day kindergarten-only classrooms during the 2010–2011 school year ($n_{children}$ = 10,800; $n_{classrooms}$ = 2,360; $n_{schools}$ = 680); therefore, although several thousand children were enrolled in classrooms that served both prekindergarteners and kindergartners across the country, these multigrade classrooms only served 2% of the 5-year-old population. Descriptive analyses also revealed that 5-year-old children were more likely to attend multigrade classrooms when they lived in larger urban communities and when they lived in the South (the significance of the group differences were estimated within a regression framework). Moreover, parents who enrolled their children in multigrade classrooms were, on average, a few years older, more likely to be married, and more advantaged across indicators of socioeconomic status (e.g., education, employment, and income). For other sociodemographic factors related to families' selection into multigrade versus kindergarten-only classrooms, see Table 1.

Not surprisingly, and as can be seen in Table 2, there were substantive differences across the two types of classrooms in terms of the average classroom age composition. Approximately 70% of children in multigrade kindergarten classrooms that served both prekindergartners and kindergarteners were 4 years of age or younger, whereas 97% of children in kindergarten-only classrooms were 5 years of age or older.

Stark differences also emerged when looking at the descriptive differences in the classroom-level processes (see left-hand panel of Table 2). On average, 5-year-olds in multigrade classrooms spent less time in whole group (5 hr less) and small group activities (1 hr less). In contrast, these children spent roughly 3 hr more per week in individual activities and over 6 hr more in child-selected activities. Five-year-olds in multigrade classrooms also experienced less language and literacy (3 hr less) and math activities (0.5 hr less) than their same-aged counterparts in kindergarten-only classrooms. Finally, a greater number of children's classmates in multigrade classrooms were considered to be below grade level in academics at year's end (36% vs. 17%). A number of these differences persisted when adjusting for other factors in the multivariate models (see right-hand panel of Table 2).

3.2 | Educational outcomes of children in multigrade classrooms

Turning to research question 2, I next examined the associations between classroom type and children's educational outcomes (see Model 1 in Table 3). Results from these models revealed that children who attended multigrade classrooms that combined pre-K and kindergarten demonstrated fewer gains in math and literacy skills throughout the school year, with effect sizes corresponding to 15% and 19% of a standard deviation, respectively. These disparities across classrooms were the equivalent of roughly 4 to 6.5 months of academic development (calculated by dividing the standardized difference in academic test scores by the regression slope of children's age; see Bradbury, Corak, Waldfogel, & Washbrook, 2011). Disparities in cognitive flexibility also emerged across classrooms, with children in multigrade classrooms demonstrating fewer gains in cognitive flexibility (18% of a standard deviation) throughout the kindergarten year. Similar, albeit marginal, patterns emerged for children's working memory (10% of a standard deviation, p = .053) and externalizing behavior problems (12% of a standard deviation, p = .050), but no significant differences were documented for children's internalizing problems or social skills. When taken together, these results indicate that children in multigrade classrooms that combined both pre-K and kindergarten exhibited less optimal academic achievement, executive function, and externalizing behavior than children in kindergarten-only classrooms.

3.3 | Mediating mechanisms of multigrade kindergarten education

Having established the disparities in children's early learning and development as a function of classroom type, the next goal of this investigation was to examine the processes underlying these differences. Accordingly, the classroom-level factors were incorporated into each of the models to test for mediation. As can be seen in the Model 2 of Table 3, the negative associations between multigrade classrooms and children's academic skills were attenuated by

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cad.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Results from unmatched and matched models predicting child outcomes as a function of multigrade versus kindergarten-only classrooms

