
Routledge
Taylor & Francis Group

Applied Developmental Science

ISSN: 1088-8691 (Print) 1532-480X (Online) Journal homepage: https://www.tandfonline.com/loi/hads20

# Does mixing age groups in early childhood education settings support children's language development?

**Laura M. Justice, Jessica A. Logan, Kelly Purtell, Dorthe Bleses & Anders Højen**

**To cite this article:** Laura M. Justice, Jessica A. Logan, Kelly Purtell, Dorthe Bleses & Anders Højen (2019) Does mixing age groups in early childhood education settings support children's language development?, Applied Developmental Science, 23:3, 214-226, DOI: 10.1080/10888691.2017.1386100

**To link to this article:** https://doi.org/10.1080/10888691.2017.1386100

APPLIED DEVELOPMENTAL SCIENCE
2019, VOL. 23, NO. 3, 214–226
https://doi.org/10.1080/10888691.2017.1386100

Routledge
Taylor & Francis Group



# Does mixing age groups in early childhood education settings support children's language development?

Laura M. Justice[a], Jessica A. Logan[a], Kelly Purtell[a], Dorthe Bleses[b], and Anders Højen[b]

[a]The Ohio State University; [b]Aarhus University

**ABSTRACT**

As early childhood education programming expands across the globe, there is an increased need to understand how features of these programs influence children's development. The composition of children's age within a classroom is one such feature, although it is much less studied than other features. Theoretical and empirical evidence suggests that children's development may be influenced by the age range of their classmates. This study examines the relations between classroom age variability on children's vocabulary development for 2,743 children between the ages of two years, nine months, and six years, 11 months enrolled in early childhood education settings in Denmark. Findings indicate a significant nonlinear relationship between the range of child age within a classroom and children's vocabulary development, such that classrooms with a maximum age range of 24 months were associated with the greatest gains in vocabulary growth. Results give direction to policy efforts focused on expansion of early childhood education programming.

Early childhood education (ECE) programs exist across the world to serve several important societal goals, including supporting the labor-market participation of parents of young children, especially women (Kamerman, 1991), and fostering the early developmental competencies of young children (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002). Across North America, there are currently numerous efforts at the local, state, and federal levels to improve access to ECE programs for young children to enhance labor-market participation among parents (Lefebvre, Merrigan, & Verstraete, 2009) and to improve children's development of key cognitive and noncognitive skills as well as readiness for formal schooling at age 5 years (Winsler et al., 2008). Presently, in the United States, about 50% of 3- to 5-year-olds participate in formal full-day center-based ECE (National Center for Education Statistics, 2017), and this number is likely to increase as many cities and states press for expanded access to ECE for young children. In some countries within the European Union, universal participation in ECE is the norm, including France, Spain, and Belgium, which each have near saturation of children in ECE by three years of age, with an average participation rate of about 80% among the EU 27 [Organisation for Economic Co-operation and Development (OECD), 2016]. Denmark, in which the current study was completed, has among the highest rates of ECE participation (94% for 3–5-year olds) for all OECD countries, with provision of universal child care being an important feature of the welfare state (Landersø & Heckman, 2017).

Of relevance to the research reported here, when ECE programs are created and expanded, there are certain *structural features* of such programs that can be directly affected by the policies that create them. These features include, for instance, teacher-child ratio, class size, and teacher credentials. For example, state-sponsored support of ECE programs might stipulate a 1:9 teacher-child ratio with a maximum class size of 18 children, and provision of programming by teachers with at least a bachelor's degree. An additional structural feature often affected by policies is the age composition of ECE classrooms, as creation/expansion policies often stipulate the age range of children who can participate in subsidized programming. For instance, the Commonwealth of Virginia's long-running Virginia Preschool Initiative (VPI) provides funding for more than 10,000 4-year-olds across the state to participate in center-based ECE each year (Huang, Invernizzi, & Drake, 2012). An indirect result of this policy, by stipulating participation only to 4-year-old children, is that it leads to creation of same-age (SA) ECE classrooms rather than mixed-age (MA) classrooms, as VPI-supported classrooms enroll only 4-year-old children in a SA context. In this research,

**CONTACT** Laura M. Justice justice.57@osu.edu 175 E 7th, Ohio State University, Columbus, OH 43201, US.

we view policy decisions regarding the age range of children who can participate in expanding programming as a potentially crucial mechanism for ensuring that ECE expansion meet its desired goals of enhancing children's development, as these decisions directly affect the classroom age composition of programs with respect to whether available programs will feature SA or MA programming.

Decisions about SA and MA classrooms also occur at more micro-levels. In the United States, many private- and public (e.g., Head Start) centers have the option of offering either SA or MA classrooms. Although these decisions are sometimes made for practical reasons, such as meeting class size requirements, center directors and teachers likely also have preferences for which environments they believe are most optimal for children's development. Similarly, parents have preferences and beliefs about what they believe is the best early childhood education program for their child and these beliefs shape their child care decisions (Crosnoe, Purtell, Davis-Kean, Ansari, & Benner, 2016). Parental beliefs about the utility of MA classrooms may, in part, determine the type of early education center in which they enroll their child. Thus, understanding how MA and SA environments shape children's academic development is critical to informing these decisions made at the policy, center, and parent level.

## Age composition in early childhood education

There has long been interest in the relative advantages and disadvantages of SA versus MA programming in ECE. SA classrooms enroll children who are relatively similar in age, as is the case of 4-year-old programs that serve only children between 48 and 60 months (e.g., Virginia's VPI), whereas MA classrooms enroll a broader range of children in terms of age. Potentially, ECE programs could enroll children from soon after birth until kindergarten entry, which typically occurs when children are five or even six years of age, for children whose parents delay kindergarten enrollment. Programs can utilize a variety of approaches for organizing children into classrooms as a function of age: children could be organized into SA classrooms, creating homogenous settings in which classrooms serve only children of a given age, or MA classrooms, characterized by greater heterogeneity in children's ages.

These approaches to age arrangements have historically been grounded in developmental theory. Most notably, supporters of MA classrooms often point to the work of Vygotsky (1978). In his theory of sociocultural learning, Vygotsky focuses on the role of peers in children's learning and notes that interacting with children more- and less-skilled than oneself is beneficial for learning and developing, as it provides opportunities for modeling more advanced behavior and scaffolding learning for younger peers. From this, MA classrooms may be advantageous because they increase the range of peers' skills than a child may be exposed to in their classroom. These theoretical underpinnings have supported the widespread use of mixed-age classrooms, and in particular, guide the Montessori method of schooling (Lillard & Else-Quest, 2006). However, some work has suggested that the benefits of mixed-age programming may vary for older and younger children. Winsler and colleagues (2002) described the results of a natural experiment, in which two SA classrooms (one serving 3-year-olds and the other serving 4-year-olds) were converted to two MA classrooms in a single early childhood program. The authors used this opportunity to examine a range of proximal processes as a function of the SA-MA contrast (Winsler et al., 2002). The authors observed that within MA settings, older children become more like younger children, and younger children become more like older children. Specifically, 3-year-olds in MA settings exhibited goal-directed behaviors similar to 4-year-olds in SA settings, whereas 4-year-olds in MA settings exhibited social affiliations more like 3-year-olds in SA settings. Thus, the MA programming appeared beneficial for some but not all children, with younger children seeming to benefit more than older children.

