Classroom composition and quality in early childhood education: A systematic review

Ana Lúcia Aguiar[1] (Corresponding author)

E-mail address: ana_lucia_aguiar@iscte-iul.pt

Cecília Aguiar[1]

E-mail address: cecilia.rosario.aguiar@iscte-iul.pt

[1]Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, Lisbon, Portugal

Author Note

This work was funded by Portuguese National Funds, via FCT — Fundação para a Ciência e a Tecnologia, through PhD research grant PD/BD/135346/2017, and was completed by the first author in partial fulfillment of the requirements for the degree of PhD in Psychology.

Acknowledgements

We want to thank Joana Albuquerque for her cooperation in screening and assessment of eligibility of studies.

Aguiar, A. L., & Aguiar, C. (2020). Classroom composition and quality in early childhood education: A systematic review. *Children and Youth Services Review, 115*, 105086. doi: 10.1016/j.childyouth.2020.105086

Abstract

High-quality early childhood education appears to be particularly beneficial for disadvantaged children, since it may help reduce an initial achievement gap. Yet, these children are frequently enrolled in disadvantaged classrooms with lower quality levels. Thus, classroom composition and quality may be associated, but evidence is scarce. In this review, we gathered evidence regarding classroom composition indexes and their association with observed classroom quality, reported in 25 studies that met the inclusion criteria. The majority of studies were conducted in the United States, with disadvantaged samples of children. Classroom composition indexes used were mainly calculations of the percentage, proportion, and average/mean of a particular type of characteristic at the classroom level, that generally captured classroom homogeneity. Most studies focused on minority and socioeconomic status. ECERS and CLASS were the most frequently used standardized observation measures of classroom quality. Evidence suggests that in classrooms with a high concentration of children with minority status and from low income families, quality tends to be lower, particularly on the CLASS emotional and instructional support domains. Additional research, particularly outside the USA, focused primarily on the association between different types of classroom composition and ECE quality is warranted.

Keywords: Systematic review, Early childhood education, Classroom composition, Classroom quality

1. Introduction

School systems of Western countries are serving an increasingly diverse student population (Vervaet, Van Houtte, & Stevens, 2018). Simultaneously, access to early childhood education (ECE) programs has been expanded (Vervaet et al., 2018). As a result, many young children, from diverse backgrounds, spend a considerable proportion of their days in ECE classrooms, where they experience interactions that shape their development (Pianta & Hamre, 2009).

High-quality ECE typically provides more opportunities for children to establish stimulating, warm, and supportive interactions (Mashburn et al., 2008; Votruba-Drzal, Coley, & Chase-Lansdale, 2004) with teachers and peers (Purtell & Ansari, 2018), and experience adequate and planned instruction (Pianta et al., 2009). Attending high-quality classrooms in ECE has been associated with better outcomes for children in terms of cognitive, linguistic (e.g., Pianta & Hamre, 2009), social, and behavioral development (e.g., Mashburn et al., 2008). There is also evidence suggesting that the benefits of attending high-quality classrooms may be long lasting and still visible in elementary school (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2011). Moreover, placement in special education and grade retention seem to be lower and high-school graduation rates seem to be higher among students who were enrolled in high-quality ECE programs (McCoy et al., 2017).

Attending high-quality classrooms may function as a protective factor for socially disadvantaged children, by providing positive experiences (Clements, Reynolds, & Hickey, 2004), that contribute to the development of self-regulation skills and pro-social behaviors (Sylva et al., 2011). Thus, high-quality ECE may have the potential to reduce initial achievement gaps (Bridges et al., 2004). However, there is evidence that these children are often enrolled in ECE classrooms with a high concentration of other disadvantaged children

(Reid & Kagan, 2015), and in classrooms with lower quality (Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008). This suggests that there may be an association between the composition of the classroom and ECE quality levels. However, research on how variations in classroom composition are associated with variations in classroom quality is still relatively underexplored and dispersed. Hence, with this review we intend to gather and systematize findings reported in the ECE literature about the associations between classroom composition and observed classroom quality.

1.1. Classroom composition as a structure feature and a predictor of process quality

We examined the association between classroom composition and classroom quality through the lens of the (bio)ecological theory, which postulates that child development is shaped by interaction patterns, that evolve over time, such as those that occur in ECE settings between children and their teachers (Bronfenbrenner & Morris, 2006) as well as the transactional model (Sameroff, 2009), that emphasizes the bidirectional and interdependent effects of the developing child's experience and his/her social environment (Sameroff, 2009). Thus, when applied to ECE, we consider that, during their interactions, children (individually and as a group) and teachers influence each other's behaviors. This means that children's characteristics, measured at the classroom level, and behaviors may affect teachers' responses and vice-versa (DiLalla & Mullineaux, 2008), with an impact on quality (Buyse et al., 2008).

Classroom quality can be defined as encompassing: (i) structural features, which refer to regulable characteristics (Slot, Leseman, Verhagen, & Mulder, 2015), such as class size, children-to-teacher ratio, and teacher education (Howes et al., 2008); and (ii) process quality, which relates to children's daily experiences in the classroom context, including their interactions with teachers and peers and their engagement in school activities (Howes et al., 2008; Phillipsen, Burchinal, Howes, & Cryer, 1997). In ECE, process quality seems to be a

stronger and more direct predictor of children's linguistic, cognitive, and social development than structural features, which seem to influence children´s development indirectly, through process quality (Friedman & Amadeo, 1999; Howes et al., 2008). Improving classroom process quality has therefore been the main goal of quality improvement programs (Pianta et al., 2014).

Since structural features tend to be easier to regulate (Cryer, Tietze, Burchinal, Leal, & Palacios, 1999), a growing body of research has focused on how these features impact process quality and how they can be used to promote positive change (Cryer et al., 1999). However, the evidence base about the association between structural features and process quality has been relatively inconsistent (Slot et al., 2015). Like other classroom structural features involving group characteristics, such as class size and children-to-teacher ratio, we propose that classroom composition, which encompasses the aggregated personal and family characteristics of the children in each classroom (Cueto, Léon, & Miranda, 2016; Jones, 2016), should also be examined as structural feature of ECE classrooms potentially subject to regulation.

Most literature about classroom composition effects in ECE settings has focused on the association with children´s achievement and has used children´s characteristics such as ability (e.g., Steenbergen-Hu, Makel, & Olszewski-Kubilius, 2016), age (e.g., Bell, Greenfield, & Bulotsky-Sheare, 2013; Guo, Tompkins, Justice, & Petscher, 2014; Purtell & Ansari, 2018), gender (e.g., Gottfried & Graves, 2013; Whitmore, 2005), ethnicity/race (e.g., Denton, Germino-Hausken, & West, 2000), and SES (e.g., Li et al., 2016; Reid & Ready, 2013; Weiland & Yoshikawa, 2014), to compute indexes of classroom composition. Fewer studies have focused on the associations between classroom composition and classroom quality in ECE settings.

Although scarce, there is evidence in the ECE literature supporting the idea that classroom composition may be associated with ECE quality levels. For example, some evidence suggests that children from disadvantaged backgrounds (e.g., Burchinal, Peisner-Feinberg, Pianta, & Howes, 2002; den Brok, van Tartwijk, Wubbels, & Veldman, 2010; Raver et al., 2009) can be at higher risk of developing more conflictual and distant interactions with their teachers (Saft & Pianta, 2001), when compared with their peers, as a consequence of contextual factors hindering their social and behavioral development (Raver et al., 2009). Thus, a high concentration of socially disadvantaged children in the classroom and, therefore, at higher risk of exhibiting behavioral problems can be associated with lower quality (Buyse et al., 2008). There is similar evidence for boys (e.g., Baker, 2006; Hamre & Pianta, 2001) and younger children (e.g., Shaw, Lacourse, & Nagin, 2005). In this sense, classroom composition can be an important structural feature of ECE (Reid & Ready, 2013), particularly when considering the impact of economic, sociocultural, and ethnic diversity or homogeneity on teacher-child interactions (Dronkers & Van der Velden, 2013).

1.1.2. Classroom composition indexes. Classroom composition can be analyzed to ascertain levels of heterogeneity or homogeneity. Heterogeneity or diversity is determined by the amount of differences on a given characteristic among members within a social group/community, while homogeneity is related with sameness on a given characteristic (Harrison & Sin, 2006; Solanas, Selvam, Navarro, & Leiva, 2012).

There are indexes created specifically to determine within-group distribution of differences, such as the mean Euclidean distance, the standard deviation, Teachman's index, Blau´s index, the coefficient of variation, and the Gini coefficient of concentration (see, Solanas et al., 2012). These indexes are used to ascertain levels of diversity, within three parameters: separation (i.e., differences in position or values), variety (i.e., differences in

categorical values), and disparity (i.e., differences in concentration of resources) (see Harrison & Klein, 2007).

To our knowledge, thus far, it is not common to find such conceptualizations of diversity (see Harrison & Klein, 2007) nor the calculation of such composition indexes in the education literature. In studies conducted in ECE settings, as well as in other education levels, the most common practice seems to be the calculation of the percentage/proportion and the average/share of children with a given characteristic in classroom (Veerman, van de Werfhorst, & Dronkers, 2013). A few exceptions can be found in studies, mostly at the primary and secondary levels of education, that used adaptations of the Hirschman-Herfindahl Index (Hirschman, 1964; Dronkers & van der Velden, 2012), first used in the economy literature, and Simpson's diversity index (Simpson, 1949; see Graham, 2004), first used in the ethology literature, to ascertain the school/classroom ethnic and sociocultural compositions. Both indexes vary between 0 (minimum diversity) and 1 (high diversity), but while the Herfindahl Index does not consider multiple possible categories within a given characteristic (e.g., distinguish between particular countries of origin [Stolle, Soroka, & Johnston, 2008]) (Schaeffer, 2013), Simpson's diversity index considers both the number of categories and the share of each category within a group (Graham, 2004).

This distinction between diversity and share is of importance since, in the education literature, results from average/share calculations are sometimes presented as being indicative of school/classroom diversity on a given characteristic (Veerman et al., 2013). Despite a possible overlap (Veerman et al., 2013), there are fundamental conceptual differences since the average/share involves the proportion of children within a group who share a particular characteristic (e.g., migration background), being a potential indicator of homogeneity (e.g., high proportion of migrant children in class from the same ethnic group), while diversity addresses the variety of a certain characteristic within the group (e.g., number and size of

distinct ethnic groups) (Veerman, 2014). Therefore, there may be a disconnection between how diversity has been conceptualized and its operationalization, which may impact the validity of findings (see Harrison & Klein, 2007).

Hence, gathering data about how group composition has been measured in education and, particularly, in the ECE literature, can contribute to further clarification on how variations in classroom composition in ECE may be associated with classroom quality (Steinberg & Garret, 2016). Furthermore, it may help inform future research with guidelines for an integrated conceptualization and operationalization of classroom composition, and also for avoiding key pitfalls, so knowledge about classroom composition effects can be enhanced.

1.1.3. Assessing classroom process quality. Classroom quality can be measured with a multitude of assessment tools, with emphasis on standardized observational measures. Observation measures typically focus on global quality, that is, on both the physical aspects of the environment and the social interactions in the classroom. However, there are also process quality measures, which focus primarily on teacher-child interactions and content specific measures, that focus on instructional quality within specific content areas (Burchinal, 2010). A description of standardized observation measures of classroom quality typically used in the literature is presented in Table 1. No single standardized observation measure covers all aspects of children's experiences in the classroom (Bryant, 2010), but most have demonstrated good reliability (Burchinal, 2010) and are believed to produce more valid assessments of teachers' effectiveness (Goldring et al., 2015), than non-standardized measures.