	Reading	Math	Working memory	Cognitive flexibility	Externalizing	Internalizing	Attentional control	Social skills
Model 1: No mediators, unmatched sample								
Multigrade (kindergarten only)	−0.19** [−0.31, −0.07]	−0.15* [−0.27, −0.04]	−0.10† [−0.20, 0.00]	−0.18** [−0.32, −0.05]	0.12† [0.00, 0.23]	0.16 [−0.03, 0.36]	0.01 [−0.09, 0.11]	0.04 [−0.12, 0.20]
Model 2: With mediators, unmatched sample								
Multigrade (kindergarten only)	−0.09 [−0.24, 0.06]	−0.08 [−0.21, 0.05]	−0.06 [−0.16, 0.04]	−0.21** [−0.35, −0.06]	0.10 [−0.02, 0.23]	0.11 [−0.09, 0.30]	0.03 [−0.06, 0.12]	0.06 [−0.10, 0.23]
Indirect effect via …								
Whole group activities	0.00 [−0.02, 0.02]	0.01 [−0.01, 0.03]	0.00 [−0.02, 0.02]	0.02 [−0.01, 0.05]	−0.01 [−0.02, 0.01]	0.01 [−0.02, 0.04]	0.01 [−0.01, 0.03]	0.01 [−0.02, 0.03]
Small group activities	0.00 [−0.00, 0.00]	0.00 [−0.00, 0.00]	0.00 [−0.01, 0.01]	0.00 [−0.00, 0.00]	0.00 [−0.00, 0.00]	0.00 [−0.01, 0.01]	0.00 [−0.00, 0.00]	0.00 [−0.01, 0.01]
Individual activities	−0.02 [−0.04, 0.00]	−0.01 [−0.03, 0.01]	0.00 [0.02, 0.02]	0.00 [−0.02, 0.03]	0.01 [−0.01, 0.02]	0.02 [−0.01, 0.04]	−0.01 [−0.03, 0.01]	−0.02 [−0.05, 0.01]
Child-selected activities	−0.05* [−0.10, −0.01]	−0.04* [−0.08, −0.01]	−0.02 [−0.06, 0.02]	0.03 [−0.01 0.08]	0.01 [−0.02, 0.05]	0.02 [−0.03, 0.07]	−0.02 [−0.05, 0.02]	0.01 [−0.03, 0.05]
Language arts activities	−0.01 [−0.02, 0.00]	−0.00 [−0.01, 0.01]	−0.01 [−0.02, 0.01]	−0.01 [−0.02, 0.01]	0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	−0.01 [−0.03, 0.00]
Math activities	0.00 [−0.00, 0.00]	0.00 [−0.01, 0.01]	0.00 [−0.01, 0.01]	0.00 [−0.00, 0.00]	0.00 [−0.01, 0.01]	0.00 [−0.01, 0.01]	0.00 [−0.01, 0.01]	0.00 [−0.00, 0.00]
Dealing with misbehavior	−0.00 [−0.01, 0.00]	−0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	0.00 [−0.01, 0.01]	0.00 [−0.01, 0.01]	0.01 [−0.01, 0.02]
Overall class behavior	−0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.01]	−0.00 [−0.01, 0.00]	0.00 [−0.01, 0.00]	0.00 [−0.02, 0.03]	0.00 [−0.01, 0.01]	−0.00 [−0.02, 0.02]	−0.00 [−0.02, 0.02]
% of children below grade level	−0.02 [−0.04, 0.01]	−0.02 [−0.04, 0.01]	−0.01 [−0.03, 0.01]	−0.02 [−0.05, 0.01]	−0.00 [−0.01, 0.01]	0.01 [−0.01, 0.03]	−0.01 [−0.03, 0.01]	−0.01 [−0.03, 0.02]
Total indirect effect	−0.10*** [−0.16, −0.04]	−0.07** [−0.12, −0.02]	−0.04 [−0.09, 0.01]	0.03 [−0.03, 0.08]	0.01 [−0.04, 0.07]	0.06† [−0.00, 0.12]	−0.03 [−0.08, 0.03]	−0.02 [−0.08, 0.04]

(*Continues*)

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 (Continued)