Such findings lend appeal to having classrooms serve a narrow age range of children, such as only infants or only 4-year-olds, which also may be beneficial to teachers. For instance, a teacher may provide a better quality of instruction and care if children were relatively similar in their ability to attend and learn during common classroom routines. Less effort may need to be directed toward differentiating instruction and managing children's behaviors than might occur if children varied significantly in age range.

## Age composition and children's achievement

Only a handful of studies have sought to estimate the effects of various classroom-age compositions on children's development in such settings. Referenced previously, Winsler and colleagues (2002) examined how MA classrooms influenced children's classroom behaviors separately for the younger (3-year olds) and older (4-year olds) children. Their work showed that younger children in mixed age classrooms derived benefits from being in a classroom environment with older peers, presumably who could model positive behavior and expose them to more cognitively-challenging material. The authors proposed that this MA environment resulted in a "stretching" effect for the younger children. The older children in MA classrooms, however, were found to "act younger," perhaps because they spend more time interacting with children who are less developed behaviorally and cognitively.

More recent studies involving larger samples of classrooms have sought to estimate the actual effects of these age-grouping contrasts on children's achievement, largely focusing on children's language development (Ansari, Purtell, & Gershoff, 2016; Bell, Greenfield, & Bulotsky-Shearer, 2013; Guo, Tompkins, Justice, & Petscher, 2014; Moller, Forbes-Jones, & Hightower, 2008). Language development, primarily children's vocabulary growth, is often studied in relation to classroom composition (e.g., Justice, Logan, Lin, & Kaderavek, 2014), because children's language skills are greatly affected by the characteristics of the linguistic input they experience (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). The ages of the children in a classroom seem to shape the linguistic environment of the ECE classroom. For example, a child in a classroom with a wide age range of peers may be exposed to children with both greater and lower vocabulary skills, which may facilitate or impede his or her vocabulary growth. A recent study showed direct pathways between the language skills of children's classmates on their language gains over a year in preschool, and that these operate independently of teachers' instructional pathways (Yeomans-Maldonado, Justice, & Logan, in press).

The body of work that has sought to determine the extent to which variability in classroom age composition may affect children's language development or similar outcomes provide minimal clarity with respect to the possible effects of MA versus SA groupings. Two studies (Bell et al., 2013; Moller et al., 2008) operationalized variability in classroom age composition by using the standard deviation (*SD)* of children's ages in a classroom as the independent variable, whereas Ansari et al. (2016) used the proportion of 3-year-olds in classrooms and Guo et al. (2014) used *SD* and age range of children in a classroom. Two studies found minimal to no association between variability in classroom age composition and children's development (Bell et al., 2013; Guo et al., 2014) and two reported a negative association between increased variability in classroom age composition (as would occur in MA classrooms) and children's outcomes (Ansari et al., 2016; Moller et al., 2008). The latter work suggests that having a larger range of ages in ECE settings may have negative effects on the development of older children in these classrooms. If this is the case, it would suggest that SA programming is more desirable than MA programming, at least for 3- and 4-year-old children.

There are several important limitations of these studies, which helped to motivate the present investigation. First, all have focused exclusively on investigating classroom age composition for programs serving only children from low-income backgrounds. This detracts from the potential generalizability of findings to classrooms serving a socioeconomically diverse range of students. Further, it can negatively affect internal validity in that children in such studies may show a more limited range in skills (and development over time) than would occur for a more diverse sample. Second, the age range of children within classrooms studied in these works provided only limited age variability within classrooms. For instance, Bell et al (2013) examined the relations between classroom age composition and child outcomes in Head Start classrooms. On average, children within each classroom ranged in age only up to about 12 months. The children in this study were also relatively narrow in their ages, with a mean of 48 months and a standard deviation of only seven months. As with the Ansari et al. (2016) study, which also relied on a Head Start sample, it is difficult to draw firm conclusions on the effects of MA versus SA programming on ECE participants using samples that do not represent the full range of classroom age composition possibilities nor the plausible age range of possible participants in ECE programs.

## Danish ECE context

Denmark is a welfare state that provides universal childcare (ECE) to children soon after birth, with an enrollment rate of 85% of 1–2 year olds and 94% of all 3–5 year olds (OECD Family Database, 2016). The ECE facilities are organized in public, semi-public, and private centers or in publicly organized nonparental home-based care. All types of ECE facilities are audited by the municipality and are regulated by national laws concerning educational quality and safety. Costs are highly subsidized, with a minimum of 75% of the cost covered by the municipality and larger subsidies provided for low-income households (for further discussion of this context, see Gupta & Simonsen, 2010).

Danish ECE programs are situated in most residential areas across Denmark; hence, their socioeconomic composition is solely determined by the composition of the neighborhood. Although programs with a high proportion of at-risk children do exist, Danish ECE programs typically serve a more socioeconomically diverse group of children as compared to subsidized programs in the United States, which often feature targeted enrollment of only low-income children.

Danish ECE programs are often considered to relatively high quality in terms of structural characteristics (Esping-Andersen et al., 2012). For instance, the average children-to-teacher ratio for preschoolers is 7:1. Approximately 60% of teachers have a 3.5-year pedagogical Bachelor degree and the other 40% are mainly teaching assistants who do not have a university degree. Beyond these structural characteristics, observational work suggests that the observed classroom quality is similar to the United States. Specifically, the emotional quality of classrooms is relatively high, and the instructional quality of classrooms is fairly low (Bleses et al., 2017). Emotional quality captures the general climate of the classroom in terms of warmth and sensitivity, whereas instructional quality captures the climate of the classroom in terms of cognitive challenge. The low levels of the latter may reflect to the focus of the ECE curriculum that is common to many European countries. This curriculum typically emphasizes children's social-emotional development, as opposed to a cognitive or academic orientation, and is characterized by a holistic and strong play-oriented curriculum with little time spent on formal instruction (Sylva, Ereky-Stevens, & Aricescu, 2015). Legislation implemented in 2004 requires all programs to formulate 'learning plans' focused on children's development in the areas of social competence, language, body and movement, nature and natural phenomena, and cultural expressions and values (Sylva et al., 2015).

