Classroom Age Composition and the School Readiness of 3- and 4-Year-Olds in the Head Start Program

Arya Ansari[1], Kelly Purtell[2], and Elizabeth Gershoff[1]
[1]Department of Human Development and Family Sciences, University of Texas at Austin, and [2]Department of Human Sciences, The Ohio State University

Abstract

The federal Head Start program, designed to improve the school readiness of children from low-income families, often serves 3- and 4-year-olds in the same classrooms. Given the developmental differences between 3- and 4-year-olds, it is unknown whether educating them together in the same classrooms benefits one group, both, or neither. Using data from the Family and Child Experiences Survey 2009 cohort, this study used a peer-effects framework to examine the associations between mixed-age classrooms and the school readiness of a nationally representative sample of newly enrolled 3-year-olds ($n = 1,644$) and 4-year-olds ($n = 1,185$) in the Head Start program. Results revealed that 4-year-olds displayed fewer gains in academic skills during the preschool year when they were enrolled in classrooms with more 3-year-olds; effect sizes corresponded to 4 to 5 months of academic development. In contrast, classroom age composition was not consistently associated with 3-year-olds' school readiness.

The early childhood years present a critical window of opportunity to address issues of inequality in the life course (Heckman, 2008). To this end, there has been mounting interest in publicly funded preschool programs, including the expansion of preschool education for 3-year-olds (Duncan & Magnuson, 2013), as a potential policy lever for preparing children for school. Understanding the implications of such expansion efforts is of the utmost importance, as many programs, including Head Start—the largest federally funded preschool program in the United States—often serve both 3- and 4-year-olds in the same classrooms. In fact, as of the 2009 school year, roughly 75% of all Head Start classrooms were mixed age (Moiduddin et al., 2012).

The empirical inquiry into mixed-age classrooms is not a new endeavor; it has been central to many developmental theories in the field of early childhood education. Social-learning and cognitive theories of child development suggest that one of the mechanisms through which development occurs is interactions between children and their classmates (Bandura, 1986; Vygotsky, 1978). That is, children's active engagement with other children, especially those who are older and more skilled both academically and behaviorally, can facilitate their cognitive and socioemotional development over time. Furthermore, this model suggests that for older children, mixed-age classrooms allow them to practice and develop their own academic and behavioral skills through scaffolding and modeling behaviors for younger children. From a practical and theoretical standpoint, therefore, mixed-age classrooms can have advantages for all children, regardless of their age (Katz, Evangelou, & Hartman, 1990).

Corresponding Author:
Arya Ansari, University of Texas at Austin, Department of Human Development and Family Sciences, 1 University Station A2700, Austin, TX 78712
E-mail: aansari@utexas.edu

These cognitive and social-learning theories underlie the *peer-effects* framework (Henry & Rickman, 2007; Justice, Logan, Lin, & Kaderavek, 2014; Mashburn, Justice, Downer, & Pianta, 2009) that guides this study. In contrast to the evidence on the effects of teachers on children, however, the literature on peer effects—and how children's learning is directly affected by their classmates—is far shallower, especially during the preschool years. Understanding the role of peer effects during this period is critical because much of the classroom instruction in early care and education programs occurs through peer interactions and whole-group activities as opposed to teacher-directed instruction (Ansari & Gershoff, 2015; Mashburn et al., 2009). The omission of peer effects from past evaluations of Head Start has left a critical gap in knowledge. In fact, classroom age composition may be one of the main reasons why the Head Start program has been documented as less effective for older children than for younger children (Jenkins, Farkas, Duncan, Burchinal, & Vandell, 2015; Puma et al., 2010), although this possibility has yet to be tested.

Prior research has suggested two primary reasons that peer effects may be important in early childhood classrooms. One is the direct-effects pathway (Justice et al., 2014; Mashburn et al., 2009). In this pathway, children are directly influenced by their peer's abilities. From this perspective, one would hypothesize that mixed-age classrooms would be beneficial for younger children but may be deleterious for the skill development of older children. The second is an indirect-effects pathway, whereby teachers modify their classroom practices to accommodate a wide range of skill levels, which can lead to less challenging content and potential disengagement for older children (Urberg & Kaplan, 1986). This pathway is particularly salient when considering children's behavior, as having a wide range of behaviors in the classroom or a high concentration of problem behaviors is challenging for teachers to manage (Yudron, Jones, & Raver, 2014); therefore, teachers of mixed-age classrooms may spend more time on classroom management than on activities that facilitate children's learning. While both of these pathways for peer effects seem feasible, the evidence regarding mixed-age preschool classrooms has been ambiguous, with some studies documenting positive effects (Blasco, Bailey, & Burchinal, 1993; Goldman, 1981; Guo, Tompkins, Justice, & Petscher, 2014), and others documenting null or negative associations (Bell, Greenfield, & Bulotsky-Shearer, 2013; Moller, Forbes-Jones, & Hightower, 2008; Urberg & Kaplan, 1986; Winsler et al., 2002). Given the conflicting evidence, it remains unclear whether children benefit from being in mixed-age versus restricted-age classrooms and, if there is an advantage, whether this applies to younger or older children.