	Reading	Math	Working memory	Cognitive flexibility	Externalizing	Internalizing	Attentional control	Social skills
Model 3: No mediators, matched sample								
Multigrade (kindergarten only)	−0.16** [−0.28, −0.05]	−0.15** [−0.26, −0.04]	−0.08 [−0.21, 0.04]	−0.15* [−0.30, −0.01]	0.12† [−0.02, 0.27]	0.13 [−0.07, 0.33]	−0.03 [−0.15, 0.010]	−0.03 [−0.19, 0.13]
Model 4: With mediators, matched sample								
Multigrade (kindergarten only)	−0.11 [−0.28, 0.05]	−0.05 [−0.20, 0.10]	0.02 [−0.16, 0.20]	−0.07 [−0.24, 0.11]	0.17† [−0.01, 0.34]	0.10 [−0.13, 0.33]	−0.07 [−0.26, 0.11]	−0.05 [−0.27, 0.17]

Note. All outcomes and mediators have been standardized to have a mean of 0 and standard deviation of 1, and thus, all estimates in this table correspond to effect sizes (i.e., standard deviation units). Estimates in brackets correspond to 95% confidence intervals.

***$p < .01$.

**$p < .01$.

*$p < .05$.

†$p < .10$.

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.2036 by UniversiteitsbibliotheekGent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

approximately half and were no longer statistically significant. Estimates of indirect effects revealed that this attenuation was statistically significant; that is, classroom-level processes accounted for a significant share of the difference in children's test scores across multigrade and kindergarten-only classrooms, with effect sizes of 7–10% of a standard deviation. In contrast to children's academic achievement, the disparities in the other domains of school performance remained largely the same (see Table 3).

When looking at the specific classroom-level processes that accounted for this significant attenuation, results indicate that over 25% of the attenuation in the association between kindergarten type and children's math and literacy skills was attributed to the time children spent in child-selected activities (see Table 3). Although not statistically significant (*p*s = .15–.18), roughly 10–15% of these differences in test scores were also attributed to the fact that children in multigrade classrooms that enrolled both prekindergartners and kindergartners had a greater number of classmates who were below grade level in academics. In contrast to children's academic performance, none of the mediators explained a significant share of the differences in children's cognitive flexibility or any of the other outcomes of interest.

3.4 | Robustness checks

Considering that multigrade classrooms that combined both pre-K and kindergarten were more likely to be housed in private schools and employed less educated teachers (see Table 1), ancillary analyses were estimated to determine whether the aforementioned associations were conditioned on these two factors using interaction terms. Results from these analyses (available from the author) revealed that the disparities in children's school success across multigrade and kindergarten-only classrooms were largely the same, regardless of school sector (public versus private) and teachers' years of education (continuously measured). In light of the evidence that the distribution of children's ages also has implications for child development (Ansari et al., 2016; Moller et al., 2008) and considering that the majority of kindergarten-only classrooms had at least one 6-year-old child—who were roughly a full year older than the children sampled in the ECLS-K—supplementary analyses were estimated controlling for the proportion of 6-year-olds in each class. All results were quantitatively similar as those discussed above.

Finally, to address concerns regarding selection bias, propensity scores were estimated using the psmatch2 function in the Stata program. Within each of the 50 imputed datasets, logit models—which included the entire set of variables listed in Table 1—were estimated to gauge the likelihood that children were enrolled in a multigrade classroom (vs. kindergarten only). The nearest neighbor method (with four matches) was used within a caliper of .05, which resulted in match of roughly 180 children in multigrade classrooms with approximately 480 children in kindergarten-only classrooms (sample sizes varied across imputations). The replacement method was used in order to maximize the multigrade kindergarten classrooms, while a caliper width of .05 was implemented because this method has been shown to produce more comparable comparison conditions (see Austin, 2011). Before propensity score matching, roughly 60% of the baseline characteristics were not balanced across the two conditions; as can be seen in Table A1, however, after employing propensity scores, there were no longer any significant differences across groups and the standardized differences were generally less than 10% of a standard deviation. As a further precaution, the overall balance of the matches was tested using a Hotelling test and the balance of the matched samples were checked within each quartile of the propensity scores, both of which also suggested that balance was largely achieved.