## Present study

The present study was conducted to improve our understanding of the effects of classroom age composition on children's language development, particularly their vocabulary growth, drawing upon a large, population-based sample of children attending ECE programs in Denmark during the 2013 and 2014 years. Involving the largest and most socioeconomically diverse sample of children to date in research on this topic, involving 3,340 children between the ages of 2 years, 9 months, and 6 years, 11 months, the present study exhibits several compelling strengths. First, the Danish child-care system offers an advantageous context to address questions regarding classroom age composition, as formal child-care participation rates in Denmark are among the highest of all OECD countries, as noted previously. For infants and toddlers (up to 3 years of age), about 70% of children are in formal child care (the highest rate of participation among OECD countries), and about 94% for 3- to 5-year-olds (within the top 10 of OECD countries) (OECD Family Database, 2016). With data in this study representing 13 (out of 98) municipalities across the country, the sample in this study more closely approximates a population base from which to derive estimates of classroom-age composition effects.

Second, children in Denmark participate in publicly funded child-care programs irrespective of the family's socioeconomic status or ability to pay for such services, thereby mitigating the selection of families into particular programs as occurs in the United States. Relatedly, programs are not aligned to a specific funding source, such as public-preK initiatives or Head Start, as is common in the United States. As a result, the sample included in the present study is more socioeconomically diverse than that a number of studies conducted within the United States that focus on children in targeted enrollment preschool programs.

Third, publicly funded child-care in Denmark is available to families when their children are as young as three months of age, resulting in a system in which a large number of infants, toddlers, and preschoolers attend centers together. Consequently, many classrooms in these centers enroll mixed-aged groupings, with children potentially aging in range from birth to six years. As a result of these phenomena, the classrooms included in this study enrolled children who ranged in age, on average, 24 months. This distinguishes the present study from those conducted in ECE programs in the United States, in which children in a given classroom typically ranged in age only up to about 12 months (e.g., Bell et al., 2013). This allows us to more rigorously examine the effects of classroom age composition, especially the effects of range, on children's language outcomes.

Two questions were addressed in this study. The first question asked: To what extent is variability in the age composition of ECE programs associated with children's vocabulary growth? We addressed this question in two ways. First, we predicted children's vocabulary growth from the *standard deviation* of children's age in months at the start of the year, controlling for relevant covariates, similar to Bell et al. (2013). Because we found that the within-classroom age variability was considerably wider than has been found in other similar work, we also predicted children's vocabulary growth from the *range of ages* within a classroom at the start of the year, also controlling for relevant covariates. In both approaches, we examined the *linear relationship* between classroom age composition and children's outcomes. Second, given the age range present in the classrooms involved in this study, which averaged 24 months between the youngest and oldest child, we also examined whether the relationship between classroom age composition and children's vocabulary growth approximated a *curvilinear relationship*. We examined whether the relation between age composition and language gains was dependent on the degree of variability in age composition. In addressing both of these questions, we also examined whether there was a significant interplay between children's age and classroom age composition, as some prior studies have suggested that MA programming is beneficial only for younger children (Moller et al., 2008).

## Methods

### Participants

There were 3,340 2- to 6-year-old children (54% male) who participated in this study. The children were, on average, 54 months of age ($SD = 10.3$ months), but ranged in age from 33 months (2 years, 9 months) to 83 months (6 years, 11 months). About 10% of the children were considered immigrants, the majority of whom come from eastern European (e.g., Poland) or middle-eastern countries (e.g., Turkey). The children were enrolled in publicly funded Danish child care centers from across the country; their centers were involved in one of two large-scale effectiveness trials supported by the Ministry of Education or the Strategic Research Council of Denmark. Specifically, centers across the country were invited to participate in one of two parallel effectiveness studies, with each involving three experimental conditions plus a control. Random assignment was used for allocation of centers to conditions. Results of those trials are described in several prior reports (Bleses et al., in press).

For our purposes, we used only children who were enrolled in the control condition of either trial, given that our outcome of interest (i.e., language development) was a focus within the effectiveness trials. Of those children, 27 were missing data on vocabulary measures at the pre-test or both the pre- and post-test and were therefore excluded from analyses (0.8% of the total sample). An additional 554 were missing data at the posttest due primarily to matriculation outside of the school system, and another 16 participants were missing data on one or more demographic characteristic. Information about these 570 children contributed to the classroom-based measurement (including the main predictors of interest: age composition within each classroom), but were excluded from the inferential analyses as individuals. The final analyzed dataset included 2,743 students enrolled in 227 different classrooms.

Examination of the patterns of missingness did show that missing data was associated with certain child- and family-level variables. Specifically, missing data from recruitment to posttest was higher for immigrant children ($p < .001$) as well as children with low maternal education, and children from low-income families (both $p = .004$). Again, while these children's data were included in the classroom-based measures of age composition, it was not included in the individual-level analysis; thus, the results may not generalize to all students in a classroom, but rather only to those who remain in class the whole year.

To contextualize the sample in terms of socioeconomic status, we used maternal education (highest level attained) as a proxy; these data were provided by Statistics Denmark. Fourteen percent of children's mothers had primary school as their highest level attained; 35% had high school and/or vocational education; 31% had a four-year college degree; and 20% had advanced university education. Thus, there was a near normal distribution of children's socioeconomic status represented in this study, corresponding to the universal ECE system in which children were enrolled. In terms of ethnic diversity, information from Statistics Denmark was used to identify the number of children who come from families considered immigrants. These may be first or second generation immigrants. In this sample, 10% of children were considered immigrants.

### Procedure

Children's child-care centers were enrolled in one of two parallel effectiveness trials, as noted. As part of those studies, children's centers were randomly assigned to one of four arms of the trial in which they were enrolled, with one arm comprising a true control condition. All children included in the present study were in a control condition, corresponding to "business-as-usual instruction" in their centers. Thus, any effects of classroom age composition in this study correspond to those occurring under typical educational circumstances.

As part of the larger study, children were assessed prior to and following the intervention period, with an approximately 30-week interval between assessments. Assessments were conducted by children's teachers using a familiar instrument that is part of a national language screening program already in place. Teachers receive a training manual and videos to support their use of this tool twice annually with their children. Teachers downloaded all testing materials and stimuli as well as explicit instructions in how to administer the tool and upload data to a central site. Other studies of children's development in ECE settings have relied on teacher-implemented direct assessments (Hamre et al., 2010; Huang et al., 2012).

### Child language assessment

The vocabulary skills of children were assessed pre- and postintervention using one subtest of the standardized, norm-referenced assessment, *Language Assessment of Children: 3–6* (LA; Bleses et al., 2010). The LA is administered in the vast majority of Danish child-care centers as part of a national early language and literacy screening program, and was already in use in all but one set of centers involved in the present study. The LA comprises seven subtests, one of which captures children's vocabulary skills (max score 76). The vocabulary measure was chosen because vocabulary growth has most often been studied in relation to classroom compositional factors in ECE settings (Guo et al., 2014; Justice, Petscher, Schatschneider, & Mashburn, 2011) and is influenced by classroom characteristics. The other measures focus on phonological awareness, alphabet knowledge, and general communication strategies. For the vocabulary subtest, children are presented with a picture and are asked to provide a word that describes that picture; stimuli included nouns (e.g., squirrel), verbs (e.g., writing), and adjectives (e.g., dirty). An individual child was presented with a maximum of 40 age-dependent items.