Recent studies that have tried to address these outstanding issues have been limited in a few notable ways. First, these studies have not examined these processes using a national data set, and, therefore, some of the conflicting evidence may be the result of prior studies having generally used small and localized samples (Blasco et al., 1993; Goldman, 1981; Guo et al., 2014; Winsler et al., 2002). Second, although two studies (Bell et al., 2013; Moller et al., 2008) provided much needed—albeit conflicting—evidence on the potential outcomes of mixed-age classrooms for low-income children, they relied on teachers' reports of children's academic skills. This is problematic because these reports are likely to be biased and reflect teachers' perceptions of children's abilities relative to those of their classmates (Guo et al., 2014; Mashburn et al., 2009), which may be particularly concerning when trying to tease apart the achievement of two age cohorts of children. Third, researchers have usually failed to control for other classroom processes and selection factors, thus limiting their ability to draw firm conclusions about class composition. Finally, although restricted-age classrooms may be more effective than mixed-age classrooms (Moller et al., 2008), it may not always be feasible to separately serve 3- and 4-year-olds in all areas of the country (Veenman, 1995); in other words, mixed-age classrooms are sometimes created out of necessity. Thus, an additional scientific and policy question that remains to be answered is whether there is a threshold or tipping point at which the ratio of younger to older students in mixed-age classrooms has a more deleterious or positive effect for children.

Our objective in the present investigation was to address these gaps in knowledge by using data from the Family and Child Experiences Survey (FACES) 2009 cohort to examine (a) whether classroom age composition affects 3- and 4-year-olds' school readiness and (b) whether there is a threshold at which the associations between classroom age composition and children's school readiness are stronger (or weaker). In addressing these questions, we are poised to answer both a basic developmental question regarding the role of peer effects in early care and education programs as well as a policy question regarding the implications of mixed-age classrooms for Head Start children across the nation.

Method

The FACES 2009 study followed a nationally representative sample of 3,349 three- and four-year-old first-time Head Start attendees across 486 classrooms. Children participated in the study in the Fall of 2009 and were followed-up periodically through the end of their kindergarten year (Spring 2010, Spring 2011, and for 3-year-olds, Spring 2012). To achieve a nationally representative

sample, FACES 2009 used a procedure in which the probability of being selected in the first three stages (program, center, and classroom) was proportional to size; in the fourth stage (children), equal-probability sampling was used. In total, there were 60 selected programs, two centers per program, and up to three classrooms per center. Roughly 10 children (both 3 and 4 years of age) were selected per class, with an oversampling of 3-year-olds. The sampling frame included Head Start programs in all 50 states and the District of Columbia (for more information on sampling, see Moiduddin et al., 2012). If a child left Head Start at any time before the end of the kindergarten year, he or she was no longer included in the FACES study.

For the current investigation, we focused on the first two waves of data collection (Fall 2009 and Spring 2010); thus, children who did not have a valid longitudinal weight required for our modeling techniques were excluded (*n* = 444). We also excluded 76 children who switched classrooms between the Fall and Spring of the school year in order to isolate any additional confounds that may emerge from switching classrooms. These exclusion criteria resulted in a final sample of 2,829 children (42% 4-year-olds, 58% 3-year-olds; see Table 1 for demographics). Compared with the 520 children who were excluded, the final sample of children displayed more positive social behavior (but not better academic performance) at the start of the school year and were enrolled in classrooms with more children overall but fewer 3-year-olds. Although there were no differences in family income by exclusion status, children in our final sample came from families who were less disadvantaged across other indicators compared with children who were excluded. For example, these children had mothers who were slightly older, displayed fewer depressive symptoms, were more likely to be employed, and had greater years of education. Additionally, these children were more likely to be of Latino origin and from Spanish-speaking homes. The use of longitudinal weights, however, which are discussed more fully below, addressed issues of nonrandom and cross-wave attrition.

Measures

Descriptive statistics for all variables of interest are presented in Table 1.

Classroom age composition. During the Fall of 2009, teachers reported how many children overall were in their classroom (*M* = 17.15, *SD* = 2.21) and how many were 3 years old or younger (*M* = 7.11, *SD* = 5.17), 4 years old (*M* = 9.04, *SD* = 5.66), and 5 years old (*M* = 1.00, *SD* = 1.81). Note that teachers did not report children's exact ages, and these reports were for all children in the

classroom, not just for the children who were part of the FACES study. Because there were only a small number of 5-year-olds, we dichotomized children as 3 years old or younger or 4 years or older (for a similar method, see Moiduddin et al., 2012). We then divided the number of 3-year-olds by the class size to create our focal indicator of the proportion of 3-year-olds in each classroom. Table 2 provides detailed descriptive information on this variable, at both a classroom and child level.

Children's language and literacy skills. A measure of children's language and literacy skills was created based on their scores on the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), the Woodcock-Johnson Letter-Word Identification subtest, and the Woodcock-Johnson Spelling subtest (Woodcock, McGrew, & Mather, 2001). These assessments evaluated children's verbal skills as well as their letter-word identification and writing skills. Children who came from non-English-speaking homes were screened for their English proficiency, and if they failed the test, were assessed in Spanish. Most children (82%) took the assessments in English at both times, 8% took them in Spanish at both times, and 10% took the Spanish assessment in the Fall and the English assessment in the Spring. As a precaution, we ran additional models excluding the 10% of children who switched language of assessment, and all findings were qualitatively similar. Finally, because each of the assessments was scored on a different scale, we created standardized scores for each and then averaged them to create a composite for language and literacy skills (Wave 1: α = .65, Wave 2: α = .68; for a similar approach, see Duncan et al., 2015). We ran additional models looking at each of the outcomes individually and found similar effect sizes as those reported below.