Having successfully achieved balance, the above models were replicated using the matched data in the Mplus program to test for the main effects of multigrade education on children's school outcomes. In order to adjust for any potential remaining bias from measured characteristics, these models included all covariates (see also Berger, Brooks-Gunn, Paxson, & Waldfogel, 2008; Coley & Lombardi, 2013). As can be seen in the Model 3 of Table 3, results from these analyses revealed much the same pattern: Children who attended classrooms that combined pre-K and kindergarten performed less well in areas of literacy, mathematics, and cognitive flexibility than children who attended kindergarten-only classrooms, with effect sizes ranging from 15% to 16% of a standard deviation. Although not statistically significant, the effect sizes for working memory and externalizing behavior remained comparable with the

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

unmatched estimates. The only substantive difference that emerged across the unmatched and matched models was that after accounting for the classroom factors (see Model 4 of Table 3), there were no longer any differences in children's cognitive flexibility when estimating propensity score models (effect size = 7% of a SD, ns), which was not the case when estimating models in the unmatched data (effect size = 21% of a SD, $p < .01$).

4 | DISCUSSION

Although single-grade classrooms are the dominant method of educating children in the United States, the multigrade model of education has long been an alternative philosophy that has been implemented across the country (e.g., Katz et al., 1990; Mason & Burns, 1996; Veenman, 1995); yet much of the empirical evidence behind this practice is fairly ambiguous. Given the increased number of children attending kindergarten (Davis & Bauman, 2013), we need to consider the implications of multigrade education when it is implemented during the transition to school, which has long-term implications for children's learning and development (Duncan et al., 2007; Entwisle & Alexander, 1989; Jones et al., 2015). Such inquiry is especially important in light of the "academization" of kindergarten (Bassok et al., 2016). Accordingly, the current investigation sought to contribute to this literature by conducting secondary data analysis of the ECLS-K class of 2010–2011 to answer three questions regarding: (a) the prevalence and characteristics of multigrade classrooms that combine pre-K and kindergarten classrooms; (b) the educational outcomes of kindergarteners in these classrooms; and (c) the mechanisms through which multigrade classrooms affect 5-year-olds' educational outcomes. The results of this investigation have two take-home messages.

To begin, only a small share of the overall kindergarten population in the United States attends classrooms that also serve prekindergarteners; however, at the population level, these estimates translate to roughly 57,000 children, or 2% of children, which is smaller than the share of children in multigrade classrooms at older ages (5%). In contrast to social-learning and cognitive theories of child development (Bandura, 1986; Vygotsky, 1978), the results from this investigation revealed that 5-year-old children who were enrolled in classrooms that served both prekindergarteners and kindergarteners fared less well academically than those who attended kindergarten-only classrooms. These differences in children's academic skills corresponded with 4 to 6.5 months of academic development. Children in multigrade classrooms also demonstrated less optimal executive functioning than did 5-year-olds who attended kindergarten-only classrooms, but similar to prior studies in this area of work (Ansari et al., 2016; Thomas, 2012; Veenman, 1995), children in multigrade classrooms were not consistently performing any better (or worse) than their classmates in areas of social behavior.

Although prior studies suggest that the effects of multigrade education are mixed (e.g., Mariano & Kirby, 2009; Thomas, 2012), much of this literature has sampled older children during first or second grade and beyond. This lack of empirical attention on the transition to school is notable as the skills that children demonstrate in kindergarten are predictive of their short- and long-term school success as well as their long-term well-being through the transition to adulthood (Duncan et al., 2007; Entwisle & Alexander, 1989; Jones et al., 2015). For these reasons, future studies should consider the relative age of children to determine whether the effects of multigrade classrooms change as children get older. In fact, recent analyses of preschool programs reveal much the same pattern as those reported here: Older children who were enrolled in classrooms with a greater number of younger children performed less well academically than older children in preschool classrooms with fewer younger children (Ansari et al., 2016; Moller et al., 2008). Therefore, the implementation of multigrade (and mixed-age) classrooms during early childhood and the transition to school requires greater scrutiny—just because children in multigrade classrooms do no better or worse than children in single-grade classrooms during first grade and beyond (e.g., Thomas, 2012; Veenman, 1995) does not mean that such practices produce no differences when implemented at kindergarten. When taken together, the results of this study are in line with the findings reported by Mariano and Kirby (2009), who found that multigrade classrooms were less conducive for children's academic achievement; however, the effect sizes reported herein were somewhat larger (10–20% of a standard deviation vs. 10% of a standard deviation). The effect sizes reported as part of this study

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15272719, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/rcd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

were also larger than those documented by Thomas (2012) who examined multigrade classrooms implemented in first grade (0–6% of a standard deviation).