Psychometric quality of the tool is adequate, including internal consistency (alpha ranging from .75–.91 for individual subtests). Concurrent validity for the vocabulary subtest with standardized assessments of receptive and expressive vocabulary are .55 and .42, respectively.

### Classroom age composition

Classroom age composition was calculated in two ways: (1) using the within-classroom *standard deviation* of children's ages at the start of the study (Bell et al., 2013), and (2) as the within-classroom *range* of ages at the start of the study. To test the curvilinear effect, squared products of each age composition variable were also calculated. Note that all key study variables were examined for outliers and none were identified.

## Results

### Preliminary analyses

Table 1 provides basic descriptions and inter-correlations among the key study variables at both the child level and classroom level. In examining these data, a few points warrant note regarding the classroom-level compositional variables. First, the mean age of children in a classroom is not associated with the age range of children in the classroom, using either standard deviation of age ($r = -.03$) or classroom age range ($r = .03$). This shows that there is no systematic relationship between the age of children in a classroom and their composition of that classroom; classrooms serving older children are no more heterogeneous in age range than classrooms serving younger children. Second, classroom size is associated with the age range of children in the classroom, using either standard deviation of age ($r = -.24$) or classroom age range ($r = .39$). This suggests that classrooms that serve a broader range of children in terms of age tend to serve larger numbers

**Table 1.** Means and correlations for key study variables.

| | | Child | | | | | Classroom | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Correlations | | | | | | | | | | | | |
| 1 | Child Pretest Vocabulary | 1 | 0.86 | −0.03 | 0.62 | −0.29 | 0.41 | 0.00 | 0.29 | 0.28 | −0.02 | 0.00 |
| 2 | Child Posttest Vocabulary | 0.86 | 1 | −0.03 | 0.56 | −0.31 | 0.36 | −0.01 | 0.29 | 0.22 | −0.05 | −0.04 |
| 3 | Child Male | −0.03 | −0.03 | 1 | 0.05 | −0.02 | 0.00 | 0.03 | 0.01 | −0.01 | −0.02 | −0.02 |
| 4 | Child Age | 0.62 | 0.56 | 0.05 | 1 | 0.02 | 0.63 | 0.07 | 0.48 | 0.47 | −0.03 | 0.01 |
| 5 | Child Immigrant | −0.29 | −0.31 | −0.02 | 0.02 | 1 | 0.01 | −0.06 | 0.03 | 0.04 | 0.08 | 0.07 |
| 6 | Class Mean Age | 0.41 | 0.36 | 0.00 | 0.63 | 0.01 | 1.00 | 0.12 | 0.76 | 0.76 | −0.03 | 0.03 |
| 7 | Class Size | 0.00 | −0.01 | 0.03 | 0.07 | −0.06 | 0.12 | 1 | −0.09 | 0.37 | 0.24 | 0.39 |
| 8 | Class Min Age | 0.29 | 0.29 | 0.01 | 0.48 | −0.03 | 0.76 | −0.09 | 1 | 0.29 | −0.59 | −0.57 |
| 9 | Class Max Age | 0.28 | 0.22 | −0.01 | 0.47 | 0.04 | 0.76 | 0.37 | 0.29 | 1 | 0.53 | 0.62 |
| 10 | Class *SD* Age | −0.02 | −0.05 | −0.02 | −0.03 | 0.08 | −0.03 | 0.24 | −0.59 | 0.53 | 1 | 0.93 |
| 11 | Class Age Range | 0.00 | −0.04 | −0.02 | 0.01 | 0.07 | 0.03 | 0.39 | −0.57 | 0.62 | 0.93 | 1 |
| Descriptives | | | | | | | | | | | | |
| | Mean | 0.08 | 0.55 | 0.54 | 53.95 | 0.10 | 54.00 | 19.97 | 41.91 | 66.55 | 7.77 | 24.64 |
| | *SD* | 0.97 | 0.87 | 0.50 | 10.28 | 0.30 | 6.46 | 11.63 | 8.04 | 8.45 | 3.04 | 9.81 |
| | Minimum | −2.10 | −2.10 | 0.00 | 33.00 | 0.00 | 38.60 | 4.00 | 33.00 | 42.00 | 1.51 | 3.00 |
| | Maximum | 1.74 | 1.74 | 1.00 | 83.00 | 1.00 | 68.00 | 61.00 | 65.00 | 83.00 | 14.82 | 44.00 |

*Note*. Child pretest and posttest vocabulary scores are z-scores, based on a mean of 0 and standard deviation of 1.

of children. Third, we conceptualize age composition in two ways in this investigation: The within-classroom standard deviation and the within-classroom range of ages. The two methods are strongly correlated ($r = .93$, variables 10 and 11 in Table 1), and a close investigation of these variables demonstrates the correlation between the two is heteroschedastic, strongest at the lower end of the distribution. Thus while similar to one another, these two measures are not identical.

To investigate predictive relations among the constructs of interest, hierarchical linear models (i.e., multilevel models) were fit to the data using the Mixed procedure in SAS 9.4. Models predicting children's vocabulary gains (children's posttest predicted from their pretest scores) were fit as a preliminary step to examine the variance components within and between classrooms. This model indicated that there was significant variance at the between-classroom level (17%) on children's vocabulary scores. This model also indicated that there was significant between-classroom variability on the relation of pretest with the posttest (9%), suggesting that the model providing the best fit to the data allowed each classroom to have a unique autoregressive slope. This model was used as the baseline for all future models described in this section. Prior to entering any predictors, all independent variables were z-scored (recalculated to have a mean of zero and a standard deviation of one) to allow for ease of interpretation and comparison between models.

## Classroom composition and children's vocabulary gains

The first research question sought to examine the extent to which variability in the age composition of ECE classrooms was associated with children's vocabulary growth. For this question, age composition was first represented by the within-class standard deviation of children's ages at the time of pretest. As another method of examining the first research question, and because the classrooms in the present study demonstrated far wider ranges than have been seen in previous studies, we also examined the same question conceptualizing age composition as the range of ages of children within the classroom.

### Age by composition interactions

To answer the first research question, a hierarchical linear regression model was fit to the data, predicting children's vocabulary scores from their pretest scores, the child's age, whether the child was male, and whether the child was an immigrant. We next included the classroom-factors, including the number of children per classroom and the mean age of all children in the classroom. Of primary interest, we also included both a variable representing the classroom composition, and an interaction of the child's age by composition, which allowed us to determine whether classroom composition was a more salient predictor of children's language gains for older or younger children (Bell et al., 2013). The resulting coefficients are reported in Table 2, with within-classroom standard deviation on the left, and within-classroom age range on the right. For both outcomes, none of the included classroom-specific factors were predictive of children's gains. Nor did we find evidence of an interplay between classroom composition and child's age (age standard deviation coefficient $= .0003$, $p = .982$, Age range coefficient $= -.004$, $p = .758$).