Children's math skills. The FACES 2009 data includes a composite measure of children's math skills, which was based on children's scores on the Woodcock-Johnson Applied Problems subtest (Woodcock et al., 2001) and on the Early Childhood Longitudinal Study, Birth Cohort math assessment (Snow et al., 2007; Wave 1: α = .80, Wave 2: α = .82), both of which were administered in the Fall and Spring of the school year. These assessments tapped into children's classification, comparison, and shape-recognition skills. Children not proficient in English were assessed in Spanish.

Children's behavior problems. Teachers provided reports of children's behavior problems at the beginning and end of the year using 14 items derived from the Personal Maturity Scale (Entwisle, Alexander, Cadigan, & Pallis, 1987) and the Behavior Problems Index (Peterson & Zill, 1986). The FACES 2009 data include a composite

Table 1. Descriptive Comparison of the Two Cohorts

Variable	3-year-olds ($n = 1,644$)	4-year-olds ($n = 1,185$)	Difference between cohorts (p)
Child and household characteristics			
Children's gender (proportion female)	.49	.51	> .250
Children's race			
White	.20	.20	> .250
Black	.35	.28	< .001
Latino	.37	.46	< .001
Asian/other	.09	.07	.041
Child age (mean months)	41.26 (3.65)	52.22 (3.80)	< .001
Months between assessments (mean)	5.75 (1.75)	5.92 (0.94)	.005
Language of assessment			
English (Fall), English (Spring)	.83	.82	> .250
Spanish (Fall), Spanish (Spring)	.09	.07	.021
Spanish (Fall), English (Spring)	.08	.11	.016
Mothers' marital status			
Married	.30	.30	> .250
Not married	.18	.19	> .250
Single-parent household	.52	.51	> .250
Mothers' education			
Less than high school	.33	.41	< .001
High school diploma	.35	.33	> .250
Some college	.25	.21	.033
Bachelor's degree	.07	.05	.016
Mothers' age (mean years)	28.76 (5.95)	29.21 (5.94)	.050
Household size (mean number of people)	4.58 (1.63)	4.67 (1.66)	.147
Mothers' employment			
Full time	.27	.26	> .250
Part time	.21	.22	> .250
Unemployed	.52	.52	> .250
Mothers' depressive symptoms (mean)[a]	4.94 (5.98)	4.62 (5.66)	.155
Income/poverty ratio	2.58 (1.39)	2.48 (1.35)	.058
Household language (English)	.76	.68	< .001
Classroom characteristics			
Child/teacher ratio	8.23 (2.01)	8.75 (2.46)	< .001
Child/adult ratio	7.21 (2.12)	7.34 (2.10)	.114
Class size (mean number of children)	16.68 (2.32)	17.80 (1.86)	< .001
Teachers' depressive symptoms (mean)[a]	4.52 (4.86)	3.96 (4.18)	.002
Hours of school per week	26.20 (12.47)	25.91 (12.15)	> .250
Multilingual instruction	.33	.39	< .001
Teachers' years of teaching experience (mean)	12.85 (8.51)	13.39 (8.77)	.103
Teachers' education			
High school	.06	.08	.101
Some college	.12	.08	< .001
Associate's degree	.36	.30	< .001
Bachelor's degree	.36	.40	.012
Some graduate school	.10	.14	.004
Teachers' degree in early childhood education	.93	.91	.112
Teachers' race			
White	.40	.46	.003
Black	.38	.25	< .001
Latino	.17	.26	< .001
Asian/other	.05	.03	.005

(continued)

Table 1. *(continued)*

Variable	3-year-olds (n = 1,644)	4-year-olds (n = 1,185)	Difference between cohorts (p)
Teachers' hourly salary ($)	13.31 (5.39)	14.27 (5.39)	< .001
Teachers' benefits (scale from 0–9)	6.89 (2.00)	6.52 (2.42)	< .001
Outcomes[b]			
Language and literacy skills (mean in Fall)	−0.23 (0.89)	0.30 (1.05)	< .001
Language and literacy skills (mean in Spring)	−0.24 (0.90)	0.35 (1.03)	< .001
Math skills (mean in Fall)	10.89 (4.95)	16.40 (6.61)	< .001
Math skills (mean in Spring)	15.18 (6.78)	21.83 (8.41)	< .001
Behavior problems (mean in Fall)	5.00 (4.62)	3.98 (4.27)	< .001
Behavior problems (mean in Spring)	4.62 (4.68)	3.79 (4.44)	< .001
Social skills (mean in Fall)	14.36 (4.81)	16.37 (4.82)	< .001
Social skills (mean in Spring)	16.59 (4.67)	18.19 (4.38)	< .001

Note: For 3- and 4-year-olds, values shown are proportions unless otherwise indicated (proportions might not sum to 1.00 because of rounding). Standard deviations are given in parentheses.
[a]Parents' and teachers' depressive symptoms were assessed via the short form of the Center for Epidemiological Studies–Depression Scale (Radloff, 1977). [b]See the Method section for details on the measures used to obtain scores for the four outcomes.

of these assessments, which were obtained using a 3-point Likert scale (0, *never*, to 2, *very often*) and tapped into children's aggressive, hyperactive, and withdrawn behaviors (Wave 1: α = .88, Wave 2: α = .87).