Moving to the second take-home message, the reasons that multigrade classrooms appear to have a deleterious effect on children's educational outcomes have been understudied, but this study was able to identify meditational processes for their associations with kindergartener's development. Not only do the results from this investigation demonstrate that the experiences of children in these classrooms were qualitatively different—children in multigrade classrooms that served both prekindergartners and kindergartners experienced significantly less literacy stimulation and spent less time in teacher-directed activities—but also these classroom-level processes mediated the negative associations that were documented between multigrade classrooms and children's academic achievement. Thus, one of the underlying explanations why children in multigrade classrooms performed less well academically was because they received less teacher-directed instruction. It is important to interpret these mediational findings in the context of the existing literature on classroom pedagogy, which also suggest that more time spent in quality didactic teacher-directed instruction facilitates children's early academic achievement (Barnett, 2011; Mashburn et al., 2009; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008; Stipek, Feiler, Daniels, & Milburn, 1995). At the same time, however, more time spent in didactic instruction has been found to result in less optimal social-behavioral development among young children (Stipek et al., 1995).

Nonetheless, when taken as a whole, these meditational analyses indicate that there are, in fact, "indirect" pathways between multigrade education and children's early academic learning. From a practical perspective, these indirect pathways may prove to be malleable targets for intervention, which can be a first step in addressing the long-standing concerns of multigrade classrooms (Mason & Burns, 1996). These findings are especially important when considering that some multigrade programs—and by extension mixed-age programs—are implemented owing to economic and logistical constraints, and therefore, separately serving 4- and 5-year-old children is likely to be challenging for some kindergarten programs.

Despite these potential mechanisms that underlie the associations between multigrade classrooms and children's academic achievement, the mechanisms for children's executive function were unclear. Although this study could not uncover these processes, one might further speculate that classroom management and quality, which prior studies have found to be positively associated with children's executive function (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013), are two alternative pathways for the documented associations. Unfortunately, the ECLS-K data did not have indicators of classroom quality or specific information on what teachers were doing in the classroom during the different types of activities to test such hypotheses. Considering the fact that children's executive functioning rapidly evolves during this period, continued work is necessary to explore the pathways through which multigrade classrooms may affect children's development of executive function skills.

As with any study, the results of this study need to be interpreted in light of its limitations, beyond those discussed above. Primarily, the ECLS-K class of 2010–2011 included a nationally representative sample of kindergarteners—not prekindergarteners—and, thus, these data cannot (and do not) speak to the benefits of multigrade education for the preschool-age children who were enrolled in these classrooms. Whether younger children benefit from having older children in the classroom remains contested with some scholars documenting positive effects (Guo, Tompkins, Justice, & Petscher, 2014) and others documenting null or negative associations (Ansari et al., 2016; Bell et al., 2013; Moller et al., 2008; Thomas, 2012; Veenman, 1995) and, therefore, remains a topic for future research. It should be kept in mind, however, that even if the younger children were to benefit from having kindergarteners in the classroom, these data suggest that it would be at the expense of the older children's academic achievement and executive functioning.

Second, although several precautions were taken to address potential selection bias (e.g., lagged dependent variable analyses and propensity score matching), the results reported herein need to be interpreted with caution because children were not randomly assigned to multigrade and kindergarten-only classrooms. Additional analytical strategies (e.g., random assignment and instrumental variables) must be used to build on this correlational evidence