### Nonlinear contribution of age composition

Recall that the second research question sought to examine the extent to which the relation between within-classroom age composition and vocabulary gain was nonlinear. To do so, we next added the squared product of the within-classroom age composition variable. The results of this model are presented in Table 3, and show a small and nonsignificant quadratic

**Table 2.** Predicting vocabulary growth from age composition and child's age.

| | Within-Classroom Age *SD* | | | | Within-Classroom Age Range | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Standard Error | *t* | *p* | Estimate | Standard Error | *t* | *p* |
| Intercept | 0.585 | 0.019 | 31.18 | <001 | 0.585 | 0.019 | 31.22 | <001 |
| Pretest Vocabulary | 0.687 | 0.015 | 44.36 | <001 | 0.687 | 0.016 | 44.31 | <001 |
| Child Male | −0.019 | 0.015 | −1.23 | 218 | −0.019 | 0.015 | −1.24 | 214 |
| Child Bilingual | −0.196 | 0.031 | −6.27 | <001 | −0.194 | 0.031 | −6.24 | <001 |
| Child's Age | 0.053 | 0.014 | 3.82 | <001 | 0.055 | 0.014 | 3.98 | <001 |
| Children per Class | −0.031 | 0.020 | −1.51 | 131 | −0.026 | 0.022 | −1.18 | 236 |
| Class mean Age | 0.004 | 0.020 | 0.20 | 844 | 0.001 | 0.020 | 0.04 | 966 |
| Age Composition | −0.003 | 0.016 | −0.16 | 870 | −0.010 | 0.017 | −0.60 | 551 |
| Child Age*Age composition | 0.000 | 0.012 | −0.02 | 982 | −0.004 | 0.012 | −0.31 | 758 |
| Variance Components | | | | | | | | |
| Tau (Intercept) | 0.034 | 0.005 | 6.77 | <001 | 0.033 | 0.005 | 6.73 | <001 |
| Tau (Pretest Slope) | 0.018 | 0.004 | 5.00 | <001 | 0.018 | 0.004 | 5.01 | <001 |
| Sigma Squared (Error) | 0.139 | 0.004 | 33.98 | <001 | 0.139 | 0.004 | 33.98 | <001 |

**Table 3.** Predicting vocabulary growth from nonlinear classroom age composition.

| | Within-Classroom Age *SD* | | | | Within-Classroom Age Range | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Standard Error | *t* | *p* | Estimate | Standard Error | *t* | *p* |
| Intercept | 0.612 | 0.024 | 25.06 | <001 | 0.628 | 0.025 | 25.57 | <001 |
| Pretest Vocabulary | 0.686 | 0.016 | 44.26 | <001 | 0.685 | 0.016 | 44.11 | <001 |
| Child Male | −0.019 | 0.015 | −1.27 | 204 | −0.019 | 0.015 | −1.27 | 204 |
| Child Immigrant | −0.197 | 0.031 | −6.31 | <001 | −0.194 | 0.031 | −6.24 | <001 |
| Child Age | 0.054 | 0.012 | 4.50 | <001 | 0.054 | 0.012 | 4.52 | <001 |
| Class Size | −0.034 | 0.020 | −1.64 | 102 | −0.023 | 0.021 | −1.07 | 287 |
| Class Mean Age | 0.005 | 0.017 | 0.33 | 744 | 0.004 | 0.016 | 0.24 | 807 |
| Age Composition | −0.008 | 0.016 | −0.47 | 638 | −0.025 | 0.018 | −1.43 | 153 |
| Squared Age Composition | −0.026 | 0.015 | −1.67 | 095 | −0.041 | 0.016 | −2.64 | 008 |
| Variance Components | | | | | | | | |
| Tau (Intercept0 | 0.033 | 0.005 | 6.74 | <001 | 0.032 | 0.005 | 6.67 | <001 |
| Tau (Pretest Slope) | 0.018 | 0.004 | 5.03 | <001 | 0.018 | 0.004 | 5.06 | <001 |
| Sigma Squared (Error) | 0.139 | 0.004 | 34.00 | <001 | 0.139 | 0.004 | 34.01 | <001 |

relation between within-classroom standard deviation and children's language gain (estimate = −.026, $p = .095$), and a slightly larger but statistically significant effect of within classroom range (estimate = −.041, $p = .008$).

To better understand this effect, we graphed the fitted relation between class range and language gain (Figure 1). In Figure 1, the x-axis is the classroom age range in months, and the y-axis is the expected additional gain in language scores for each additional month of student age range. For example, at two standard deviations below the mean of classroom age range (6 months), the expected gain in language is .10 points, and the positive slope of the line suggests that one additional month of age range would expect to result in an additional 0.15 points of gain from fall to spring (through a centering procedure, we find that this difference is significantly different from zero, $p = .01$). At the mean (two-year age range; 24 months), we expect to see .20 points of gain. Examining the function on Figure 1, the asymptote of the function is located at 24 months, which shows that each additional month of age range does not change the predicted estimate (expected additional change per month = −.03, $p = .15$). At two standard deviations above the mean of age range (42 months), age range again shows a small relation (estimate = .05), but each additional month of age range would correspond to a negative change (−.21, $p = .01$). It is clear from this depiction the interaction manifests such that some age-range is beneficial to children, but too much age range can be detrimental to language gains.

### *Exploratory interactions*

In addition to the a-priori tests of classroom composition and children's vocabulary gain, we also tested an additional exploratory interaction effect. First, whether the quadratic effect of classroom age composition was dependent on children's age, again testing whether classroom composition is more or less

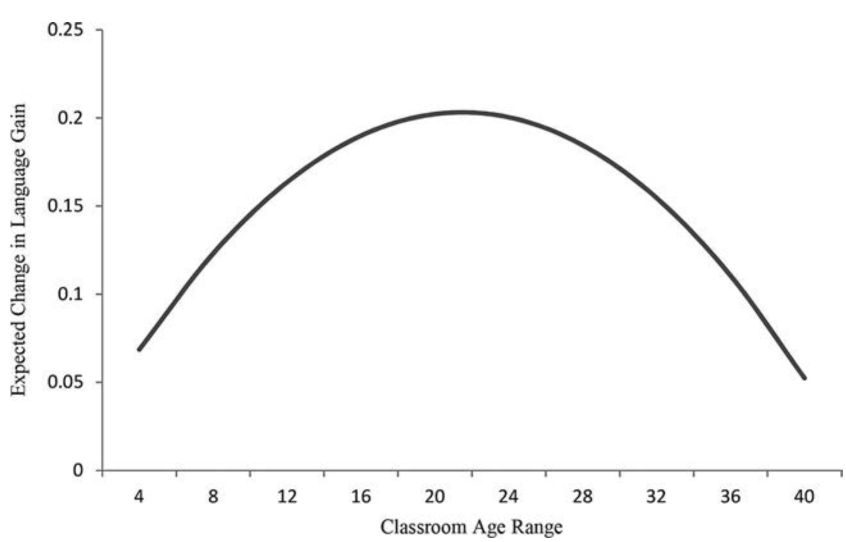


**Figure 1.** Graphical representation of results depicted in Table 3: Additional expected language gain for each observed value of within-classroom age range in months.

important for some ages than others. We did so by adding the three-way interaction term with child age interacting with the squared age composition quadratic term. In each case the standardized coefficients were near zero and nonsignificant (estimates = .002, $p = .16$ and −.001, $p = .31$, respectively). The remaining model was unchanged.