Children's social skills. During both the Fall and Spring semesters, teachers also reported on children's social skills (e.g., how often children followed directions, helped put things away, followed rules) using 12 items drawn from the Personal Maturity Scale (Entwisle et al., 1987) and the Social Skills Rating System (Gresham & Elliott, 1990). The FACES 2009 data includes a composite of these items, which were obtained using a 3-point Likert scale (0, *never*, to 2, *very often*), with higher numbers indicative of more optimal social skills (Wave 1: α = .89, Wave 2: α = .89).

Covariates. To reduce the possibility of spurious associations, we adjusted for a theoretically relevant set of child-, household-, teacher-, and classroom-level variables. Child and household covariates were children's gender, children's race/ethnicity, children's age at the start of school, months between assessments, language of assessment, mothers' education, mothers' age, mothers' employment status, mothers' marital status, mothers' depressive symptoms, ratio of income to poverty level, household size, and household language. We also controlled for a full set of classroom and teacher characteristics, namely teacher-child ratio, adult-child ratio, class size, whether the classroom was multilingual (English only vs. English and Spanish), teachers' race/ethnicity, teachers' depressive symptomology, teachers' education, whether teachers' degrees were in early childhood education, teachers' years of experience, teachers' benefits (e.g., paid vacation, sick leave), and teachers' hourly salary. Finally, all models adjusted for children's baseline skills; that is, we estimated whether classroom age composition was associated with changes in children's

Table 2. Distribution of Classrooms, 3-Year-Olds, and 4-Year-Olds According to the Percentage of 3-Year-Olds in the Classroom

Classrooms by percentage of 3-year-olds	Proportion of all classrooms	Proportion of 3-year-olds in each type of classroom	Proportion of 4-year-olds in each type of classroom
Classrooms with 0% 3-year-olds	.14	.00	.26
Classrooms with 1–19% 3-year-olds	.18	.10	.27
Classrooms with 20–39% 3-year-olds	.27	.26	.30
Classrooms with 40–59% 3-year-olds	.16	.18	.11
Classrooms with 60–79% 3-year-olds	.08	.11	.04
Classrooms with 80–99% 3-year-olds	.08	.16	.02
Classrooms with 100% 3-year-olds	.09	.19	.00

school-readiness outcomes, which is one of the strongest adjustments for omitted-variable bias (National Institute of Child Health and Human Development Early Child Care Research Network & Duncan, 2003).

Plan of analysis

Ordinary-least-squares (OLS) regression analyses were conducted using the Stata program (StataCorp, 2011). To address issues of missing data (5–18%), we imputed 50 data sets through the chained-equations method. We accounted for the nesting of children in classrooms using robust standard errors clustered at the classroom level (see also Duncan et al., 2015; Weiland & Yoshikawa, 2014). Similar to multilevel modeling, clustered standard errors correct for the nonindependence of observations due to multiple children being in the same classroom. We also included longitudinal weights to adjust for bias that may arise because of cross-wave attrition, to account for sampling stratification, and to ensure that the data were nationally representative. Because we had hypothesized that classroom age composition would differentially affect younger and older children (Moller et al., 2008), all analyses were conducted separately for each cohort. Thus, the comparisons were not across age groups, but rather within the two cohorts; therefore, we examined the associations between classroom age composition and 3- and 4-year-olds' school-readiness gains. In our first set of models, we examined classroom age composition as a linear variable, whereas in our second set of models, we examined threshold effects. These thresholds, which are described more fully in the Results section, were created and examined—within each cohort—to determine whether there was a tipping point at which classroom age composition was more strongly or weakly associated with children's school-readiness outcomes.

As an additional step, we tested propensity-score-matching (PSM) models, which are a strong method of controlling for selection on observable factors and thereby increasing confidence in causal inference (Rosenbaum & Rubin, 1983). PSM involves four steps. First, we conducted logit models in each of the 50 imputed data sets to estimate the likelihood that children were enrolled in a classroom that was above the threshold of interest. In these models, we included the entire set of child, household, and teacher variables listed in Table 1 as covariates, in addition to assessments of children's school-readiness skills at the start of the Head Start year. Next, for our matching models, we used the nearest-neighbor method (with four matches) within a caliper of .01, which ensured a sufficient overlap of propensity scores between the comparison conditions. We then examined the quality of the matches in several ways. First, we regressed each covariate on the indicator

variable that distinguishes cases above the threshold from the comparison cases using the propensity-score weight (see also Jenkins et al., 2015), which allowed us to examine the propensity-score-adjusted means across groups. We also checked the standardized mean differences for all of our covariates across each comparison group, using the |.10| benchmark for assessing balance and the joint significance of overall balance using a Hotelling test. Finally, even if the full sample achieved balance, it is possible that differences might have been found within differing levels of the propensity scores; thus, as a final check of balance, we separated the comparison conditions into quartiles on the basis of children's propensity scores. If there are no significant differences across these tests, then balance is achieved. After creating the matched samples, we replicated our OLS models using the propensity-matched data, in which we included the full set of covariates in order to adjust for any potential remaining bias from measured characteristics (Berger, Brooks-Gunn, Paxson, & Waldfogel, 2008; Coley & Lombardi, 2013).