to test for causal associations; even so, it is encouraging that the outcomes associated with multigrade classrooms were largely the same across model specifications. Moreover, although the sample provided enough statistical power to detect meaningful differences, the small sample of children in multigrade classrooms limited the types of analyses that could be conducted. Future studies, therefore, should consider oversampling children in multigrade classrooms, which, in turn, will allow for a stronger assessment of the underlying mechanisms for these documented associations. In doing so, future studies can also consider potential moderating variables to determine under what context or conditions these effects are amplified (or attenuated). Relatedly, future studies should inspect of the long-term implications of multi- versus single-grade education and would greatly benefit from stronger and more detailed measures of classroom processes, which would allow for a better evaluation of the degree of difference across classrooms. Finally, closer attention must be paid to variation within multigrade classrooms in order to determine whether, and why, some multigrade programs are more successful than others. For example, understanding whether programs were implemented for pedagogical or logistical reasons is necessary because it is likely that pedagogical (as compared with logistical) implementation will account for distinct outcomes, whereas logistical implementation may indicate a lack of training.

In sum, this study provided a large-scale national portrait of multigrade education during the kindergarten year that can be built on by more intensive multimethod community-based approaches in the future. The results of this study also provided some of the first evidence to suggest that, at the national level, 5-year-olds who attended multigrade kindergarten classrooms with prekindergartners were performing less well academically than 5-year-olds who attended kindergarten-only classrooms. Classroom-level processes largely explained these differences. If future studies can address the limitations of this study and replicate the findings reported herein, then this research would have potential developmental and policy implications for children's learning during the transition to school. Specifically, with the small number of programs that combine prekindergarten and kindergarten classrooms in the United States, these results suggest that, when possible, their implementation may need to be re-evaluated.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Institute of Child Health and Human Development (R24 HD42849, PI Mark Hayward; T32 HD007081-35, PI R. Kelly Raley and Elizabeth Gershoff; and R01HD069564, PI Elizabeth Gershoff) awarded to the Population Research Center at the University of Texas at Austin and by a grant from the Institute of Education Sciences, U.S. Department of Education (R305B130013) awarded to the University of Virginia. I thank Elizabeth Gershoff and Kelly Purtell for their helpful comments on and assistance with prior versions of this article.

REFERENCES

Ansari, A., Purtell, K. M., & Gershoff, E. T. (2016). Classroom age composition and the school readiness of 3- and 4-year-olds in the Head Start program. *Psychological Science*, *27*, 53–63. https://doi.org/10.1177/0956797615610882

Austin, P.C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*, 399–424. https://doi.org/10.1080/00273171.2011.568786

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*, 40–49. https://doi.org/10.1002/mpr.329

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, *333*, 975–978. https://doi.org/10.1126/science.1204534

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

Bell, E. R., Greenfield, D. B., & Bulotsky-Shearer, R. J. (2013). Classroom age composition and rates of change in school readiness for children enrolled in Head Start. *Early Child Research Quarterly*, *28*, 1–10. https://doi.org/10.1016/j.ecresq.2012.06.002

Berger, L., Brooks-Gunn, J., Paxson, C., & Waldfogel, J. (2008). First-year maternal employment and child outcomes: Differences across racial and ethnic groups. *Children and Youth Services Review*, *30*, 365–387. https://doi.org/10.1016/j.childyouth.2007.10.010

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development*, *85*, 1346–1356. https://doi.org/10.1111/cdev.12192

Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2011). Inequality during the early years: Child outcomes and readiness to learn in Australia, Canada, United Kingdom, and United States. IZA Discussion Paper No. 6120.

Burns, R. B., & Mason, D. A. (2002). Class composition and student achievement in elementary schools. *American Educational Research Journal*, *39*, 207–233. https://doi.org/10.3102/00028312039001207

Coley, R. L., & Lombardi, C. M. (2013). Does maternal employment following childbirth support or inhibit low-income children's long-term development? *Child Development*, *84*, 178–197. https://doi.org/10.1111/j.1467-8624.2012.01840.x

Purl, M. C, & Curtis, J. (1970). *A look at combination class Effect at Emerson Elementary School.* Riverside, CA: Riverside Unified School District. (ERIC document reproduction service no. ED 042 863)

Davis, J., & Bauman, K. (2013). *School enrollment in the United States: 2011*. Washington, DC: U.S. Census Bureau.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Entwisle, D. R., & Alexander, K. L. (1989). Early schooling as a "critical period" phenomenon. *Sociology of Education and Socialization*, *8*, 27–55.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. American Guidance Service.