### *Curvilinear relations*

As an additional way to understand the curvilinear relationship between classroom age composition and children's outcomes, we next created a categorical representation of whether a given class was SA or MA. To do so, we examined the distribution of within-classroom age ranges (see Figure 2), and identified three equally-spaced age groupings (each 13 months) that characterized the classroom sample. Classrooms in which the age range was within 13 months between the oldest and youngest child were coded as *same age* classrooms (SA; $n = 543$); classrooms in which the age range between the oldest and youngest child was between 13 and 26 months were coded as moderately mixed age classrooms (MMA; $n = 927$); and classrooms in which the age range exceeded 26 months between the oldest and youngest child were coded as widespread mixed age classrooms (WMA; $n = 1300$). These categorical codes were entered into the same multiple regression analysis previously described, with the results presented in Table 4.

In Table 4, we see that there were no differences in the expected vocabulary gains for the SA group vs. the WMA group, but there was a small but significant difference in vocabulary gains favoring the moderate mixed age group over the wide mixed age group (estimate = .84, $p = .021$). As with the previous question, we also investigated the extent to which

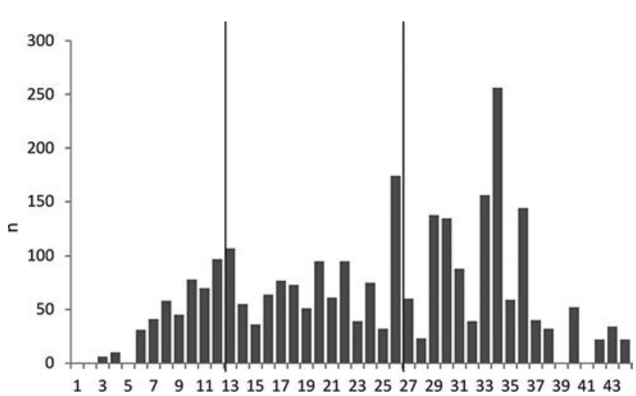


**Figure 2.** Histogram (n on Y-axis) of classroom age ranges with vertical lines representing designated cut points for the three groups of classroom age ranges (cut points at 12 and 26 months).

**Table 4.** Same age (SA) and mixed age (MA) classrooms to predict children's vocabulary gains.

| | Estimate | Standard Error | *t* | *p* |
|---|---|---|---|---|
| Intercept | 0.221 | 0.137 | 1.62 | 108 |
| Pretest | 0.700 | 0.016 | 44.33 | <001 |
| Child Male | −0.019 | 0.015 | −1.29 | 198 |
| Child Immigrant | −0.195 | 0.031 | −6.26 | <001 |
| Child Age | 0.005 | 0.001 | 4.45 | <001 |
| Class Size | −0.002 | 0.002 | −1.27 | 206 |
| Class Mean Age | 0.002 | 0.003 | 0.59 | 557 |
| SA vs WMA | 0.002 | 0.043 | −0.04 | 972 |
| SA vs MMA | 0.086 | 0.042 | 2.06 | 040 |

*Note.* SA = same age classrooms; MMA = moderate mixed age classrooms; WMA = wide mixed age classrooms.

children's ages interacted with any main effects of SA vs. MA classrooms on children's vocabulary growth, but we did not find any evidence to support these ideas (absolute value of all estimates ≤.001, $p$-values >.80).

## Discussion

Many initiatives across North America seek to increase young children's access to and participation in ECE. For instance, Quebec legislation provided publicly funded ECE programs to all 4-year-old children across the Canadian province at a very low cost to parents in 1997, with the program expanded to include all children from birth to age 5 by 2000 (Haeck, Lefebvre, & Merrigan, 2014). Importantly, the nature of policies enacted regarding ECE expansion have direct effects on the nature of classroom composition, including the age range of children served within a given program. For instance, state-funded programs that provide ECE to only 4-year-olds yield classrooms serving children who range in age relatively narrowly (i.e., all are four years of age), whereas other programs, such as the federal Head Start program, yield classrooms that serve children ranging in age more broadly, as the program supports participation of children ranging in age from 3 to 5 years of age. Given that classroom age composition is a structural characteristic of ECE classrooms that can be readily affected by public policy, it is valuable to consider whether the age range of children in a classroom may have effects on their development within that setting.

For many years, educational researchers and theorists have argued the value of MA programming for young children. This advocacy stems from prominent developmental theories, including those of Vygotsky (1978), which assert that children's development and learning is contingent on access to more capable peers. Two-year-old children who have opportunities to engage with three- and four-year-old children in ECE settings, for instance, are allowed significant enrichment opportunities in which they observe their elder peers in engage in more complex play and talk, which "stretches" their development (Winsler et al., 2002); for older children in these settings, their interactions with younger children provide them crucial opportunities to model advanced skills and serve as mentors (Whaley & Kantor, 1992).

Recent, larger-scale developmental research has raised questions regarding the developmental benefits of mixed-age groupings, however, especially for older children. For instance, Ansari and colleagues (2016) found that 4-year-olds in Head Start programs made fewer gains in language and literacy skills when their classrooms contained a relatively high density of 3-year-olds. This finding seems to argue against the value of mixing ages, at least for 4-year-olds, and would provide impetus for advancing ECE programs that serve only 4-year-old children. Guo and colleagues' (2014) study of classroom age composition for 3-, 4-, and 5-year-old children showed that mixed-age programming had positive effects on the vocabulary skills of young children, and negative effects on the vocabulary of older children. This finding complicates the issue, as it would provide impetus for advancing mixed-age ECE programming to the benefit of younger children, while recognizing that it could have negative impacts for older children.

A limitation of the classroom-age composition literature, and any interpretation of results, is that all studies have focused only at-risk, low-income children enrolled in classrooms with relatively little overall variation in classroom-age composition. To this end, such work has not had the benefit of exploring age composition effects with an unselected sample of children attending programs that vary widely in the ages of children enrolled, as we did in this study. The primary finding of this study, which fills a critical gap in the literature, is showing that classroom age composition does relate to children's development. Specifically, with respect to children's vocabulary growth, we found that mixed-age programming yielded the best outcomes for children when compared to SA programming, but only when the mixed-age programming results in classroom configurations in which the youngest and older child differ in age by at least 14 months but no more than 26 months. This result suggests that children in ECE classrooms in which classmates range in age from 36 months to 56 months of age, as an example, would show greater vocabulary gains than children whose classmates range from 36 months to 45 months of age (range is too limited) or those whose classmates range from 36 months to 66 months of age (range is too broad). Importantly, we did not find that these relations varied by children's own age. In other words, both the older children and the younger children benefited when the age range was between 14 and 26 months.