Results

Descriptive statistics for classroom age composition (both at the classroom and child level, separated by cohort) are provided in Table 2. Seventy-seven percent of all Head Start classrooms were considered to be mixed age, while 14% contained only 3-year-olds, and 9% contained only 4-year-olds. The 3-year-olds were slightly more likely than the 4-year-olds to be enrolled in a classroom with at least one peer of a different age (81% vs. 74%). Additionally, on average, 3-year-olds were in classrooms where 59% of their classmates were 3 years old or younger, whereas 4-year-olds were in classrooms with a fewer percentage of 3-year-olds (21%). Thus, there is considerable heterogeneity in the distributions of 3- and 4-year-olds across Head Start classrooms.

Regression models

We began by examining the associations between the continuous scale of classroom age composition and the school-readiness gains of 3- and 4-year-olds. Although classroom age composition was not associated with growth in the academic skills of 3-year-olds (Table 3), a higher proportion of 3-year-olds was negatively associated with gains in math and in language and literacy skills among 4-year-olds (7% of a standard deviation; Table 3). There was, however, no relation between classroom age composition and children's social or behavioral skills for either age group.

Next, we conducted threshold analyses to determine whether there was a point at which the association

Table 3. Results of Models Using Classroom Age Composition to Predict Gains in Children's School Readiness

School-readiness outcome	3-year-olds			4-year-olds		
	b	*p*	R^2	*b*	*p*	R^2
Language and literacy skills	−0.00 [−0.06, 0.05]	> .250	.52	−0.07 [−0.13, −0.02]	.006	.65
Math skills	0.03 [−0.03, 0.09]	> .250	.56	−0.07 [−0.12, −0.02]	.009	.66
Behavior problems	−0.04 [−0.10, 0.02]	.201	.49	0.02 [−0.05, 0.09]	> .250	.45
Social skills	0.06 [−0.02, 0.13]	.160	.39	0.03 [−0.03, 0.09]	> .250	.36

Note: Values in brackets are 95% confidence intervals. All variables were standardized, and therefore the unstandardized regression coefficients in this table correspond to effect sizes (i.e., standard-deviation units). Models were adjusted for the clustering of children in classrooms and all covariates listed in Table 1.

between mixed-age classrooms and children's school-readiness outcomes was stronger. As recently discussed by Weiland and Yoshikawa (2014), there is no agreed-on method for selecting a threshold, but possibilities include inflection points, conceptually defined points, empirically identified points, and nonlinear regression analyses. In this study, we tested several possible thresholds but focused on those that corresponded to the mean percentage of 3-year-olds in a classroom as well as 1 standard deviation above and below the mean within the samples of 3-year-olds (1 *SD* below the mean: ≥ 30%; mean: ≥ 60%; 1 *SD* above the mean: ≥ 90%) and 4-year-olds (1 *SD* below the mean: > 0%; mean: ≥ 20%; 1 *SD* above the mean: ≥ 45%; see Table 4). In other words, for each cohort, we tested three different thresholds in three separate models that correspond to low (1 *SD* below the mean), medium (mean), and high (1 *SD* above the mean) percentages of 3-year-olds. For example, the analyses of the low threshold compared children who were in classrooms that were 1 standard deviation below the mean with those who were above this threshold, whereas the analyses of the high threshold compared children who were 1 standard deviation above the mean with those children below this threshold. We also tested OLS models with quadratic terms, but we found no evidence for a nonlinear association so these models are not discussed further.

Results from these analyses demonstrated a similar but stronger association between classroom age composition and the academic skills of 4-year-olds (see Table 5). Specifically, 4-year-olds who were enrolled in classrooms at or above the mean percentage of 3-year-olds demonstrated fewer gains in both math skills and language and literacy skills (11–12% of a standard deviation) than 4-year-olds who were enrolled in classrooms with fewer 3-year-olds. These associations were considerably stronger at the high threshold, corresponding to 26 to 29% of a standard deviation and over 4 months of academic development (calculated by dividing the standardized difference in academic test scores by the regression slope of children's age; Bradbury, Corak, Waldfogel, & Washbrook, 2011). Thus, the negative consequences of mixed-age classrooms for 4-year-olds' academic skills were greatest when classrooms had a nearly equal distribution of 3- and 4-year-olds. No thresholds were identified for the links between classroom age composition and 4-year-olds' social and behavioral skills (see Table 6), nor were there any consistent patterns that emerged for 3-year-olds' school-readiness outcomes.