Guo, Y., Tompkins, V., Justice, L., & Petscher, Y. (2014). Classroom age composition and vocabulary development among at-risk preschoolers. *Early Education and Development*, *25*, 1016–1034. https://doi.org/10.1080/10409289.2014.893759

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social–emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*. https://doi.org/10.2105/AJPH.2015.302630

Justice, L. M., Logan, J. A., Lin, T. J., & Kaderavek, J. N. (2014). Peer effects in early childhood education testing the assumptions of special-education inclusion. *Psychological Science*, *25*, 1722–1729. https://doi.org/10.1177/0956797614538978

Katz, L. G., Evangelou, D., & Hartmann, J. A. (1990). *The case for mixed-age grouping in the early years*. Washington, DC: National Association for the Education of Young Children.

Little, A. W. (2001). Multigrade teaching: Towards an international research and policy agenda. *International Journal of Educational Development*, *21*, 481–497. https://doi.org/10.1016/S0738-0593(01)00011-6

Mariano, L. T. & Kirby, S. N. (2009). Achievement of students in multigrade classrooms: Evidence from the Los Angeles Unified School District. RAND Education Working Paper WR-685-IES.

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development*, *80*, 686–702. https://doi.org/10.1111/j.1467-8624.2009.01291.x

Mason, D. A., & Burns, R. B. (1996). "Simply no worse and simply no better" may simply be wrong: A critique of Veenman's conclusion about multigrade classes. *Review of Educational Research*, *66*, 307–322. https://doi.org/10.3102/00346543066003307

Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K., … Zelazo, P. D. (2012). Executive function skills and school success in young children experiencing homelessness. *Educational Researcher*, *41*, 375–384. https://doi.org/10.3102/0013189X12459883

Mather, N., & Woodcock, R. W. (2001). *Woodcock-Johnson III tests of achievement: Examiner's manual*. Riverside Pub.

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Child Research Quarterly*, *28*, 314–324. https://doi.org/10.1016/j.ecresq.2012.07.008

Moller, A. C., Forbes-Jones, E., & Hightower, A. D. (2008). Classroom age composition and developmental change in 70 urban preschool classrooms. *Journal of Educational Psychology*, *100*, 741–753. https://doi.org/10.1037/a0013099

National Institute of Child Health and Human Development Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*, 1454–1475. https://doi.org/10.1111/1467-8624.0061

Nurmi, J. E., & Kiuru, N. (2015). Students' evocative impact on teacher instruction and teacher–child relationships: Theoretical background and an overview of previous research. *International Journal of Behavioral Development*. https://doi.org/10.1177/0165025415592514

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, *45*, 365–397. https://doi.org/10.3102/0002831207308230

Pratt, D. (1986). On the merits of multiage classrooms. *Research in Rural Education*, *3*, 111–115.

Proehl, R. A., Douglas, S., Elias, D., Johnson, A. H., & Westsmith, W. (2013). A collaborative approach: Assessing the impact of multi-grade classrooms. *Catholic Education: A Journal of Inquiry and Practice*, *16*, 417–440.

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ccd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. K. J., Nathanson, L., & Brock, L. L. (2009). The contributuion of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*, 958–972 doi.org/10.1037/a0015861

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

StataCorp (2009). *Stata user's guide, release 11*. College Station, TX: Stata Press.

Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development*, *66*, 209–223. https://doi.org/10.1111/j.1467-8624.1995.tb00866.x

Thomas, J. L. (2012). Combination classes and educational achievement. *Economics of Education Review*, *31*, 1058–1066. https://doi.org/10.1016/j.econedurev.2012.07.013

Tourangeau, K., Nord, C., Lê, T., Sorongon, A.G., Hagedorn, M.C., Daly, P., and Najarian, M. (2014). Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLSK:2011), user's manual for the ECLS-K: 2011 Kindergarten Data Files and Electronic Codebook. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Urberg, K.A., & Kaplan, M.G. (1986). Effects of classroom age composition on the play and social behaviors of preschool children. *Journal of Applied Developmental Psychology*, *7*, 403–415. https://doi.org/10.1016/0193-3973(86)90009-2

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, *65*, 319–381. https://doi.org/10.3102/00346543065004319

Vincent, S. (1999). The multigrade classroom: A resource for small, rural schools. In *Review of the research on multigrade instruction*). Portland, OR: Northwest Regional Educational Laboratory.

Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.), *Readings on the development of children* (pp. 34–41). Cambridge, MA: Harvard University Press.

Way, J. W. (1981). Achievement and self-concept in multiage classrooms. *Educational Research Quarterly*, *6*, 69–75.

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and execution function skills in an urban public prekindergarten program. *Early Child Research Quarterly*, *28*, 199–209. https://doi.org/10.1016/j.ecresq.2012.12.002

Weiland, C., & Yoshikawa, H. (2014). Does higher peer socioeconomic status predict children's language and executive function skills gains in prekindergarten? *Journal of Applied Developmental Psychology*, *35*, 422–432. https://doi.org/10.1016/j.appdev.2014.07.001

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297–301. https://doi.org/10.1038/nprot.2006.4

How to cite this article: Ansari A. Multigrade kindergarten classrooms and children's academic achievement, executive function, and socioemotional development. *Inf Child Dev*. 2017;26:e2036. https://doi.org/10.1002/icd.2036

APPENDIX A

TABLE A1 Weighted sample descriptives after propensity score matching

	Kindergarten only	Multigrade	Significant group difference
Child/household characteristics			
Age at kindergarten entry	64.15 (5.53)	63.75 (5.67)	
Child male	0.58	0.58	
Child White	0.55	0.55	
Child Black	0.18	0.13	
Child Latino	0.15	0.18	
Child Asian/other	0.13	0.14	
English home language	0.93	0.93	

(*Continues*)

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/icd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A1 (Continued)

	Kindergarten only	Multigrade	Significant group difference
First-time kindergartener	0.92	0.91	
Child diagnosed with disability	0.23	0.24	
Beginning of year skills			
Math	31.27 (10.52)	30.98 (10.45)	
Reading	38.88 (10.20)	38.73 (11.11)	
Working memory	434.35 (29.81)	435.18 (29.08)	
Cognitive flexibility	13.76 (3.70)	13.97 (3.42)	
Attentional control	3.91 (0.90)	3.92 (0.78)	
Externalizing	1.64 (0.65)	1.62 (0.58)	
Internalizing	1.45 (0.50)	1.46 (0.50)	
Social skills	2.97 (0.66)	2.97 (0.60)	
Parent education	15.55 (2.36)	15.52 (2.39)	
Parent age	35.93 (6.44)	36.17 (6.30)	
Parents' married	0.74	0.74	
Children in house	2.32 (1.02)	2.28 (0.93)	
Mom employed full time	0.57	0.55	
Mom employed part time	0.17	0.18	
Mom unemployed	0.26	0.27	
Household income	13.44 (4.86)	13.59 (5.27)	
School/class characteristics			
Class size	16.57 (9.41)	16.96 (9.54)	
Teacher experience	17.29 (9.96)	17.25 (11.52)	
Teacher years of education	16.23 (0.99)	16.14 (1.12)	
Teacher change	0.07	0.10	
Public school	0.35	0.36	
Percent minority	0.45	0.44	
Percent eligible for free/reduced lunch	0.20	0.20	
School size	283.45 (210.11)	278.99 (233.24)	
Northeast	0.23	0.25	
Midwest	0.05	0.06	
South	0.56	0.51	
West	0.16	0.18	
City	0.39	0.34	
Suburb	0.39	0.43	
Town	0.02	0.04	
Rural	0.19	0.19	
Sample size[a]	~480	~180	

Note.

[a]Sample sizes vary across the imputations; the sample size provided corresponds to the average number of children across the 50 imputed datasets.

15227219, 2017, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cd.2036 by Universiteitsbibliotheek Gent, Wiley Online Library on [16/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License