Some studies that focused on the association between classroom structural features and standardized observation measures of process quality reported a significant association, for example, between classroom quality and teacher's education and training (e.g., Burchinal,

Cryer, Clifford, & Howes, 2002), teacher-child ratios, and group size (e.g., Cryer et al., 1999). However, evidence is mixed (see Resnick, 2010).

1.2. This Review

High-quality ECE has been consistently linked to children's positive developmental outcomes (e.g., Burchinal, Kainz, & Cai, 2011; Camilli, Vargas, Ryan, & Barnett, 2010; Pianta et al., 2009), with some studies suggesting that this association may be more significant for particular groups of children, specifically, for those in social and economic disadvantage (e.g., Zaslow et al., 2010). Further, child characteristics and classroom composition may influence teacher behavior and classroom quality, in an apparent two-way interaction (DiLalla & Mullineaux, 2008).

Existing reviews and meta-analysis addressing classroom composition effects have focused on its association with student outcomes at different school levels. We identified a review about the effects of within-class grouping in primary and secondary schools (Kutnick et al., 2005); another about between-class ability grouping (i.e., tracking/streaming), in grades 6 to 12 (Belfi, Goos, De Fraine, & Van Damme, 2012); and two meta-analyses on the relationship between peer group composition and students' achievement in primary and secondary schools (Van Ewijk & Sleegers, 2010a,b).

Despite the potential practical and research implications, to our knowledge, there are no other reviews addressing the associations between classroom composition and classroom quality in ECE. Therefore, in this systematic review, we aimed to identify classroom composition indexes used in the ECE literature and to examine the associations between classroom composition in ECE and observed classroom quality. By systematically gathering and examining the current evidence base on classroom composition in ECE, we aimed to inform future research on existing gaps in knowledge regarding the associations between

structural features of ECE classrooms and process quality and help inform decision-making processes regarding the organization of classrooms.

2. Method

2.1. Eligibility Criteria

Inclusion and exclusion criteria were defined using the SPIDER tool (Sample, Phenomenon of Interest, Design, Evaluation, and Research type; Cooke, Smith, & Booth, 2012). To be eligible for qualitative synthesis, studies had to meet the following criteria:

i. Sample: Focus on teachers of children aged between 3 and 5/6 years old, enrolled in ECE center-based programs (i.e., preschool or kindergarten).

ii. Phenomenon of Interest: Classroom composition, including ethnic, racial, sociocultural, socioeconomical, and linguistic heterogeneity/diversity or homogeneity (e.g., proportion/percentage/ratio of children from minority groups or children in disadvantaged/at-risk).

iii. Design: Any type of study (e.g., correlational, longitudinal, experimental) providing empirical evidence on observed classroom quality.

iv. Evaluation: Standardized observations of classroom processes, specifically, of teacher-child relationship/interactions, of teacher-child conflict, of teacher-child proximity, and/or of teacher practices as outcomes, measured systematically and translated into quantitative data. If testing the implementation of specific interventions, studied needed to provide pre-treatment scores and/or scores from control/ “business as usual” /no intervention groups.

v. Research type: Any type of empirical research using standardized observation measures, both global and content specific, of classroom quality with a quantitative approach to data analyses.

Studies were excluded if the sample consisted of teachers serving in other types of early child care services (e.g., family-centered care, residential care facilities), caregivers other than teachers (e.g., parents), and teachers of younger (infants, toddlers) or older children (from primary school onwards). The focus on children aged between 3 and 6 was related with the goals of the broader project in which this review is included, and also because ECE coverage and attendance rates are considerably higher for preschool-aged children (European Commission/EACEA/Eurydice, 2019). Furthermore, studies were excluded if composition indexes were provided only at the school level (e.g., school ethnic composition, school socioeconomic composition). We decided to focus on the classroom level so that potential variations in quality between classrooms within the same centers would not be overlooked (e.g., Karoly, Zellman, & Perlman, 2013) and also because process quality is typically measured and reported at the classroom level. Systematic reviews, meta-analyses, and qualitative studies were not included. Studies with naturalistic observations of classroom quality with a qualitative approach to data analyses, studies that employed non-standardized observation measures (despite adopting a quantitative approach to data analyses), studies using teachers' self-reported interactions with children and pedagogical practices, and studies reporting only post-treatment scores (if testing the implementation of specific interventions), were excluded. Only studies written in English and Portuguese were considered. We did not define restrictions regarding scientific discipline or year of publication.

2.2. Search Procedures

An electronic systematic search of the literature was conducted to identify all potential eligible, published and unpublished, empirical studies providing data on the association between classroom composition and classroom quality in ECE. EBSCO databases such as Academic Search Complete, ERIC, PsycARTICLES, PsycINFO, Psychology and Behavioral Sciences Collection, as well as Scopus and Web of Science were searched. To ensure an

appropriate balance between sensitivity and specificity (Hempel, Xenakis, & Danz, 2016), we limited our search to studies that contained the selected search terms in the title, abstract, key terms, and/or topic. Three search strings, regarding the population, the phenomenon of interest, and the method of evaluation, were developed and combined. Each string was composed of a vast array of search terms, representing both more general and more specific concepts, to capture the multiplicity of existing classroom composition indexes and of observation measures of classroom quality used in ECE contexts, while narrowing search results. Examples of search terms included in each string follow: (a) "early childhood education and care" OR "center-based child care" OR preschool* OR "3-to-5-year* old*" AND teacher* OR educator* OR professional* AND (b) "class* composition" OR "class* characteristics" OR "class* heterogeneity" OR "group homogeneity" AND (c) "class* observations" OR "observed interaction*" OR "observed practice*" OR "process quality". For a full scope on the search strategy see the Appendix.

To guarantee the identification of records that might have been missed on the initial electronic database search, a hand-search of reference lists from already known empirical and theoretical literature was conducted, as well as a legacy search, based on the reference lists of all eligible studies.

2.3. Screening and Study Selection

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (Liberati et al., 2009), a sequential examination process, illustrated in Figure 1, was conducted, in order to select studies of interest. The initial electronic database search generated 1095 unique records, after duplicate entries were eliminated (n = 2335). Additionally, 21 records were identified through manual search. Peer-reviewed articles, book chapters, dissertations, theses and reports, were retrieved by October 5, 2018. Subsequently, a pair of independent raters conducted title and abstract screenings of these records, using

Rayyan, a web and mobile app (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016), reaching 88% agreement at this phase. Conflicting decisions in the exclusion process (*n* = 138) were resolved by a third rater. Most disagreements were on studies not using observation methods of classroom quality in ECE. One hundred and twenty studies qualified for the next phase, a full-text examination, after meeting at least one of the inclusion criteria. Of those, seven could not be retrieved and were excluded without examination. The remaining 113 studies were reviewed in full. Inter-rater agreement for final selection based on full text analysis was 84%. Conflicting decisions in the exclusion process (*n* = 21) were, again, resolved by a third rater. Disagreements were mostly related to studies that did not address directly the association between classroom composition and classroom quality and to studies that focused on children's individual characteristics and not on group level characteristics. Thirty-one studies that provided data on the association between classroom composition and observed classroom quality were identified. However, of those, nine (29%) were excluded because they used non-standardized observation measures of classroom quality. Twenty-five peer-reviewed articles, 15 resulting from electronic database search and 10 from hand-search, were deemed eligible and were selected for qualitative syntheses.

2.4. Coding and Syntheses

For qualitative analysis, the first author extracted from all eligible studies information on: (i) the theoretical framework, (ii) the sample (e.g., sample size, age range), (iii) the study design, (iv) the classroom composition index, (v) the observation measure used to assess classroom quality, (vi) the results on the associations between classroom composition and classroom quality, and (vii) covariates. Studies were categorized by the type of classroom composition index used and are presented in the results section accordingly. Studies that report data on the association between more than one classroom composition index and classroom quality were allocated to all adequate categories.

3. Results

3.1. Description of Studies

Information extracted from selected studies is presented in Tables 2 to 6. The level of detail in the information presented in the tables matches that of the included studies. For each study, we presented the terminology used by the respective authors regarding sample characteristics used to compute classroom composition indexes and covariates, so that data extraction was as truthful as possible.

Most studies (*n* =18, 72%) were published after 2010 and only one (Sontag, 1997) was published before 2000. Five studies were conducted in Europe (one in Denmark [Slot, Bleses, Justice, Markussen-Brown, & Højen, 2018], one in Finland [Pakarinen et al., 2010], two in Germany [Bihler et al., 2018; Kuger, Kluczniok, Kaplan, & Rossbach, 2016], one in the Netherlands [Broekhuizen, Slot, van Aken, & Dubas, 2017]) and the remaining 20 were conducted in the USA[1].

3.1.1. Theoretical framework. In several studies (*n*= 11, 44%), the theoretical framework was not clearly stated. Among those which made it explicit (*n*= 14, 56%), around half were framed by the ecological theory. The remaining studies were grounded on different theories and conceptual frameworks, including sociocultural (Vygotsky, 1978) and social-learning theory (Bandura, 1986); input effects on bilingual language development (Unsworth, 2016); transactional model of coercive cycles of adult–child conflict (Snyder, Cramer, Afrank, & Patterson, 2005) and of stress and coping (Lazarus, 1991); culturally responsive teaching (Gay, 2000); and social–interactionist theories of language acquisition (e.g., Baumwell, Tamis-LeMonda, & Bornstein, 1997), among others.

3.1.2. Sample characteristics. More than half of the studies (*n* = 14, 56%) relied on data from large-scale studies, such as the National Center for Early Development and

[1] This information concerns the countries where data were collected.

Learning's (NCEDL) Multi-State Study of Pre-Kindergarten, the Study of State-Wide Early Education Programs (SWEEP) (Downer et al., 2012; LoCasale-Crouch et al., 2007; Reid & Ready, 2013; Sanders & Downer, 2012; Valentino, 2018), and the Early Childhood Longitudinal Study, Birth Cohort (Fram & Kim, 2012; Iruka & Morgan, 2014). Of those 14 studies, six relied on data from both the NCEDL Multi-State Study and SWEEP. Almost all studies included classrooms from state-funded programs, such as Pre-K and Head Start, that served a considerable percentage of children at-risk, due to social/economic constraints. The majority of studies (*n* = 20, 80%) were conducted in preschools; two studies involved kindergarten classrooms (Fram & Kim, 2012; Pakarinen et al., 2010); three studies (Stipek, 2004; Maxwell, McWilliam, Hemmeter, Ault, & Schusterb, 2001; Debnam, Pas, Bottiani, & Cash, 2015) were conducted in K-2nd, K-3rd, and K-8th classrooms, respectively.

3.1.3. Study design. Most studies (*n* = 18, 72%) were cross-sectional and all, except for two (Sawyer et al., 2016; Slot et al., 2018), were correlational. In terms of number of data collection points, two longitudinal studies (Ansari & Pianta, 2018; Kuger et al., 2016) assessed simultaneously classroom composition and quality, at least in two distinct moments, and three other studies (Dotterer et al., 2013; Friedman-Krauss et al., 2014; Sontag, 1997) assessed classroom quality in more than one moment, over time, but assessed classroom composition only one time. Two studies were part of randomized control trials (Sawyer et al., 2016; Slot et al., 2018).