PSM models

We utilized PSM methods to address potential selection bias as a function of preexisting differences. After fully balancing the comparison conditions (Hotelling *F*s = 0.08–1.30, n.s.; full results are available on request), we found that the PSM models confirmed the conclusions drawn from the OLS models (see Tables 5 and 6). As

Table 4. Distribution of 3- and 4-Year-Olds in Classrooms Meeting the Different Thresholds of Classroom Age Composition

Threshold	3-year-olds		4-year-olds	
	Threshold criterion	Proportion of 3-year-olds	Threshold criterion	Proportion of 4-year-olds
Low (≥ 1 *SD* below the mean)	30%+ 3-year-olds	.76	0%+ 3-year-olds	.74
Medium (≥ mean)	60%+ 3-year-olds	.47	20%+ 3-year-olds	.47
High (≥ 1 *SD* above the mean)	90%+ 3-year-olds	.27	45%+ 3-year-olds	.21

Table 5. Results of Threshold Models Using Classroom Age Composition to Predict Gains in Children's Academic Skills

	Language and literacy skills				Math skills			
	OLS model		PSM model		OLS model		PSM model	
Age cohort and threshold group	b	p	b	p	b	p	b	p
3-year-olds								
Low (≥ 1 SD below the mean)	−0.04 [−0.14, 0.06]	> .250	0.02 [−0.09, 0.13]	> .250	0.04 [−0.08, 0.15]	> .250	0.16 [0.04, 0.29]	.010
Medium (≥ mean)	0.01 [−0.10, 0.12]	> .250	0.00 [−0.11, 0.12]	> .250	0.05 [−0.06, 0.16]	> .250	0.07 [−0.04, 0.19]	.210
High (≥ 1 SD above the mean)	0.05 [−0.07, 0.17]	> .250	0.02 [−0.10, 0.13]	> .250	0.11 [−0.01, 0.22]	.067	0.10 [−0.01, 0.22]	.064
4-year-olds								
Low (≥ 1 SD below the mean)	−0.11 [−0.22, 0.01]	.073	−0.11 [−0.21, −0.01]	.032	−0.08 [−0.19, 0.03]	.148	−0.05 [−0.15, 0.05]	> .250
Medium (≥ mean)	−0.12 [−0.22, −0.01]	.026	−0.13 [−0.24, −0.02]	.017	−0.11 [−0.21, −0.01]	.034	−0.10 [−0.20, −0.00]	.047
High (≥ 1 SD above the mean)	−0.26 [−0.40, −0.11]	< .001	−0.30 [−0.44, −0.15]	< .001	−0.29 [−0.43, −0.14]	< .001	−0.17 [−0.31, −0.03]	.017

Note: Values in brackets are 95% confidence intervals. See Table 4 for explanation of thresholds. All variables were standardized, and, therefore, the unstandardized regression coefficients in this table correspond to effect sizes (i.e., standard-deviation units). Models were adjusted for the clustering of children in classrooms and all covariates listed in Table 1. OLS = ordinary least squares; PSM = propensity-score matching.

before, classroom age composition was significantly and consistently associated with gains in math skills and language and literacy skills among 4-year-olds, and the effects were strongest at the high threshold. These effect sizes were comparable with the OLS models at the medium threshold; however, at the high threshold, these estimates slightly increased for children's language and literacy skills (from 26% to 30% of a standard deviation) and slightly decreased for children's math skills (from 29% to 17% of a standard deviation). Propensity-score analyses also revealed that 3-year-olds exhibited somewhat greater social and math skills when they were

Table 6. Results of Threshold Models Using Classroom Age Composition to Predict Gains in Children's Behavior Problems and Social Skills

	Behavior problems				Social skills			
	OLS model		PSM model		OLS model		PSM model	
Age cohort and threshold group	b	p	b	p	b	p	b	p
3-year-olds								
Low (≥ 1 SD below the mean)	−0.01 [−0.12, 0.10]	> .250	−0.07 [−0.20, 0.06]	> .250	0.05 [−0.08, 0.19]	> .250	0.14 [−0.03, 0.31]	.116
Medium (≥ mean)	−0.12 [−0.26, 0.02]	.085	−0.15 [−0.32, 0.03]	.103	0.11 [−0.07, 0.29]	.233	0.22 [0.05, 0.40]	.013
High (≥ 1 SD above the mean)	−0.03 [−0.16, 0.10]	> .250	−0.08 [−0.22, 0.06]	.248	0.13 [−0.05, 0.30]	.150	0.17 [−0.01, 0.35]	.068
4-year-olds								
Low (≥ 1 SD below the mean)	0.12 [−0.05, 0.28]	.169	0.15 [−0.02, 0.31]	.086	0.01 [−0.15, 0.17]	> .250	−0.03 [−0.20, 0.13]	> .250
Medium (≥ mean)	0.05 [−0.09, 0.18]	> .250	0.07 [−0.07, 0.21]	> .250	0.05 [−0.08, 0.18]	> .250	0.03 [−0.11, 0.16]	> .250
High (≥ 1 SD above the mean)	−0.07 [−0.25, 0.11]	> .250	−0.07 [−0.25, 0.11]	> .250	0.13 [−0.05, 0.31]	.164	0.04 [−0.15, 0.23]	> .250

Note: Values in brackets are 95% confidence intervals. See Table 4 for explanation of thresholds. All variables were standardized, and, therefore, the unstandardized regression coefficients in this table correspond to effect sizes (i.e., standard-deviation units). Models were adjusted for the clustering of children in classrooms and all covariates listed in Table 1. OLS = ordinary least squares; PSM = propensity-score matching.

enrolled in classrooms with a greater number of 3-year-olds, but the threshold was not consistent (see Tables 5 and 6).