A few other findings of interest warrant note. First, study results showed that vocabulary growth for children in same-age programming, in which children are no more than 13 months apart in age, was similar to that seen for children in mixed-age classrooms in which children varied widely in their ages. In the widespread mixed-age classrooms, which enrolled the largest number of children in this study, there were more than 26 months between the oldest and young children in the classroom. While we can only speculate as to why these classroom configurations seem to equivalent influence on children's language development, we speculate that in the widespread mixed-age classrooms, children may segregate themselves into smaller groupings exhibiting age homophily. Studies consistently show that children begin to self-segregate into smaller groupings during the preschool years, with segregation related to such factors as children's gender, behavioral competence, and skill levels (Chen, Lin, Justice, & Sawyer, 2017; Daniel, Santos, Peceguina, & Vaughn, 2013). These homophily effects seem to reflect children's interest in engaging with others who are like them. While age has not typically been included in studies of social networks in preschool classrooms, children's do show homophily in their friendships based on skill levels, which often correlate closely with age (Lin, Justice, Paul, & Mashburn, 2016). It may be that in widely mixed age classrooms, children interact primarily with those similar in age, thus mirroring what occurs in same-age classrooms. If this is the case, children's language growth may be similar in both types of settings.

Second, the study results did not find there to be an interaction between children's age and the classroom composition variables of range and *SD,* as has been found in several prior studies (Ansari et al., 2016; Moller et al., 2008). Such work, conducted in the United States, indicated that having a larger range of ages in ECE settings may have negative effects on the development of older children in these classrooms, possibly because they begin "act younger" (Winsler et al., 2002). In the present study, we found no evidence that older children are negatively affected by being in mixed-age classrooms, potentially because mixed-age settings allow older children to serve as mentors of younger children; theoretically, engaging with less-skilled children provides opportunities for older children to learn and develop (Vygotsky, 1978).

Third, the study results suggest that while there are benefits to mixing ages within early childhood classrooms, there is a levelling off of this benefit when children's age range exceeds a certain point; that is, when the difference in age between the youngest and oldest child in a classrooms exceeds 24 months, the benefits of mixing ages diminishes. We can only speculate as to why this occurs, but suspect that it reflects age homophily effects that occur in early childhood classrooms, in which children tend to prefer interacting with children who are similar in age. In a classic study of 1- to 12-year-old children's interactions with others at home and in the neighborhood, researchers showed a developmental increase in children's preference to interact with others who are closer in age (Ellis, Rogoff, & Cromer, 1981). Specifically, from the toddler years into early adolescence, children showed a gradual, significant increase in interacting with peers who are closer (<1 year age difference) in age to them. It may be that in preschool settings in which children vary significantly in age, children create social networks comprising those who are relatively close in age. If this is the case, children would be clustered in same-age social networks within these classrooms, with limited opportunities to benefit from interactions with those who are significantly younger or older. Research that examines micro-level patterns in children's social networks as a product of same- and mixed-age programming is needed to confirm this possibility.

It is worthwhile to point out that the contribution of classroom age composition on children's vocabulary development was relatively small in magnitude (standardized estimate = .04). This effect can be contextualized in light of a recent integrative analysis of four large-scale datasets by Keys et al. (2013), which examined the relations between a number of ECE classroom characteristics and 6,240 3- to 5-year-old children's development in language, math, and socioemotional skills. Application of meta-analytic techniques indicated that the average relation between classroom quality and children's language outcomes was .05 (standardized coefficient), which is similar to that reported in this study. While it can take extensive efforts to improve classroom quality, including considerable professional development of teachers (Early, Maxwell, Ponder, & Pan, 2017), modifying the classroom age composition of ECE classrooms could take relatively minimal effort and yield positive benefits to children.

The findings of this work provide empirical support for the studies that have identified crucial, proximal processes in MA classrooms that would seem to yield better benefits for children. For instance, Winsler and colleagues (2002) examined children's goal-directed activities and peer interactions in MA and SA classrooms in a longitudinal study in which one ECE program transitioned from SA to MA classrooms. In the SA classrooms, the age range of children in the classrooms was about 13 months; whereas, it was 18 months in the MA classrooms. The authors showed there to be significantly more cross-age interaction in the MA classrooms as compared to SA classrooms, which presumably provides young children the opportunity to learn important skills from older children. Indeed, 3-year-olds in MA classrooms spent a greater amount of time in on-task goal-oriented activities than 3-year-olds in SA classrooms. Such work suggests that the broadened age range and therefore developmental competencies of children that occur in MA classrooms enhance the learning opportunities available within such settings. As a result, we would argue that provision of MA classrooms is an important policy lever by which to maximize the benefits of ECE expansion for children.

The results of this study should help to inform policies related to ECE access and expansion, especially efforts that maximize the potential for MA programming. For instance, rather than providing funding to programs to serve children of only a given age (e.g., 4-year-olds), expansion policies can allow ECE funds to be layered with other funds to create programs that serve children across a range of ages. In Ohio, for instance, the state's effort to expand participation in ECE focuses on providing funding to serve 4-year-olds specifically, but ECE programs can layer that funding into a broader preschool program that serves children varying widely in ages, thus allowing for MA programming. An alternative approach is for policies to support a broader range of ages, as does Early Head Start/Head Start (birth to five years) and early childhood special education (three to five years). However, while these programs allow potential for MA programming, these also represent targeted enrollment programs that can result in children being tracked into classrooms that serve only children who are poor or only children with disabilities. It is unclear that the results presented in this study, which involved a diverse, un-selected sample, would generalize effectively to targeted-enrollment MA programming.

There are limitations of this work that should be recognized. First, this study took place in the well-established universal child-care system of Denmark. Given universal participation of children in ECE in this country, important opportunities are available to learn about how to configure ECE structural characteristics in ways to maximize developmental benefits for children, as we do in this study. While we would argue that the results presented here provide important truths regarding how classroom age composition interfaces with children's development in ECE settings, the Danish context varies in important ways from other contexts, thus raising questions about generalization. Second, we examined only a single outcome of interest, namely children's vocabulary growth. We did not assess whether the age-composition effect transcended different dimensions of development. Relatedly, this assessment was completed by teachers rather than independent research staff. Examining classroom-composition effects on other measures, including those implemented by research professionals is an important direction for future research. Finally, the research presented here is correlational in nature. Experimental research that examines the developmental benefits of SA versus MA programming for children is necessary to ensure that the results observed this in and other studies are not simply a result of selection bias, given the nonrandom enrollment of children into ECE classrooms. As ECE programs expand, there should be opportunities to insert experimental approaches into initiatives to assess causal relations between classroom age composition and children's cognitive and noncognitive development, as has been possible in other major societal initiatives (Leventhal & Brooks-Gunn, 2003).