3.1.4. Standardized observation measures of classroom quality. Eight standardized observation measures of classroom quality in ECE were used. Details on these measures are presented in Table 1. Two global quality measures, one content specific measure, and five process quality measures were extracted. Most measures include a set of items which can be scored into specific quality factors or averaged into a global score (Bryant, 2010). The Early Childhood Environment Rating Scale (ECERS, ECERS-R; Harms & Clifford, 1980; Harms,

Clifford, & Cryer, 1998, 2008), a global quality measure, and the Classroom Assessment Scoring System (CLASS, CLASS PRE-K; Pianta, LaParo, & Hamre, 2008) were the most frequently used standardized observation measures of classroom quality (n = 6, 24% and n = 17, 68%, respectively). Four studies used both the ECERS and CLASS (Dotterer et al., 2013; Pianta et al., 2005; Reid & Ready, 2013; Valentino, 2018). The ECERS was also used in combination with the Caregiver Interaction Scale (CIS; Arnett, 1989) in one study to create three quality profiles (Iruka & Morgan, 2014).

All associations between classroom composition and the ECERS were significant. Most studies reported scores on domains of language/interactions and provisions for learning separately. Regarding CLASS scores, results were not so consistent. Most studies reported scores on domains of quality, separately, even though classroom organization was not assessed as frequently as emotional and instructional support. Few associations between the CLASS total score and classroom composition were significant (four out of 11). Emotional support was assessed 23 times in association with classroom composition and 15 of those associations were significant. Instructional support was assessed 19 times and 12 associations were significant. Five of the nine associations tested between classroom organization and composition were significant.

3.2. Classroom Level Characteristics and Classroom Composition Indexes

Five types of children´s characteristics, measured at the classroom level, were used to compute classroom composition indexes: ability (n = 4, 16%), age (n = 5, 20%), gender (n = 3, 12%), minority status (n = 17, 68%), and SES (n = 11, 44%). Twelve out of the 25 studies included two types of characteristics (48%) and one included three. Of these, all except for one, which focused simultaneously on age and gender, focused on minority status and one other index. The most common overlap was between minority status and SES (n = 7, 28%).

Under minority status, we coded all studies that operationalized classroom composition indexes based on the concentration of children identified as belonging to a particular ethnicity or race, as having an immigration background, and as being a dual-language learner (DLL). First, we found that the aforesaid characteristics frequently coexisted, that is, children often accumulated some of these characteristics (e.g., Hispanic/Latino children from immigrant families attending ECE in the USA generally learn both the Spanish and English languages), so aggregation was a possibility. Second, in the USA education system it is common to gather information on children´s ethnic and racial identifications separately. However, ethnicity tends to be related almost exclusively with being or not part of the Hispanic/Latino culture, while race is associated with children´s country of origin/ancestry, such as being American Indian or Alaska Native, Asian, Black or African American, a Native Hawaiian or other Pacific islander, or White (National Forum on Education Statistics, 2008). Therefore, the distinction between the two concepts can become blurry. Both ethnicity and race are socially constructed concepts (Markus, 2008), often used to distinguish between social groups (Johnson-Bailey & Drake-Clark, 2010). Thus, in our view, independently of the terminology used, these concepts are primarily related with perceptions of belongingness to a given social group that often represents having a minority status (Khanna & Harris, 2009).

In 24 of the studies (96%), classroom composition indexes were calculated based on the percentage/proportion and the average/mean of children with a given characteristic in the classroom and, therefore, measured mostly classroom homogeneity. One study (Ansari & Pianta, 2018) used Simpson's Diversity Index (1949) to calculate classroom age diversity. More detail on how composition indexes were computed in each study, for each characteristic of children in the classroom is presented next.

3.3. Associations Between Classroom Composition and Observed Classroom Quality

Only four of the studies (16%) defined specific research hypotheses regarding the potential direction of the association between classroom composition and quality (Ansari & Pianta, 2018; Sawyer et al., 2016; Slot et al., 2018; Stipek, 2004). The remaining studies though providing data on the association between classroom composition and quality, focused primarily on the association between classroom composition and children's developmental outcomes. A synthesis of the main findings regarding the association between classroom composition indexes and process quality is presented in Table 7. The magnitude of effects was generally small.

3.3.1. Classroom ability composition and classroom quality. Four studies defined ability in terms of the presence or absence of disabilities in children or the percentage of children with IEPs (see Table 2). One longitudinal study reported that teachers in inclusive classrooms (i.e., including both children with and without disabilities) used significantly more disapprovals of children's behavior compared with teachers in segregated classrooms (i.e., all children with disabilities), based on assessments with the ESCAPE (Sontag, 1997). A cross-sectional study found a positive association between a higher number of children with disabilities in the classroom and the quality of literacy focus, but no association was found with language modeling, two scales included in the CLASS (Justice et al., 2008). Two other cross-sectional studies found no association between the number of children with disabilities in the classroom and the APEEC (Hemmeter, Maxwell, Ault, & Schuster, 2001) and quality profiles defined by a combination of the ECERS and the CIS (Iruka & Morgan, 2014).

3.3.2. Classroom age composition and classroom quality. Out of five studies on the association between classroom age composition and classroom quality (see Table 3), two found significant associations. One longitudinal study reported a negative association between higher age diversity and the CLASS emotional support, classroom organization, and instructional support domains, compared with less diverse classrooms in terms of children's

age (Ansari & Pianta, 2018). Furthermore, this study reported a decrease in classroom organization and emotional support scores in year two following an increase in age diversity. Another study found that a higher mean age was positively associated with the ECERS total score (Kuger et al., 2016). Conversely, three studies, two cross-sectional (Pakarinen et al., 2010; Purtell & Ansari, 2018) and one randomized control trial (Slot et al., 2018), found no association between the proportion of children in the classroom within a determined age range or the classroom mean age and CLASS scores.

3.3.4. Gender composition and classroom quality. The three studies that focused on the associations between gender composition and classroom quality, reported no significant associations between the percentage of boys or girls in the classroom and the CLASS emotional and behavioral support scores (Toddler version; Broekhuizen et al., 2017), the CLASS emotional climate domain (Friedman-Krauss et al., 2014), and the ASSIST scores (Debnam et al., 2015) (see Table 4).

3.3.3. Classroom minority status and classroom quality.

Seven cross-sectional studies found no association between the concentration of Hispanic/Latino children learning both the English and Spanish language in the classroom and ELLCO-DLL scores (Sawyer et al., 2016), CIS scores (Fram & Kim, 2012), and CLASS scores (Bassok & Galdo, 2016; Bihler et al., 2018; Downer et al., 2012; Sanders & Downer, 2012; Valentino, 2018). One more study reported no association between the percentage of children with low proficiency in the English language in the classroom and the quality of literacy focus and of language modeling, two scales included in the CLASS (Justice et al., 2008). Another cross-sectional study found no association between the percentage of White children in the classroom and the use of culturally responsive teaching, assessed with the ASSIST (Debnam et al., 2015).

Conversely, two cross-sectional studies reported that a high concentration of Hispanic/Latino children in the classroom, compared with a high concentration of majority children, was associated with lower global quality in the ECERS total score, language and interactions, and provisions for learning (Valentino, 2018) and with the use of fewer constructivist teaching strategies measured with the ECCOM (Stipek, 2004). Similar results were reported in these two studies in classrooms with a high concentration of Black/African-American children, again, in comparison with classrooms with a higher concentration of majority children. This type of composition was also associated with lower quality in the CLASS total score and in emotional and instructional supports in classrooms with higher (Bassok & Galdo, 2016), and with lower CIS scores (Fram & Kim, 2012). One longitudinal study (Friedman-Kraus et al., 2014), also reported a similar association between the percentage of Black children and the CLASS emotional climate scores. Note, however, that the longitudinal study by Dotterer et al. (2014) reported lower quality in the ECERS language and interactions, and provisions for learning subscales, and the CLASS instructional support domain in universal programs with higher percentages of White children in the classroom.

Four more studies reported significant associations: in one study conducted in the USA both classrooms with higher quality and lower quality, measured with a combination of ECERS and CIS, had a higher percentage of non-English-speaking children compared with classrooms with medium quality (Iruka & Morgan, 2014); in German ECE settings, a proportion of 100% migrant children (with low proficiency in the German language) in the classroom was negatively associated with the ECERS total score, that was about .75 points lower than in classrooms with a proportion of 0%; also, from year 1 to year 2 an increased proportion of children from migrant families was associated with a decrease in ECERS scores (Kuger et al., 2016); in a Danish study, a pre-intervention assessment revealed that a higher proportion of non-Danish children in the classroom was associated with lower quality scores

in all of the CLASS domains, particularly with Classroom Organization (Slot el at., 2018); lastly, one study conducted in the Netherlands reported lower emotional and behavioral support in the CLASS in classrooms with a higher proportion of non-Dutch children (Broekhuizen et al., 2017) (see Table 5).

3.3.5. Socioeconomic composition and classroom quality. Under SES we included studies that operationalized this index based on indicators such as family income, maternal education, and average of family income and maternal education. Out of 11 studies focusing on socioeconomic composition and classroom quality, nine reported significant associations (see Table 6). Two studies found no association between the percentage of children living in poverty in the classroom and the CLASS total score (Bihler et al., 2018; Phillips et al., 2009). Two more studies found no association with the CLASS emotional support (Dotterer et al., 2013) and the CLASS instructional support (Reid & Ready, 2013).

Conversely, four studies, three cross-sectional and one longitudinal, reported a negative association between a higher concentration (i.e., percentage or proportion) of children living in poverty in the classroom and the CLASS total score (LoCasale-Crouch et al., 2007; Sanders & Downer, 2012; Valentino, 2018) and emotional and instructional support scores (Bassok & Galdo, 2016; Pianta et al., 2005; Valentino, 2018). Of these studies, two were conducted with subsamples from the same larger-scale studies. A negative association was also found with the ECERS total score (Valentino, 2018), the ECERS interactions and provisions for learning (Pianta et al., 2005; Valentino, 2018) and the use of constructivist teaching strategies, measured with the ECCOM (Stipek, 2004). Conversely, one longitudinal study reported that in classrooms from targeted programs, with more children living in poverty, scores in the ECERS interactions and provisions for learning scales and the CLASS instructional support were higher, compared with classrooms from universal programs with a lower percentage of economically disadvantaged children (Dotterer et al., 2013).

Two cross-sectional studies, conducted with subsamples from the same larger-scale projects, focused on the association between classroom mean level of maternal education and classroom quality and reported that in classrooms with higher mean levels of maternal education, the CLASS total score was higher (LoCasale-Crouch et al., 2007; Sanders & Downer, 2012). In classrooms with a higher average of family income and maternal education, the ECERS total score and the CLASS emotional support score were also higher **(**Reid & Ready, 2013).

3.3.6. Covariates. There was wide variation in the number and type of covariates considered in the association between classroom composition and process quality. In more than half of the studies, this association was not assessed considering the presence of covariates. In the remaining studies the number of covariates considered varied between two (Debnam et al., 2015; Maxwell et al., 2001; & Stipek, 2004) and 21 (Purtell & Ansari, 2018). Covariates were related with program, teacher, classroom and child characteristics. The most common covariates were associated with teacher characteristics, mainly, teacher education, years of experience, and training; and with classroom characteristics, such as composition, size, and teacher-child ratio. There were no substantial differences in terms of significant associations between classroom composition and quality reported in studies that considered covariates (seven out of 12 reported at least one significant association) and those that did not (nine out of 13 reported at least one significant association). Since we did not formally conduct a meta-analysis, we can only mention that the size of effects appeared to be, in general, small.

4. Discussion

We set out to identify the types of classroom composition indexes used in the ECE literature and their association with observed classroom quality, based on the premise that the characteristics of the children in the classroom shape their experiences (e.g., Pianta et al.,

2005). Even though there is a growing interest in classroom composition effects, particularly over the last two decades, most screened studies focused on the association between classroom composition and children's outcomes and only a small number was eligible for this review. Thus, more empirical research is needed to inform policies and decision-making processes, regarding the organization of classrooms in center-based ECE.