Robustness checks

Although PSM methods addressed issues of selection bias among the measured variables, concerns regarding unmeasured variables persisted. To assess the potential confounding role of omitted variables, we conducted impact-threshold-for-confounding-variables (ITCV; Frank, 2000) analyses. ITCV analyses quantify the degree to which an unknown variable would have to correlate with both the predictor and outcome variables of interest to negate the observed associations. Results from these analyses indicated that associations between classroom age composition and the academic skills of 4-year-olds were unlikely to be negated by unobserved confounds (ITCV estimates = .04). As one example, an unmeasured factor would have to correlate with both classroom age composition and children's academic skills at a minimum of .20 to negate the documented associations, thereby lending confidence to our conclusions.

Discussion

Although mixed-age classrooms have long been a part of the American educational system (Katz et al., 1990; Moiduddin et al., 2012), the potential costs and benefits of mixing children of different ages has received little empirical attention, especially in early-care and education programs serving low-income children. This study addressed these gaps in the extant literature using regression models and propensity-score matching with data from the recently released and nationally representative FACES 2009 cohort.

Our primary finding was that mixed-aged classrooms appear to have negative implications for the academic achievement of 4-year-olds. Specifically, 4-year-olds who were enrolled in mixed-aged classrooms demonstrated fewer gains in math and in language and literacy skills than did 4-year-olds who were in classrooms with fewer 3-year-olds; these differences were statistically significant and consistent across a number of model specifications. In contrast, classroom age composition was not consistently associated with 3-year-olds' school readiness, nor was it related to the social behaviors of either 3- or 4-year-olds. This might be because the magnitude of differences across age cohorts were larger for children's academic skills (50–85% of a standard deviation) compared with their social-behavioral skills (20–40% of a standard deviation; see Table 1). Moreover, these findings are in line with past research that has shown that young children's academic skills are affected more by early-care

and education programs than are their social behaviors (Forry, Davis, & Welti, 2013; Puma et al., 2010; Winsler et al., 2008).

These findings fill a critical gap in the existing literature, which has offered conflicting and inconclusive evidence on the associations between mixed-age classrooms and children's school success (Bell et al., 2013; Guo et al., 2014; Moller et al., 2008; Winsler et al., 2002). Just as important, these results may point to one of the reasons why the national evaluation of Head Start yielded very modest impacts for 4-year-olds (Puma et al., 2010); given that 75% of Head Start classrooms are mixed age (Moiduddin et al., 2012), 4-year-olds in the program have a strong chance of being in a classroom environment that does not appear to promote their academic learning. Future evaluators of Head Start should consider the role of mixed-age classrooms and peer effects when examining the program's impact on children's school readiness.

Previous research on the role of peer effects in mixed-age classrooms had not identified thresholds at which these factors yielded less optimal outcomes for children, in part because of small sample sizes (Blasco et al., 1993; Goldman, 1981; Guo et al., 2014; Winsler et al., 2002). In the present study, we found that even a moderate number of 3-year-olds in Head Start classrooms (i.e., threshold of 20%) resulted in less optimal academic achievement for older children, corresponding to a loss of roughly 2 months of academic development; however, these negative associations were considerably larger at the high threshold (i.e., 45%), which translates to roughly 4 to 5 months of academic development. These results provide compelling evidence that 4-year-olds enrolled in mixed-age Head Start classrooms are less likely to enter school ready to learn in the domains of math and of language and literacy. Although there was some evidence from the PSM models to suggest that 3-year-olds demonstrated greater gains in math and social skills when they were enrolled in classrooms with fewer older children, the threshold was not consistent.

Although we examined the associations between classroom age composition and children's school readiness, we did not examine the underlying mechanisms. Previous literature suggests two plausible pathways. In the direct-effects pathway, children are directly influenced by their classmates' abilities through classroom interactions (Justice et al., 2014; Mashburn et al., 2009), which may be beneficial for younger peers in the classroom, but not older children (Moller et al., 2008). The second possibility is an indirect-effects pathway, whereby teachers modify their classroom practices to accommodate a wide range of skill levels, which results in potential disengagement among older and more advanced children (Urberg & Kaplan, 1986). It is likely that both pathways underlie the documented associations, but the

present findings cannot speak to whether one or both pathways are at play. Future studies are needed to examine these pathways within the Head Start setting, as they have the potential to inform the design and implementation of Head Start programs.

The findings reported here must be interpreted in light of two limitations. First, this study did not include data on the exact ages of children's classmates. The FACES 2009 data reveals that, on average, 4-year-olds were 52 months of age, whereas 3-year-olds were 41 months of age; thus, there is roughly a full year of difference between 3- and 4-year-olds. In other words, it is likely that 3-year-olds were 2 years away from kindergarten entry, whereas 4-year-olds were 1 year away. Second, the external validity of our findings is limited, as it is only strictly applicable to Head Start programs and not necessarily to other child care or formal education settings.