In sum, classroom age composition is a malleable feature of ECE classrooms, representing one that can be easily affected through public policies and practical initiatives. Despite the importance of this characteristic, there is little research to guide policymakers and practitioners on the topic. Our study uses a large, representative sample to shed light on topic and finds a complex, but practical finding. Children's develop vocabulary best when exposed to peers of different ages, but not to peers who are too different in age. Taking this into consideration when designing children's classroom environment in future ECE programming has the potential to increase children's vocabulary growth and, ultimately, their school readiness.

## Funding

This work was supported by the Denmark National Board of Social Services [Grant Number x] and This work was supported by the The Danish Council for Strategic research [Grant Number x].

## References

Ansari, A., Purtell, K., & Gershoff, E. (2016). Classroom age composition and the school readiness of 3-and 4-year-olds in the Head Start Program. *Psychological Science*, *27*(1), 53–63. doi:10.1177/0956797615610882

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), 1–10. doi:10.1016/j.ecresq.2012.06.002

Bleses, D., Hogen, A., Slot, P., & Justice, L. M. (2017). Relations between structural quality aspects and process quality in Danish preschools: Evidence for compensating factors. Presentation at the International Association for the Study of Children Language, Lyon France.

Bleses, D., Højen, A., Justice, L. M., Dale, P. S., Dybdal, L., Piasta, S. B., … & Haghish, E. F. (in press). The Effectiveness of a Large-Scale Language and Preliteracy Intervention: The SPELL Randomized Controlled Trial in Denmark. *Child Development*.

Bleses, D., Vach, W., Jørgensen, R. N., & Worm, T. (2010). The internal validity and acceptability of the Danish SI-3: A language screening instrument for 3-year-olds. *Journal of Speech, Language and Hearing Research*, 53, 490–507.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the abecedarian project. *Applied Developmental Science*, *6*(1), 42–57. doi:10.1207/S1532480XADS0601_05

Chen, J., Lin, T.-J., Justice, L., & Sawyer, B. (2017). The social networks of children with and without disabilities in early childhood special education classrooms. *Journal of Autism and Developmental Disorders*, 1–16.

Crosnoe, R., Purtell, K. M., Davis-Kean, P., Ansari, A., & Benner, A. D. (2016). The selection of children from low-income families into preschool. *Developmental Psychology*, *52*(4), 599. doi:10.1037/dev0000101

Daniel, J. R., Santos, A. J., Peceguina, I., & Vaughn, B. E. (2013). Exponential random graph models of preschool affiliative networks. *Social Networks*, *35*(1), 25–30. doi:10.1016/j.socnet.2012.11.002

Early, D. M., Maxwell, K. L., Ponder, B. B., & Pan, Y. (2017). Improving teacher-child interactions: A randomized control trial of making the most of classroom interactions and my teaching partner professional development models. *Early Childhood Research Quarterly*, *38*, 57–70. doi:10.1016/j.ecresq.2016.08.005

Ellis, S., Rogoff, B., & Cromer, C. C. (1981). Age segregation in children's social interactions. *Developmental Psychology*, *17*(4), 399. doi:10.1037/0012-1649.17.4.399

Esping-Andersen, G., Garfinkel, I., Han, W.-J., Magnuson, K., Wagner, S., & Waldfogel, J. (2012). Child care and school performance in Denmark and the United States. *Children and Youth Services Review*, *34*(3), 576–589. doi:10.1016/j.childyouth.2011.10.010

Guo, Y., Tompkins, V., Justice, L. M., & Petscher, Y. (2014). Classroom age composition and vocabulary development among at-risk preschoolers. *Early Education and Development*, *25*(7), 1016–1034. doi:10.1080/10409289.2014.893759

Gupta, N. D., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics*, *94*(1), 30–43. doi:10.1016/j.jpubeco.2009.10.001

Haeck, C., Lefebvre, P., & Merrigan, P. (2014). The distributional impacts of a universal school reform on mathematical achievements: A natural experiment from Canada. *Economics of Education Review*, *41*, 137–160. doi:10.1016/j.econedurev.2014.03.004

Hamre, B., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of my teaching partner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly*, *25*(3), 329–347. doi:10.1016/j.ecresq.2009.07.002

Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly*, *27*(1), 33–45. doi:10.1016/j.ecresq.2011.03.006

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*, 337–374. doi:10.1016/S0010-0285(02)00500-5

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*(9), 1722–1729. DOI: 10.1177/0956797614538978.

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, *82*(6), 1768–1777. doi:10.1111/j.1467-8624.2011.01665.x

Kamerman, S. B. (1991). Child care policies and programs: An international overview. *Journal of Social Issues*, *47*(2), 179–196. doi:10.1111/j.1540-4560.1991.tb00294.x

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., & Howes, C. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development*, *84*(4), 1171–1190. doi:10.1111/cdev.12048

Landersø, R., & Heckman, J. J. (2017). The Scandinavian fantasy: The sources of intergenerational mobility in Denmark and the US. *The Scandinavian Journal of Economics*, *119*(1), 178–230.

Lefebvre, P., Merrigan, P., & Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a Canadian natural experiment on low-fee universal child care. *Labour Economics*, *16*(5), 490–502. doi:10.1016/j.labeco.2009.03.003

Leventhal, T., & Brooks-Gunn, J. (2003). Moving to opportunity: An experimental study of neighborhood effects on mental health. *American Journal of Public Health*, *93*(9), 1576–1582.

Lillard, A., & Else-Quest, N. (2006). The early years: Evaluating Montessori education. *Science*, *313*(5795), 1893–1894. doi:10.1126/science.1132362

Lin, T.-J., Justice, L. M., Paul, N., & Mashburn, A. J. (2016). Peer interaction in rural preschool classrooms: Contributions of children's learning-related behaviors, language and literacy skills, and problem behaviors. *Early Childhood Research Quarterly*, *37*, 106–117. doi:10.1016/j.ecresq.2016.04.001

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*(4), 741. doi:10.1037/a0013099

National Center for Education Statistics. (2017). *The Condition of Education: Preschool and Kindergarten Enrollment*. Washington, DC: National Center for Education Statistics.

Organisation for Economic Co-operation and Development. (2016). *Enrolment in childcare and pre-school*. Retrieved from https://www.oecd.org/els/soc/PF3_2_Enrolment_childcare_preschool.pdf.

Sylva, K., Ereky-Stevens, K., & Aricescu, A. (2015). *Overview of European ECEC curricula and curriculum* template: CARE, Utrecht, The Netherlands.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*.

Whaley, K., and Kantor, R. (1992). Mixed-age grouping in infant-toddler childcare: Enhancing developmental processes. *Child and Youth Care Forum*, *21*, 369–384.

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*(3), 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*(3), 314–329. doi:10.1080/03004430.2011.622755

Yeomans-Maldonado, G., Justice, L. M., & Logan, J. (In Press). The mediating role of classroom quality on peer effects and language gain in pre-kindergarten ECSE classrooms. *Applied Developmental Science*.