4.1. Theoretical framework

The lack of a clearly stated framework in many studies does not mean that these studies do not have a valid rationale, built upon a substantive theory, or a conceptual framework (Camp, 2001). Nonetheless, defining a clear theoretical framework helps in the definition of the research design, contributes with new knowledge to a specific theoretical community, and clarifies the assumptions underlying the problem under investigation to readers (Camp, 2001). Since studies varied substantially in their research aims and designs, it is not possible to identify contributions to one specific theoretical string or to fully integrate findings reported in this review. Nevertheless, ecological frameworks seem to be salient in the empirical research reviewed, suggesting an acknowledgement of the role of classroom composition as an important feature of the classroom microsystem (Bronfenbrenner & Morris, 2006).

4.2. Study design

Given that this review was framed by the transactional model (Sameroff, 2009) and (bio)ecological theory (Bronfenbrenner & Morris, 2006), it was of interest to analyze if eligible studies considered the passage of time in the association between classroom composition and quality, based on the premise that interactional processes between children and teachers can change over time, as a function of classroom composition. Only two studies (Ansari & Pianta, 2018; Kuger et al., 2016) measured both classroom composition and quality over time and reported noteworthy findings. Both studies reported differences from

year to year in classroom quality associated with variations in classroom composition, regarding age diversity (Ansari & Pianta, 2018) and concentration of children with minority status (Kuger et al., 2016). These results are indicative of both the importance of investigating how classroom composition may be associated with the quality of education children receive (e.g., Snell, Hindman, & Belsky, 2015) and of doing so over time (Ansari & Pianta, 2018). Multiple assessments over the year(s) can help identify what and how any type of change in classroom composition may constitute an additional challenge and hinder teachers' conditions to establish good quality interactions with children, as well as the strategies and supports needed to help teachers overcome them (Ansari & Pianta, 2018).

4.3. Observation measures of classroom quality

Even though structural features have been considered preconditions of process quality (e.g., Philips et al., 2000; Pianta et al., 2005), the evidence base about the association between structural features and process quality has been inconsistent (Slot et al., 2015). Quality scores on the ECERS and the CLASS were those with more associations with classroom composition (see Table 7). We found relatively consistent negative associations across studies, between disadvantaged classroom compositions, from a social and economic perspective, and the ECERS scores. Even though a recent meta-analysis about the relationship between ECERS and child outcomes reported that, in general, ECERS scores tend to be low across programs and that little variance in quality measured with the ECERS can impact the level of significance found in associations (Brunsek et al., 2017), these results should be cause for concern. Moreover, although associations with CLASS scores were not so consistent across studies, negative associations between higher proportions of children from disadvantaged backgrounds and emotional and instructional support were found frequently.

Mixed results for the CLASS may arise, for example, from distinct operationalizations of classroom composition indexes and from the diversity in number and type of covariates used in the studies (Perlman et al., 2016). Nevertheless, the significant associations reported in this review should not be overlooked. Evidence from the ECE literature indicates that while emotional support is frequently of medium-high to high-quality (Pianta et al., 2008), instructional support is frequently of low-quality, both in American (e.g., Hamre et al., 2014) and European classroom samples (e.g., Aguiar, Aguiar, Cadima, Correia, & Fialho, 2019; Bihler et al., 2018). Hence, the association between disadvantaged classroom compositions and lower-quality emotional support is particularly relevant, although both raise concerns. In classrooms with high-quality emotional support teachers are sensitive and responsive to children´s emotional states and needs (Pianta, Hamre, & Allen, 2012), and children experience positive and warm interactions with teachers and peers (Pianta et al., 2008). Ultimately, teachers in these classrooms are able to promote the social and emotional functioning of children (Pianta et al., 2008). In classrooms with high-quality instructional support, teachers are able to implement activities in a way that promotes the learning of useful knowledge (Pianta et al., 2008) and contributes to children´s cognitive and linguistic development (e.g., Pianta & Hamre, 2009). Together, these findings indicate that specific groups of disadvantaged children are enrolled in lower-quality classrooms, meaning that potential benefits of high-quality ECE may not be reaching the children most in need.

4.4. Classroom Level Characteristics and Classroom Composition Indexes

Sociodemographic variables are often divided into two or more categories, except age, that can have multiple values (Steel & Tranmer, 2011). This was the case in multiple studies included in this review, that focused mostly on grouping children according to a shared category in a given sociodemographic variable (Steel & Tranmer, 2011), and then contrasting groups of children who fit a different category within the same sociodemographic variable

(e.g., groups of DLL vs. non-DLL; poor vs. non-poor; 100% proportion migrant vs. 0% proportion migrant; Caucasians vs. non-Caucasians; high average maternal education vs. low average maternal education). Consequently, these studies portraited classroom composition in terms of relative homogeneity. Results add to the still scarce evidence that disadvantaged classroom compositions can be associated with lower quality. Conversely there was little evidence about the association between classroom diversity and quality. Only in one study addressing age composition (Ansari & Pianta, 2018) there was a clear consideration of within-group heterogeneity. This study reported a significant association with classroom quality and is illustrative of how a diversity index can be used in the study of diversity regarding distinct demographic characteristics.

Researchers in the education field may not be very familiar with existing diversity indexes (e.g., Roberson, Sturman, & Simons, 2007) that can potentially be adapted to the study of classroom composition or, as reported in other fields of study (see Harrison & Klein, 2007), the concept of diversity may not yet be refined to the point that choices about the most adequate operationalization methods can be clearly made (Harrison & Klein, 2007). However, the development of studies that assess classroom composition diversity is crucial not only to produce in-depth knowledge on the association between classroom composition and quality, but also to adequately inform policies and decision-making processes regarding the organization of classrooms.

No study included in this review used the Herfindahl index, presented in the introduction section, to compute classroom composition diversity. Nonetheless, this index has already been used in the field of education. For example, Dronkers and van der Velden (2012), in a study with 15 year-olds, used this index with complementary calculations of the average/share of children from a set of particular countries of origin to compute the school ethnic composition, so that a combined effect of ethnic diversity and share on students

outcomes could be examined. Diversity and average/share can, thus, be used as separate and complementary group composition indicators (Dronkers & vand der Velden, 2012). Other composition indexes, mostly used in studies outside the education literature, should be examined in future research about the association between classroom composition and quality in ECE. The mean Euclidean distance, the standard deviation, Teachman's index, Blau's index, the coefficient of variation, and the Gini coefficient of concentration have all been used to determined differences in the distribution of demographic characteristics such as age, gender, ethnicity, and education level, within groups (e.g., Harrison & Klein, 2007). These indexes allow a direct and simple calculation of diversity effects, but they do not account for group size or differences in the number of categories between characteristics (Solanas et al., 2012). Thus, group variances must be corrected to account for the effects of differences in group size, when aggregating different groups with respect to a given category, to prevent systematic bias (e.g., Biemann & Kearney, 2010). Bias-corrected formulas have been proposed for each of these measures (see Biemman & Kearney, 2010).

In sum, there are some group composition indexes with good potential that can be used to ascertain levels of diversity within ECE classrooms. However, the choice of the index must be guided by a clear definition of diversity (Harrison & Klein, 2007). In this review, we discussed some alternatives to ascertain diversity at the school and classroom levels, as well a broader conceptualization of diversity, considering parameters of separation, variety, and disparity. They may help researchers choose the most adequate operationalization method, accordingly with the research aim. If correctly operationalized, diversity indexes can produce valid and robust evidence (Biemman & Kearney, 2010) on classroom composition effects.

4.5. Associations Between Classroom Composition and Observed Classroom Quality

Overall, we found evidence that supports the importance of examining the association between classroom composition and process quality. The focus of most studies on minority

status likely illustrates the political and research agendas prioritizing the needs of groups of children experiencing early achievement gaps (Bridges et al., 2004). Although, in general, evidence indicates that classrooms with higher proportions of children with minority status attended lower quality classrooms, results were somewhat mixed. Apparent inconsistencies found across studies included in this review are in line with evidence about the quality of programs serving children in social and economic disadvantage (see Magnuson, Meyers, Ruhm, & Waldfogel, 2004).

In studies conducted in the USA, results varied, particularly in the association between classrooms with a high concentration of Hispanic/Latino children. A couple of studies reported lower quality in classrooms with more Hispanic children, but most did not find a significant association. Confounding effects can help explain this lack of significant results, since only one of these studies (Iruka & Morgan, 2014) modeled for other structural indicators. The study reported that teacher's education, training, and enjoyment of their job were associated with classroom quality (Iruka & Morgan, 2014). Hispanic/Latino children are often dual language learners; so the lack of significant associations may be due to interactions with other factors believed to be associated with the use of bilingual practices, such as teachers' motivation and preparedness to teach DLL's or administrator support (e.g., Sawyer et al., 2016), which can derive, for example, from the development of new models of ECE that target the specific needs of the Hispanic/Latino communities (Downer et al., 2012).

Conversely, examined studies seem to indicate that Black/African-American children and children with other migration backgrounds are more likely to be enrolled in ECE classrooms with lower process quality, particularly when considering the CLASS emotional and instructional support domains. Conversely, one study (Dotterer et al., 2014) found higher instructional support and global quality in classrooms from targeted programs that served mostly children with minority status. One possible explanation for this contradictory result is

related with differences in investment across states and, consequently, in the quality of programs (Cryer et al., 1999) that minority children attend. Pre-K and Head Start programs frequently provide better quality education and care, compared with other community programs (Magnuson et al., 2004), so some minority children may be experiencing modest to good classroom quality (Iruka & Morgan, 2014).

An association between higher concentrations of children with a migration background and lower process quality was also reported in four of the five studies conducted in Europe. One European study (Kuger et al., 2016) reported a negative association between a higher proportion of children with a migration background and low proficiency in the language of the host country and classroom quality measured with the ECERS and two others (Broekhuizen et al., 2017; Slot et al., 2018) reported a similar association with the CLASS domains, with particular emphases on emotional support. Furthermore, one of these studies reported that quality tended to decrease from year to year, as concentration levels increased (Kuger et al., 2016). These classrooms may be more challenging for teachers because of communication limitations and increased difficulties in structuring learning activities (Kuger et al., 2016). Also, the accumulation of such challenges over time can, perhaps, be reflected in process quality levels (Ansari & Pianta, 2018). Providing professional development opportunities and assuring a more balanced adult-to-child ratio, for example, may help mitigate these negative associations (Kuger et al., 2016).

As expected, we found studies that reported negative associations between lower SES classroom compositions and process quality. However, we note that risk factors such as poverty and minority status group often overlap (e.g., Williams, Priest, & Anderson, 2016). In socioeconomic disadvantaged ECE classrooms, teachers are often less experienced than those allocated to classrooms with high-SES compositions (see Kalogrides & Loeb, 2013; Kalogrides, Loeb, & Beteille, 2013; Reid & Ready, 2013) and are more likely to have

insufficient training and lack the necessary support to effectively manage groups of children with increased emotional and behavioral difficulties (see Raver et al., 2008; Raver et al., 2009). Teachers in classrooms serving children from disadvantaged backgrounds also seem to hold less child-centered views compared with teachers in classrooms with more favorable sociocultural compositions (Lee & Ginsburg, 2007). At least one study considered a reasonable array of covariates at the teacher, classroom, and child levels, and still reported lower quality on both ECERS and CLASS in classrooms with a higher concentration of children living in poverty, which indicates that classroom SES composition can also be a predictor of classroom quality (Pianta et al., 2005).