With these limitations in mind, this study provides much-needed insight into the implications of mixed-age classrooms in the Head Start program. Ultimately, these findings reveal that mixed-age classrooms are associated with smaller academic gains for older children, which translates to 4 to 5 fewer months of academic development compared with their peers in classrooms with fewer younger children. Thus, despite the enthusiasm regarding the potential for mixed-age classrooms to facilitate children's school readiness, these findings underscore the need for greater caution and continued scientific investigation into the potential mechanisms underlying these findings.

Author Contributions

A. Ansari helped conceptualize the study, conducted all statistical analyses, and drafted the initial manuscript. K. Purtell helped conceptualize the study and analyses and helped in writing and revising the manuscript. E. Gershoff helped conceptualize the study and reviewed and revised the manuscript. All authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

References

Ansari, A., & Gershoff, E. (2015). Learning-related social skills as a mediator between teacher instruction and child achievement in Head Start. *Social Development*, *24*, 699–715. doi:10.1111/sode.12124

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

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 Childhood Research Quarterly*, *28*, 1–10. doi: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. doi:10.1016/j.childyouth.2007.10.010

Blasco, P. M., Bailey, D. B., & Burchinal, M. A. (1993). Dimensions of mastery in same-age and mixed-age integrated classrooms. *Early Childhood Research Quarterly*, *8*, 193–206. doi:10.1016/S0885-2006(05)80090-0

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* (Institute for the Study of Labor Discussion Paper No. 6120). Retrieved from http://www.econstor.eu/bitstream/10419/58643/1/690078234.pdf

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. doi:10.1111/j.1467-8624.2012.01840.x

Duncan, G. J., Jenkins, J. M., Auger, A., Burchinal, M., Domina, T., & Bitler, M. (2015). *Boosting school readiness with preschool curricula*. Retrieved from http://inid.gse.uci.edu/files/2011/03/Duncanetal_PreschoolCurricula_March-2015.pdf

Duncan, G. J., & Magnuson, K. A. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, *27*, 109–132. doi:10.1257/jep.27.2.109

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture and Vocabulary Test, Third Edition: Examiner's manual and norms booklet*. Circle Pines, MN: American Guidance Service.

Entwisle, D. R., Alexander, K. L., Cadigan, D., & Pallis, P. (1987). The emergent academic self-image of first graders: Its response to social structure. *Child Development*, *58*, 1190–1206. doi:10.1111/j.1467-8624.1987.tb01451

Forry, N. D., Davis, E. E., & Welti, K. (2013). Ready or not: Associations between participation in subsidized child care arrangements, pre-kindergarten, and Head Start and children's school readiness. *Early Childhood Research Quarterly*, *28*, 634–644. doi:10.1016/j.ecresq.2013.03.009

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research*, *29*, 147–194. doi:10.1177/0049124100029002001

Goldman, J. A. (1981). Social participation of preschool children in same- versus mixed-age groups. *Child Development*, *52*, 644–650. doi:10.2307/1129185

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System*. Circle Pines, MN: 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. doi:10.1080/10409289.2014 .893759

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, *46*, 289–324. doi:10.1111/j.1465-7295.2008.00163.x

Henry, G. T., & Rickman, D. K. (2007). Do peers influence children's skill development in preschool? *Economics of Education Review*, *26*, 100–112. doi:10.1016/j.econedurev .2005.09.006

Jenkins, J. M., Farkas, G., Duncan, G. J., Burchinal, M., & Vandell, D. L. (2015). Head Start at ages 3 and 4 versus Head Start followed by state pre-K: Which is more effective? *Educational Evaluation and Policy Analysis*. Advance online publication. doi:10.3102/0162373715587965

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. doi:10.1177/0956797614538978

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

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. doi:10.1111/j.1467-8624.2009.01291.x

Moiduddin, E., Aikens, N., Tarullo, L. B., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009*. Washington, DC: Office of Planning, Research, and Evaluation, U.S. Department of Health and Human Services.

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. doi: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. doi:10.1111/1467-8624.00617

Peterson, J. L., & Zill, N. (1986). Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family*, *48*, 295–307. doi:10.2307/352397

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., . . . Spier, E. (2010). *Head Start Impact Study* (Final Report). Washington, DC: U. S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401. doi:10 .1177/014662167700100306

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

Snow, K., Thalji, L., Derecho, A., Wheeless, S., Lennon, J., Kinsey, S., Rogers, J., . . . Park, J. (2007). *Early Childhood Longitudinal Study, Birth cohort (ECLS-B), preschool year data file user's manual*. Washington, DC: National Center for Education Statistics.

StataCorp. (2011). *Stata statistical software: Release 12*. College Station, TX: Author.

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. doi: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. doi:10.3102/00346543065004319

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.

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

Winsler, A., Caverly, S. L., Willson-Quayle, A., Carlton, M. P., Howell, C., & Long, G. N. (2002). The social and behavioral ecology of mixed-age and same-age preschool classrooms: A natural experiment. *Journal of Applied Developmental Psychology*, *23*, 305–330. doi:10.1016/S0193-3973(02)00111-9

Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*, 314–329. doi:10.1016/j.ecresq.2008.02.003

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishing.

Yudron, M., Jones, S. M., & Raver, C. C. (2014). Implications of different methods for specifying classroom composition of externalizing behavior and its relationship to social–emotional outcomes. *Early Childhood Research Quarterly*, *29*, 682–691. doi:10.1016/j.ecresq.2014.07.007