Most studies did not report associations between classroom age composition and process quality. However, based on two studies, classrooms with higher age diversity and with more younger children seem to have lower quality. The two studies that reported an association between classroom age composition and process quality considered an array of covariates, associated with teacher and classroom characteristics, including other classroom composition indexes, such as gender, ability (Ansari & Pianta, 2018), and migration background (Kuger et al., 2016), which can increase the accuracy of findings. These results may indicate that attending to the needs of children in these classrooms can be more demanding, particularly for less experienced teachers and for those with teacher-centered views (Ansari & Pianta, 2018). Although heterogeneous classrooms are increasingly common, there is no substantial empirical evidence supporting that this model is associated with better process quality (Ansari & Pianta, 2018). Further exploring the association between classroom age composition and process quality can have practical implications, for example, by informing enrollment policies about age cutoff points (Ansari & Pianta, 2018), if adequate, or determine more favorable funding of staff based on the classroom age composition (Kuger et al., 2016). Mechanisms to regulate classroom age composition should

be dependent on how the national ECE system in question is organized (Fuller, Kagan, Loeb, & Chang, 2004).

Evidence was not clear about the association between classroom ability composition and process quality. All studies used different quality observation measures. Two reported significant associations, but in one of them (Sontag, 1997) the authors discussed a potential artifact, associated with a specific classroom. In the other study, teachers in classrooms with more children with disabilities provided higher-quality literacy instruction. Teachers in these classrooms may benefit from additional supports from early childhood intervention and early childhood special education professionals and, therefore, may have additional resources to individualize their literacy instruction practices, thus increasing observed quality (e.g., Coombs-Richardson & Mead, 2001). These teachers can also have more experience working with children with disabilities and, consequently, have greater knowledge in the application of such practices (e.g., Küçüker, Acarlar, & Kapci, 2006). More research about the association between classroom quality and classroom ability composition is clearly needed.

Lastly, we address the lack of significant associations between classroom gender composition and process quality. The three studies examined used distinct quality observation measures. One of these measures was associated with culturally responsive teaching and might not be the most adequate to investigate the association with classroom gender composition. The lack of significant associations in the remaining two studies, that assessed emotional and behavioral support, is of particular interest, since we expected to find lower quality in classrooms with more boys (e.g., Baker, 2006; Hamre & Pianta, 2001). It might have been that confounding effects were at play. Although one of the studies considered a few teacher and classroom level covariates, other indicators frequently associated with quality levels, such as teacher´s education, training, or experience (e.g., Phillipsen et al.,

1997) were not included. Another possibility is that the variance in the percentage of boys and girls in the samples was not sufficient to produce statistically significant associations.

Even though we proposed classroom composition as a relevant structural feature and a predictor of process quality in ECE classrooms, this association may not always be linear. Investigating the impact of a single or a couple of structural features may be limited and insufficient to capture variations in process quality (Cryer et al., 1999), since variation may result from multiple factors and interactions among them (Slot et al., 2018). Indicators at the classroom and center levels (e.g., financial resources, type of program, center size), as well as more distal structural indicators, at the national and community levels (Cryer et al., 1999) (e.g., subsidies, regulatory mechanisms [Schechter & Bye, 2007], quality monitoring systems [Blau, 2001], community economic well-being [Cryer et al., 1999]) can interact with classroom composition to explain variations in process quality. Nevertheless, this review presents initial evidence that supports further investigation of which classroom composition indexes in ECE may be associated with quality and under which circumstances.

4.6. Limitations

First, we discuss limitations associated with the review process. This review may have been limited by the search strategy used. Although we defined a multitude of key terms and search strings regarding the most commonly studied classroom composition indexes, we limited this search to the title, abstract, key terms, and topic of studies. Thus, while we did this to ensure both sensitivity and specificity in our approach (Hempel et al., 2016), we might have failed to capture literature that could contribute to a deeper understanding of the association between the composition of the classroom and observed quality. Moreover, the fact that the large majority of studies included in this review were conducted in the USA may be due to a biased search strategy and to our inability to review studies in languages other than English and Portuguese. Our decision to only include studies that assessed classroom

quality with standardized observation measures may also have narrowed our scope. However, these measures tend to produce more reliable data, compared to non-standardized measures (Burchinal, 2010; Goldring et al., 2015). Similarly, our pole of studies could have been more substantial if studies with younger children and at the center level were included. Nonetheless, we felt our decisions regarding both issues were justified by practical and substantive reasons. Lastly, this synthesis is fundamentally descriptive, since conducting a meta-analysis did not seem appropriate due to the variability in sample characteristics, classroom composition indexes, study designs, standardized observation measures of classroom quality, covariates and statistics (e.g., Ahn & Kang, 2018).

Regarding limitations associated with the characteristics of the studies included in this review, 20 of the 25 studies were cross-sectional and collected data on only one occasion. Thus, these studies provide a static picture of ECE classrooms (Curby et al., 2011; Kuger et al., 2016) that may not represent accurately the predominant interaction patterns (Sawyer et al., 2016). Also, it is not possible to disentangle the direction of the associations or outline a more comprehensive scope of the challenges teachers face associated with more disadvantaged classroom compositions and with changes in composition (Ansari & Pianta, 2018), in order to determine the aspects and mechanisms associated with stability or change in quality levels over time (Kuger et al., 2016). Furthermore, 23 out of the 25 studies were correlational, therefore, no causal associations can be drawn (Read & Ready, 2013)

In this review, effect sizes appeared to be generally small, as it is common to find in studies conducted in ECE settings (e.g., NICHD ECCRN, 2002; Pianta, La Paro, Payne, Cox, & Bradley, 2002), but may have important practical implications since many disadvantaged children may be experiencing lower-quality ECE, which can have a substantial adverse effect on children´s development (see Melhuish et al., 2015). However, any estimates must be interpreted with caution due to potential selection effects (Hill, Rosenman, Tennekoon, &

Mandal, 2013). Variability in this review may be restricted (Perlman et al., 2016), since multiple studies relied on data from the same large-scale studies, mainly conducted in the USA, with samples that seem to overrepresent disadvantaged programs. Although some of the large-scale studies, such as the NCEDL and SWEEP studies, selected programs randomly, more than 20% of the invited programs for the NCEDL did not participate and parental consent was around 60%. This means that the samples from these studies may not be entirely representative (Perlman et al., 2016).

Also, since multiple studies reported zero-order correlations and simple mean comparisons between two groups regarding the association of classroom composition and quality, results are potentially exposed to the influence of confounders. Finally, considering that no single standardized observation measure can cover all relevant aspects of classroom quality (Bassok & Galdo, 2016), most studies were limited by the use of only one standardized observation measure. For example, quality measures such as APEEC, ECCOM, ECERS, or CIS, do not cover instructional support/practices, an essential dimension of teaching, associated with children's social, language, and academic outcomes (e.g., Hamre, Hatfield, Pianta, & Jamil, 2014). Thus, complementing these measures with others that capture teachers' instructional practices (e.g., CLASS) can mitigate limitations inherent to the use of one single measure (Maxwell et al., 2011).

4.7. Implications for Practice

The results of the studies examined in this review indicate that in classrooms with higher percentages of children with minority status and low SES, process quality is lower. These results are in line with previous evidence suggesting that there may be a trend for children to be enrolled in classrooms with peers from similar backgrounds (Reid & Kagan, 2015) which becomes problematic when quality gaps become large, as those reported by Valentino (2018). Creating mechanisms that ensure a more balanced sociocultural

composition in ECE classrooms can have practical implications when it comes to reduce process quality gaps (de Haan, Elbers, Hoofs, & Leseman, 2013).

Furthermore, teacher allocation processes should consider classroom composition, so that more qualified teachers are assigned to classrooms serving higher percentages of children from minority and low SES backgrounds, in an attempt to raise the quality within particularly challenging groups (Ansari & Pianta, 2018). But more than teacher allocation, it is important to design and implement training and professional development programs for all teachers, that address the main difficulties experienced in their interactions with more challenging groups of children (e.g., Pianta et al., 2009; Valentino, 2018).

In order to improve classroom quality for all children, evaluation and certification processes should adopt an holistic perspective of quality in ECE (Kuger et al., 2016). This involves a focus on the identification of key factors that may be associated with interaction patterns and teaching practices that can benefit all children (Ansari & Pianta, 2018; Maxwell et al., 2010). Beyond the regulation of administrative procedures, quality rating improvement systems should focus on teachers' ability to support the social and academic development of children through their daily interactions in the classroom (Pianta et al., 2008).

4.8. Implications for Future Research

More research focused primarily on the association between classroom composition and process quality is clearly needed. Additionally, studies regarding the quality of ECE programs are conducted primarily in the USA. Differences in policy, regulatory mechanisms, and investment in ECE across countries (Vermeer, van Ijzendoorn, Cárcamo, & Harrison, 2016), reflect the cultural values about childhood of a given society (e.g., Bertram et al., 2016 and cannot be overlooked. Hence, more research in different cultural contexts, such as the European, could contribute to a deeper understanding of how ECE policies and service models may be associated with differences in quality (Vlasov et al., 2016).

Given the recent emphases on the potential benefits of classroom heterogeneity (see European Commission, 2018; Reid & Ready, 2013), research using diversity indexes to study the association between classroom heterogeneity and quality is warranted. Future research focused, for example, on the skills and practices of highly qualified teachers and teachers who endorse child-centered views, can contribute to the design of more efficient quality improvement programs (Ansari & Pianta, 2018; Purtell & Ansari, 2018).

Studies analyzed principal effects of classroom composition indexes. Future research considering the potential interactions between indexes and other structural features of the classroom context can further our understanding of classroom composition effects. Longitudinal studies examining variations in the association between classroom quality levels and context factors, as a function of fluctuations in quality and/or context, over time, are also warranted (Kuger et al., 2016).

In sum, the evidence gathered in this review supports the proposition that classroom composition may be a key component to consider in the assessment of classroom structural features as well as in the definition of strategies aiming to improve ECE quality (Reid & Ready, 2013). It underpins the need for future research regarding the association between different types of classroom composition and quality in ECE.

References

Aguiar, A. L., Aguiar, C., Cadima, J., Correia, N., & Fialho, M. (2019). Classroom quality and children's social skills and problem behaviors: Dosage and disability status as moderators. *Early Childhood Research Quarterly, 49*, 81-92. doi:10.1016/j.ecresq.2019.05.005

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology, 71*(2), 103. doi:10.4097/kjae.2018.71.2.103

Arnett, J. (1989). Caregivers in day-care centers: Does training matter?. *Journal of Applied Developmental Psychology, 10(4),* 541-552. doi:10.1016/0193-3973(89)90026-9

*Ansari, A., & Pianta, R. C. (2018). The role of elementary school quality in the persistence of preschool effects. *Children and Youth Services Review, 86,* 120-127. doi:10.1016/j.childyouth.2018.01.025

Baker, J. A. (2006). Contributions of teacher–child relationships to positive school adjustment during elementary school. *Journal of School Psychology, 44*(3), 211-229. doi:10.1016/j.jsp.2006.02.002

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

*Bassok, D., & Galdo, E. (2016). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development, 27(1),* 128-144. doi:10.1080/10409289.2015.1057463

Baumwell, L., Tamis-LeMonda, C. S., & Bornstein, M. H. (1997). Maternal verbal sensitivity and child language comprehension. *Infant Behavior and Development, 20*(2), 247-258. doi:10.1016/S0163-6383(97)90026-6

Belfi, B., Goos, M., De Fraine, B., & Van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic

self-concept: A literature review. *Educational Research Review*, *7*(1), 62-74. doi:10.1016/j.edurev.2011.09.002

Bell, E. R., Greenfield, D. B., & Bulotsky-Sheare, 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

Bertram, T., Pascal, C., Cummings, A., Delaney, S., Ludlow, C., Lyndon, H., & Stancel-Piatak, A. (2016). *Early childhood policies and systems in eight countries: Findings from IEA's early childhood education study.* Retrieved from https://www.iea.nl/sites/default/files/2019-04/ECES-policies_and_systems-report.pdf

Biemann, T., & Kearney, E. (2010). Size does matter: How varying group sizes in a sample affect the most common measures of group diversity. *Organizational Research Methods*, *13*(3), 582-599. doi:10.1177/1094428109338875

*Bihler, L. M., Agache, A., Schneller, K., Willard, J. A., & Leyendecker, B. (2018). Expressive morphological skills of dual language learning and monolingual German children: Exploring links to duration of preschool attendance, classroom quality, and classroom composition. *Frontiers in Psychology*, *9*. doi:10.3389/fpsyg.2018.00888

Blau, D. M. (2001). *The Child Care Problem: An Economic Analysis*. New York: The Russell Sage Foundation.

Bridges, M., Fuller, B. C., Rumberger, R., & Tran, L. (2004). *Preschool for California's children: Promising benefits, unequal access.* PACE Policy Brief, 04-3, Berkeley, CA: Policy Analysis for California Education (PACE). Retrieved from http://pace.berkeley.edu/packard_foundation_study.html

* Broekhuizen, M. L., Slot, P. L., van Aken, M. A., & Dubas, J. S. (2017). Teachers' emotional and behavioral support and preschoolers' self-regulation: Relations with

social and emotional skills during play. *Early Education and Development, 28*(2), 135-153. doi:10.1080/10409289.2016.1206458

Bronfenbrenner, U. (1992). *Ecological systems theory*. Jessica Kingsley Publishers.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (6th ed., pp. 793– 828). New York, NY: Wiley.

Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta-analysis. *PloS ONE, 12*(6). doi: 10.1371/journal.pone.0178512

Bryant, D. (2010). *Observational measures of quality in center-based early care and education programs* (Research-to-policy, research-to-practice brief OPRE 2011-10c). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Canadian Centre for Knowledge Mobilisation.

Burchinal, M. (2011). Differentiating among measures of quality: Key characteristics and their coverage in existing measures. *Research to Policy, Research to Practice Brief OPRE*. Retrieved from https://www.acf.hhs.gov/sites/default/files/opre/differ_measures.pdf

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*(1), 2-11. doi:10.1207/S1532480XADS0601_01

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale

studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11-31). Baltimore, MD, US: Paul H Brookes Publishing.

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, F. (2008). Classroom problem behavior and teacher-child relationships in kindergarten: The moderating role of classroom climate. *Journal of School Psychology, 46,* 367-391. doi:10.1016/j.jsp.2007.06.009

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

Camp, W. (2001). Formulating and evaluating theoretical frameworks for career and technical education research. *Journal of Vocational Education Research, 26*(1), 4-25. doi:10.5328/JVER26.1.4

Carta, J.J., Greenwood, C. R., & Atwater, J. (1992). *ESCAPE: Eco-behavioral System for Complex Assessments of Preschool Environments*. Kansas City: University of Kansas, Juniper Gardens Children's Project.

Castro, D. C. (2005). Early language and literacy Classroom observation: addendum for English language learners. *Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute*.

Clements, M. A., Reynolds, A. J., & Hickey, E. (2004). Site-level predictors of children's school and social competence in the Chicago Child–Parent Centers. *Early Childhood Research Quarterly, 19*(2), 273-296. doi:10.1016/j.ecresq.2004.04.005

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research, 22(10),* 1435–1443. doi:10.1177/1049732312452938

Coombs-Richardson, R., & Mead, J. (2001). Supporting general educators' inclusive practices. *Teacher Education and Special Education, 24*(4), 383-390. doi:10.1177/088840640102400412

Cryer, D., Tietze, W., Burchinal, M., Leal, T., & Palacios, J. (1999). Predicting process quality from structural quality in preschool programs: A cross-country comparison. *Early Childhood Research Quarterly, 14*(3), 339-361. doi:10.1016/S0885-2006(99)00017-4

Cueto, S., León, J., & Miranda, A. (2016). Classroom composition and its association with students' achievement and socioemotional characteristics in Peru. *Assessment in Education: Principles, Policy & Practice*, *23*(1), 126-148. doi:10.4324/9781351257169-7

Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Downer, J., ... & Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade: Implications for children's experiences and conducting classroom observations. *The Elementary School Journal*, *112*(1), 16-37. doi: 10.1086/660682

*Debnam, K. J., Pas, E. T., Bottiani, J., Cash, A. H., & Bradshaw, C. P. (2015). An examination of the association between observed and self-reported culturally proficient teaching practices. *Psychology in the Schools*, *52*(6), 533-548. doi:10.1002/pits.21845

de Haan, A., Elbers, E., Hoofs, H., & Leseman, P. P. M. (2013). Targeted versus mixed preschools and kindergartens: effects of class composition and teacher-managed activities on disadvantaged children's emergent academic skills. *School Effectiveness and School Improvement*, *24*(2), 177-194. doi:10.1080/09243453.2012.749792

den Brok, P., Van Tartwijk, J., Wubbels, T., & Veldman, I. (2010). The differential effect of the teacher–student interpersonal relationship on student outcomes for students with

different ethnic backgrounds. *British Journal of Educational Psychology*, *80*(2), 199-221. doi:10.1348/000709909X465632

Denton, K., Germino-Hausken, E., & West, J. (2000). *America's kindergartners*. Washington, DC: National Center for Education Statistics.

DiLalla, L. F., & Mullineaux, P. Y. (2008). The effect of classroom environment on problem behaviors: A twin study. *Journal of School Psychology, 46,* 107-128. doi:10.1016/j.jsp.2007.02.001

*Dotterer, A. M., Lowe, K., & McHale, S. M. (2014). Academic growth trajectories and family relationships among African American youth. *Journal of Research on Adolescence, 24(4),* 734-747. doi:10.1111/jora.12080

*Downer, J. T., López, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher–child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly*, *27*(1), 21-32. doi:10.1016/j.ecresq.2011.07.005

Dronkers, J., & Van der Velden, R. (2013). Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using PISA data. In *Integration and inequality in educational institutions* (pp. 71-98). Springer, Dordrecht.

European Comission. (2018). *Proposal for a Council recommendation on high quality early childhood education and care systems*. Retrieved from https://eur-lex.europa.eu/resource.html?uri=cellar:05aa1e50-5dc7-11e8-ab9c-01aa75ed71a1.0003.02/DOC_1&format=PDF

European Commission/EACEA/Eurydice (2019). *Key data on early childhood education and care in Europe – 2019 edition. Eurydice report.* Luxembourg:

Publications Office of the European Union.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research, 94*(5), 275-282. doi:10.1080/00220670109598763

*Fram, M. S., Kim, J., & Sinha, S. (2012). Early care and prekindergarten care as influences on school readiness. *Journal of Family Issues, 33(4),* 478-505. doi:10.1177/0192513X11415354

Friedman, S., & Amadeo, J. (1999). The child-care environment: Conceptualizations, assessments, and issues. In S.L. Friedman & T.D. Wachs (Eds.), *Measuring environment across the lifespan* (pp. 127–165). Washington, DC: American Psychological Association.

*Friedman-Krauss, A. H., Raver, C. C., Morris, P. A., & Jones, S. M. (2014). The role of classroom-level child behavior problems in predicting preschool teacher stress and classroom emotional climate. *Early Education and Development*, *25*(4), 530-552. doi:10.1080/10409289.2013.817030

Fuller, B., Kagan, S. L., Loeb, S., & Chang, Y. W. (2004). Child care quality: Centers and home settings that serve poor families. *Early Childhood Research Quarterly*, *19*(4), 505-527. doi:10.1016/j.ecresq.2004.10.006

Gay, G. (2000). *Culturally Responsive Teaching: Theory, Research, and Practice*. New York: Teachers College Press.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher, 44*(2), 96-104. doi:10.3102/0013189X15575031

Gottfried, M. A., & Graves, J. (2014). Peer effects and policy: The relationship between

classroom gender composition and student achievement in early elementary school. *The BE Journal of Economic Analysis & Policy*, *14*(3), 937-977. doi:10.1515/bejeap-2013-0123

Graham, S. (2004). Ethnicity and peer harassment during early adolescence. In T. Urdan, T. & F. Pajares (Eds.). *Educating Adolescents: Challenges and Strategies* (pp. 85-108). Charlotte, NC: IAP.

Guo, Y., Tompkins, V., Justice, L., & 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

Hamre, B. K., Goffin, S. G., & Kraft-Sayre, M. (2009). *Classroom Assessment Scoring System implementation guide: Measuring and improving classroom interactions in early childhood settings.* Center for Advanced Study of Teaching and Learning. Retrieved from http://curry.virginia.edu/research/centers/ castl/publications

Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children's development. *Child development*, *85*(3), 1257-1274. doi:10.1111/cdev.12184

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Rating Scale, revised edition (ECERS)*. Williston, VT: Teachers College Press.

Harrison, D. A., & Klein, K. J. (2007) What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review, 32,* 1199-1228. doi:10.5465/amr.2007.26586096

Harrison, D. A., & Sin, H. P. (2006). What is diversity and how should it be measured? In A. M. Konrad, P. Prasad, & J. K. Pringle (Eds.), *Handbook of workplace diversity* (pp. 191-217). Thousand Oaks, CA: Sage.

Hemmeter, M.L., Maxwell, K. L., Ault, M. J., & Schuster, J.W. (2001). *Assessment of practices in early elementary classrooms (APEEC)*. New York: Teachers College Press.

Hempel, S., Xenakis, L., & Danz, M. (2016). *Systematic reviews for occupational safety and health questions: Resources for evidence synthesis*. Retrieved from https://www.jstor.org/stable/10.7249/j.ctt1d9nnzq.11?seq=1#metadata_info_tab_contents

Hill, L. G., Rosenman, R., Tennekoon, V., & Mandal, B. (2013). Selection effects and prevention program outcomes. *Prevention science*, *14*(6), 557-569. doi: 10.1007/s11121-012-0342-x

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early childhood research quarterly*, *23*(1), 27-50. doi:10.1016/j.ecresq.2007.05.002

*Iruka, I. U., & Morgan, J. (2014). Patterns of quality experienced by African American children in early education programs: Predictors and links to children's preschool and kindergarten academic outcomes. *The Journal of Negro Education*, *83*(3), 235-255. doi:10.7709/jnegroeducation.83.3.0235

Jones, S. (2016). How does classroom composition affect learning outcomes in Ugandan primary schools?. *International Journal of Educational Development*, *48*, 66-78. doi:10.1016/j.ijedudev.2015.11.010

Johnson-Bailey, J., & Drake-Clark, D. (2010). Book Review: Denzin, N.K., Lincoln, Y.S., & Smith, L.T. (Eds.). (2008). Handbook of Critical and Indigenous Methodologies. *Adult Education Quarterly*, *60*(4), 405-407. doi:10.1177/0741713609350411

* Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. Early childhood research quarterly, 23(1), 51-68. doi:10.1016/j.ecresq.2007.09.004

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher, 42(6),* 304-316. doi:10.3102/0013189X13495087

Kalogrides, D., Loeb, S., & Béteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education, 86*(2), 103-123. doi:10.1177/0038040712456555

Karoly, L. A., Zellman, G. L., & Perlman, M. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly, 28*(4), 645-657. doi:10.1016/j.ecresq.2013.05.001

Khanna, N., & Harris, C. A. (2009). Teaching race as a social construction: Two interactive class exercises. *Teaching Sociology*, *37*(4), 369-378. doi: 10.1177/0092055x0903700405

*Kuger, S., Kluczniok, K., Kaplan, D., & Rossbach, H. G. (2016). Stability and patterns of classroom quality in German early childhood education and care. *School Effectiveness and School Improvement, 27(3)*, 418-440. doi:10.1080/09243453.2015.1112815

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist, 46*, 819–834. doi:10.1037=0003-066X.46.8.819

Li, Y., Xu, L., Liu, L., Lv, Y., Wang, Y., & Huntsinger, C. S. (2016). Can preschool socioeconomic composition moderate relationships between family environment

and Chinese children's early academic and social outcomes?. *Children and Youth Services Review, 60,* 1-10. doi:10.1016/j.childyouth.2015.11.009

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine, 151*(4), 65-94. doi:10.1371/journal.pmed.1000100

*LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., ... & Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3-17. doi:10.1016/j.ecresq.2006.05.001

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal, 41*(1), 115-157. doi:10.3102/00028312041001115

Markus, H. R. (2008). Pride, prejudice, and ambivalence: Toward a unified theory of race and ethnicity. *American Psychologist, 63*(8), 651-670. doi:10.1037/0003-066X.63.8.651

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749. doi:10.1111/j.1467-8624.2008.01154.x

*Maxwell, K. L., McWilliam, R. A., Hemmeter, M. L., Ault, M. J., & Schuster, J. W. (2001). Predictors of developmentally appropriate classroom practices in kindergarten through third grade. *Early Childhood Research Quarterly, 16(4),* 431-452. doi:10.1016/S0885-2006(01)00118-1

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., ... & Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher*, *46*(8), 474-487. doi:10.3102/0013189X17737739

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., …, & Leseman, P. (2015). *A review of research on the effects of early childhood education and care on child development*. CARE Project Report. Retrieved from http://ecec-care.org/fileadmin/careproject/Publications/reports/new_version_CARE_WP4_D4_1_Review_on_the_effects_of_ECEC.pdf

National Institute of Child Health and Human Development Early Child Care Research Network (2002). Child-care structure→ process→ outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science, 13(3),* 199-206. doi:10.1111/1467-9280.00438

National Forum on Education Statistics (2008). *Managing an identity crisis: Forum guide to implementing new federal race and ethnicity categories*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from: https://nces.ed.gov/pubs2008/2008802.pdf

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. doi:10.1186/s13643-016-0384-4

*Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. E. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development*, *21*(1), 95-124. doi:10.1080/10409280902858764

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PLoS ONE*, *11*(12). doi:10.1371/journal.pone.0167660

*Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly*, *24*(3), 213-228. doi:10.1016/j.ecresq.2009.05.002

Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, *12*(3), 281-303. doi:10.1016/s0885-2006(97)90004-1

Pianta, R. C., DeCoster, J., Cabell, S., Burchinal, M., Hamre, B. K., Downer, J., ... & Howes, C. (2014). Dose–response relations between preschool teachers' exposure to components of professional development and increases in quality of their interactions with children. *Early Childhood Research Quarterly*, *29*(4), 499-508. doi:10.1016/j.ecresq.2014.06.001

*Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions?. *Applied Developmental Science*, *9*(3), 144-159. doi:10.1207/s1532480xads0903_2

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), 109-119. doi:10.3102/0013189x09332374

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom

interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365–386). New York: Springer.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Baltimore, MD, US: Paul H Brookes Publishing.

Pianta, R. C., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal*, *102*(3), 225-238. doi:10.1086/499701

*Purtell, K. M., & Ansari, A. (2018). Classroom Age Composition and Preschoolers' School Readiness: The Implications of Classroom Quality and Teacher Qualifications. *AERA Open*, *4*(1). doi:10.1177/2332858418758300

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly*, *23*(1), 10-26. doi:10.1016/j.ecresq.2007.09.001

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B.(2009). Targeting children´s behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *77*(2), 302-316. doi:10.1037/a00153

Reid, J. L., & Kagan, S. L. (2015). *A Better Start: Why Classroom Diversity Matters in Early Education.* Poverty & Race Research Action Council. Retrieved from https://files.eric.ed.gov/fulltext/ED571023.pdf

*Reid, J. L., & Ready, D. D. (2013). High-quality preschool: The socioeconomic composition of preschool classrooms and children's learning. *Early Education and Development, 24,* 1082 – 1111. doi:10.1080/10409289.2012.757519

Resnick, L. B. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher*, *39*(3), 183-197. doi:10.3102/0013189X10364671

Roberson, Q. M., Sturman, M. C., & Simons, T. L. (2007). Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indexes. *Organizational Research Methods*, *10*(4), 564-588. doi:10.1177/1094428106294746

Rusby, J. C., Taylor, T., & Milchak, C. (2001). *Assessing school settings: Interactions of students and teachers (ASSIST) observation system*. Unpublished manual.

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16(2),* 125–141. doi:10.1521/scpq.16.2.125.18698

Sameroff, A. (2009). The transactional model. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 3-21). Washington, DC, US: American Psychological Association.

*Sanders, K., & Downer, J. (2012). Predicting acceptance of diversity in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *27*(3), 503-511. doi:10.1016/j.ecresq.2011.12.001

*Sawyer, B. E., Hammer, C. S., Cycyk, L. M., López, L., Blair, C., Sandilos, L., & Komaroff, E. (2016). Preschool teachers' language and literacy practices with dual language learners. *Bilingual Research Journal*, *39*(1), 35-49. doi:10.1080/15235882.2016.1138904

SCImago (2019). SJR — SCImago journal & country rank. Retrieved from http://www.scimagojr.com.

Schechter, C., & Bye, B. (2007). Preliminary evidence for the impact of mixed-income preschools on low-income children's language growth. *Early Childhood Research Quarterly, 22*(1), 137-146. doi:10.1016/j.ecresq.2006.11.005

Schaeffer, M. (2013). Can competing diversity indices inform us about why ethnic diversity erodes social cohesion? A test of five diversity indices in Germany. *Social Science Research, 42*(3), 755-774. doi:10.1016/j.ssresearch.2012.12.018

Shaw, D. S., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry, 46*(9), 931-942. doi:10.1111/j.1469-7610.2004.00390.x

*Slot, P. L., Bleses, D., Justice, L. M., Markussen-Brown, J., & Højen, A. (2018). Structural and process quality of Danish preschools: Direct and indirect associations with children's growth in language and preliteracy skills. *Early Education and Development, 29*(4), 581-602. doi:10.1080/10409289.2018.1452494

Slot, P. L., Leseman, P. P., Verhagen, J., & Mulder, H. (2015). Associations between structural quality aspects and process quality in Dutch early childhood education and care settings. *Early Childhood Research Quarterly, 33*, 64-76. doi:10.1016/j.ecresq.2015.06.001

Snell, E. K., Hindman, A. H., & Belsky, J. (2015). Child effects and child care: Implications for risk and adjustment. *Development and psychopathology, 27*, 1059-1076. doi: 10.1017/S0954579415000681

Snyder, J., Cramer, A., Afrank, J., & Patterson, G. R. (2005). The contributions of ineffective discipline and parental hostile attributions of child misbehavior to the development of conduct problems at home and school. *Developmental Psychology, 41*(1), 30-41. doi:10.1037/0012-1649.41.1.30

Solanas, A., Selvam, R. M., Navarro, J., & Leiva, D. (2012). Some common indices of group diversity: Upper boundaries. *Psychological Reports, 111*(3), 777-796. doi:10.2466/01.09.21.PR0.111.6.777-796

*Sontag, J. C. (1997). Contextual factors influencing the sociability of preschool children with disabilities in integrated and segregated classrooms. *Exceptional Children, 63*(3), 389-404. doi:10.1177/001440299706300307

Steel, D. G., & Tranmer, M. D. (2011). Measuring and analyzing the within group homogeneity of multi-category variables. *Journal of Statistical Theory and Practice*, *5*(4), 649-658. doi:10.1080/15598608.2011.10483736

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research, 86*(4), 849-899. doi:10.3102/0034654316675417

Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure?. *Educational Evaluation and Policy Analysis*, *38*(2), 293-317. doi:10.3102/0162373715616249

*Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly*, *19*(4), 548-568. doi:10.1016/j.ecresq.2004.10.010

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19(3),* 375-397. doi:10.1016/j.ecresq.2004.07.007

Stolle, D., Soroka, S., & Johnston, R. (2008). When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions. *Political Studies, 56*(1), 57-75. doi:10.1111/j.1467-9248.2007.00717.x

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research, 9*(2), 109-124. doi:10.1177/1476718X10387900

*Valentino, R. (2018). Will public pre-K really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal, 55(1*), 79-116. doi:10.3102/0002831217732000

Van Ewijk, R., & Sleegers, P. (2010a). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*(2), 134-150. doi:10.1016/j.edurev.2010.02.001

Van Ewijk, R., & Sleegers, P. (2010b). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement, 21*(3), 237-265. doi:10.1080/09243451003612671

Veerman, G.-J.M., Van de Werfhorst, H.G., & Dronkers, J. (2013). Ethnic composition of the class and educational performance in primary education in The Netherlands. *Educational Research and Evaluation, 19,* 370-401. doi:10.1080/13803611.2013.788851

Vervaet, R., Van Houtte, M., & Stevens, P. A. (2018). Multicultural school leadership, multicultural teacher culture and the ethnic prejudice of Flemish pupils. *Teaching and Teacher Education, 76*, 68-77. doi:10.1016/j.tate.2018.08.009

Vlasov, J., Hujala, E., Essary, J., & Lenskaya, E. (2016). Longitudinal study of changes in teachers' views of early childhood education in the USA, Russia, and Finland. *Journal of Early Childhood Education Research, 5*(1), 202-222.

Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children´s development: Direct and moderated effects. *Child Development, 75*(1), 296-312. doi:10.1111/j.1467-8624.2004.00670.x

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.

Unsworth, S. (2016). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of Child Language, 43*, 603-634. doi:10.1017/S030500091500080X

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*(5), 422-432. doi:10.1016/j.appdev.2014.07.001

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review, 95*(2), 199-203. doi:10.1257/000282805774670158

Williams, D. R., Priest, N., & Anderson, N. B. (2016). Understanding associations among race, socioeconomic status, and health: Patterns and prospects. *Health Psychology*, *35*(4), 407-411. doi:10.1037/hea0000242

*Z*aslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L. and Burchinal, M. (2010). *Quality dosage, thresholds, and features in early childhood settings: A review of the literature, OPRE 2011-5.* Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

* References marked with an asterisk indicate studies included in the systematic review.

Appendix

Search terms based on the SPIDER tool (Sample, Phenomenon of Interest, Design, Evaluation, and Research type; Cooke, Smith, & Booth, 2012) entered in EBSCO databases and Scopus. Search conducted in title, abstract, key terms, and/or topic.

Sample	"early education" OR "early childhood education" OR "early childhood education and care" OR ecec OR "child care" OR childcare OR preschool* OR kindergarten* OR "center-based child care" OR "center-based childcare" OR "center-based programs" OR daycare OR "day care" OR preschooler* OR kindergartener* OR "three year*-old*" OR "3 year*-old*" OR "four year*-old*" OR "4 year*-old*" OR "five year*-old*" OR "5 year*-old*" OR "3-to-5-year* old*" OR "age* between three and five" OR "age* between 3 and 5" OR "age* 3" OR "age* 4" OR "age* 5" AND teacher* OR professional* OR adult* OR educator* OR caregiver*
Phenomenon of interest	AND "group composition" OR "group characteristics" OR "group level" OR "classroom level" OR "class level" OR "classroom composition" OR "class composition" OR "classroom characteristics" OR "class characteristics" OR "ethnic* composition" OR "ethnic* group composition" OR "ethnic* classroom composition" OR "ethnic* class composition" OR "group ethnic* composition" OR "classroom ethnic* composition" OR "class ethnic* composition" OR "sociocultural composition" OR "sociocultural group composition" OR "sociocultural classroom composition" OR "sociocultural class composition" OR "group sociocultural composition" OR "classroom sociocultural composition" OR "class sociocultural composition" OR "cultural composition" OR "cultural group composition" OR "cultural classroom composition" OR "cultural class composition" OR "group cultural composition" OR "classroom cultural composition" OR "class cultural composition" OR "racial composition" OR "racial group composition" OR "racial classroom composition" OR "racial class composition" OR "group racial composition" OR "classroom racial composition" OR "class racial composition" OR "socioeconomic status composition" OR "socioeconomic composition" OR "socio-economic status composition" OR "socio-economic composition" OR "SES composition" OR "socioeconomic status group composition" OR "socio-economic status group composition" OR "SES group composition" OR "socio-economic status classroom composition" OR "SES classroom composition" OR "socioeconomic status class composition" OR "socio-economic status class composition" OR "SES class composition" OR "group socioeconomic status composition" OR "group

socioeconomic status composition" OR "group socio-economic status composition" OR "group SES composition" OR "classroom socioeconomic status composition" OR "classroom socio-economic status composition" OR "classroom SES composition" OR "class socioeconomic status composition" OR "class socio-economic status composition" OR "class SES composition" OR “socioeconomic status average” OR “socio-economic status average” OR “SES average” OR “heterogeneous group*” OR "heterogeneous classroom*" OR "heterogeneous class*" OR “group heterogeneity” OR "classroom heterogeneity" OR "class heterogeneity" OR "heterogeneity in classroom*" OR "heterogeneity in class*" OR "heterogeneity in group*" OR "homogeneous classroom*" OR "homogeneous class*" OR "classroom homogeneity" OR "class homogeneity" OR "homogeneity in classroom*" OR "homogeneity in class*" OR "homogeneous group*" OR "group homogeneity" OR "homogeneity in group*" OR "group diversity" OR "diversity in group*" OR "diverse group*" OR "diversity within group*" OR "classroom diversity" OR "diversity in classroom*" OR "diverse classroom*" OR "diversity within classroom*" OR "class diversity" OR "diversity in class*" OR "diverse class*" OR "diversity within class*" OR "ethnic* divers*" OR "sociocultural* divers*" OR "cultural* divers*" OR "socioeconomic* divers*" OR "socio-economic* divers*" OR "SES divers*" OR “proportion of minority” OR “percentage of minority” OR “ratio of minority” OR "proportion of ethnic* minorit*" OR "percentage of ethnic* minorit*" OR "ratio of ethnic* minorit*" OR "proportion of sociocultural minorit*" OR "percentage of sociocultural minorit*" OR "ratio of sociocultural minorit*" OR "proportion of cultural minorit*" OR "percentage of cultural minorit*" OR "ratio of cultural minorit*" OR "proportion of racial minorit*" OR "percentage of racial minorit*" OR "ratio of racial minorit*" OR "proportion of language minority" OR "percentage of language minority" OR "ratio of language minority" OR “proportion of bilingual*” OR “percentage of bilingual*” OR “ratio of bilingual*” OR "proportion of dual language learners" OR "percentage of dual language learners" OR "ratio of dual language learners" OR "proportion of DLL" OR "percentage of DLL" OR "ratio of DLL" OR "proportion of English language” OR "percentage of English language” OR "ratio of English language” OR "proportion of non-native speakers” OR "percentage of non-native speakers” OR "ratio of non-native speakers” OR "proportion of native speakers” OR "percentage of native speakers” OR “ratio of native speakers” OR “proportion of disadvantaged” OR “percentage of disadvantaged” OR “ratio of disadvantaged” OR "proportion of children in disadvantage*" OR "percentage of children in disadvantage*" OR "ratio of children in disadvantage*" OR "proportion of at-risk children" OR "percentage of at-risk children" OR "ratio of at-risk children" OR "proportion of children at-risk" OR "percentage of children at-risk" OR "ratio of children at-risk" OR "proportion of children low-income" OR "percentage of children low-income" OR "ratio of children low-income" OR "proportion of children low income" OR "percentage of children low income"

	OR "ratio of children low income" OR "proportion of poor children" OR "percentage of poor children " OR "ratio of poor children " OR "proportion of children poverty" OR "percentage of children poverty" OR "ratio of children poverty" OR "proportion of immigrant* children" OR "percentage of immigrant* children" OR "ratio of immigrant* children" OR "proportion of migrant* children" OR "percentage of migrant* children" OR "ratio of migrant* children" OR "proportion of children from immigrant famil*" OR "percentage of children from immigrant famil*" OR "ratio of children from immigrant famil*" OR "proportion of non-white children" OR "percentage of non-white children" OR "ratio of non-white children" OR "proportion of white children" OR "percentage of white children" OR "ratio of white children" OR "herfindal index" OR "composition index*"
Evaluation	AND "observed relation*" OR "observed interaction*" OR "observed practice*" OR "observation measures" OR "observation* of" OR "class* observation*" OR "process quality" OR "classroom organization" OR "instructional support" OR "emotional support" OR "Assessment Profile for Early Childhood Programs" OR APECP OR "Classroom Assessment of Supports for Emergent Bilingual Acquisition" OR "CASEBA" OR "Child Caregiver Interaction Scale" OR CCIS OR "Arnett Caregiver Interaction Scale" OR CIS OR "Classroom Assessment Scoring System" OR CLASS OR "Classroom Language and Literacy Environment Observation" OR CLEO OR "Caregiver Observation Form and Scale" OR COFAS OR "Classroom Practices Inventory" OR CPI OR "Early Childhood Classroom Observation Measure" OR ECCOM OR "The Early Childhood Environment Rating Scale" OR ECERS* OR "Early Literacy Observation Tool" OR "E-LOT" OR "Observation Measures of Language and Literacy" OR OMLIT OR "The Preschool Classroom Implementation Rating Scale" OR PCI OR "Preschool Mental Health Climate Scale" OR PMHCS OR "Preschool Program Quality Assessment" OR PQA OR "Preschool Rating Instrument for Science and Math" OR PRISM OR "Quality of Early Childhood Care Settings: Caregiver Rating Scale" OR QUEST OR "Ramey and Ramey Observation of Learning Essentials" OR ROLE OR "Teacher Behavior Rating Scale" OR TBRS OR "Teacher Instructional Engagement Scale" OR TIES OR "Teaching Pyramid Observation Tool for Preschool Classrooms" OR TPOT OR "Teaching style rating system" OR TSRS

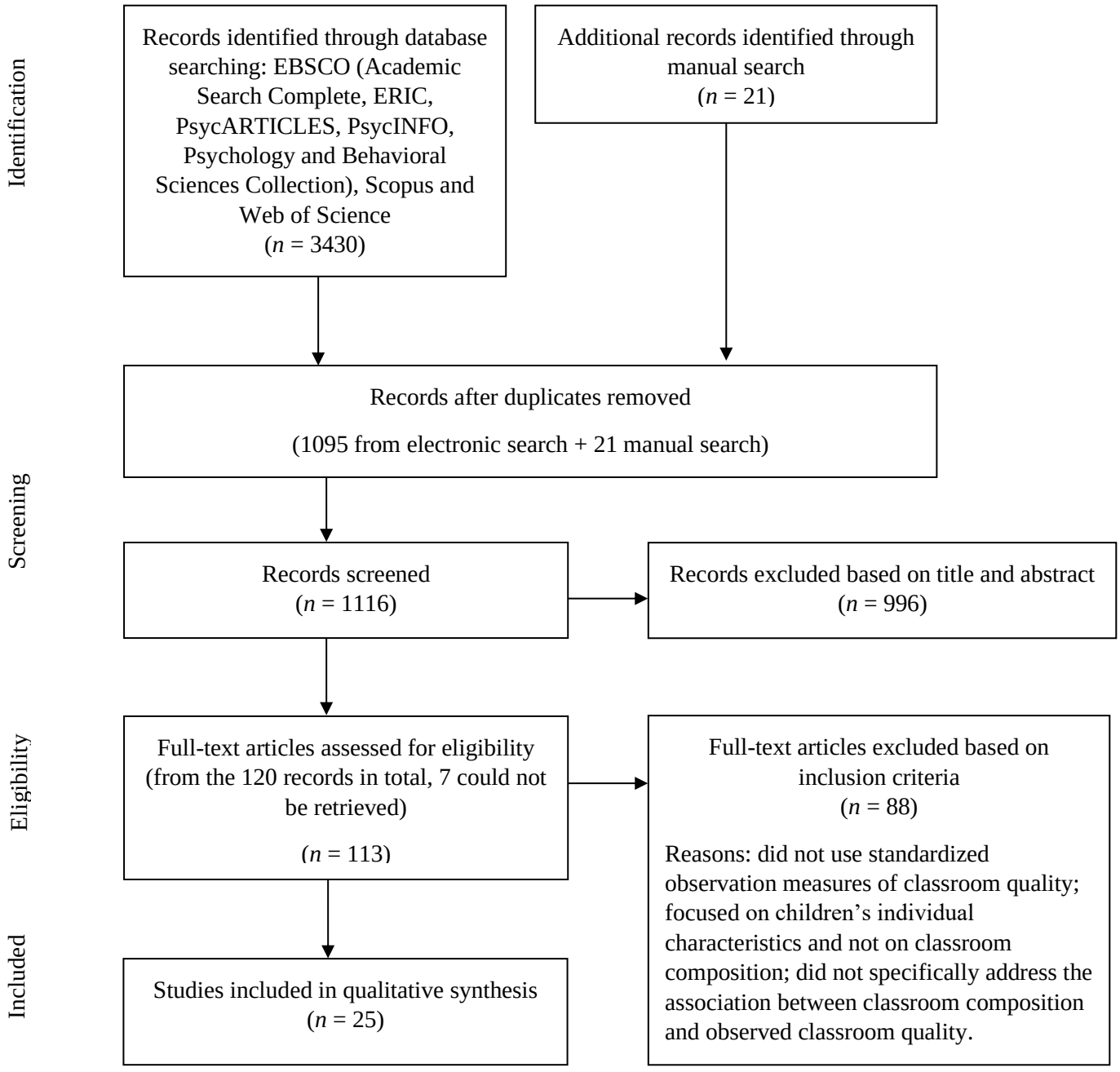

Fig. 1. Results of the search strategy based on the PRISMA statement (Liberati et al., 2009).

Table 1

Description of standardized observation measures of classroom quality by type of quality

Type of quality	Measure	Domains /Subscales	Author/year
Global	Assessment of Practices in Early Elementary Classrooms (APEEC)	16 items (e.g., room arrangement, accessibility, use of resources, teacher-child language, instructional methods, children participation)	Hemmeter, Maxwell, Ault, and Schuster, 2001
	Early Childhood Rating Scale (Revised) (ECERS-R)	7 subscales: space and furnishings, personal care, language and reasoning, activities, interactions, program structure, parents/staff	Harms, Clifford, and Cryer, 1998
Content specific	Early Language and Literacy Classroom Observation (ELLCO)	3 components: literacy environment checklist, classroom observation, and literacy activities rating scale	Castro, 2005
Process	Assessing School Settings: Interactions of Students and Teachers (ASSIST)	5 subscales: teacher's control, anticipation and responsiveness, monitoring, proactive behavior management, teacher/student meaningful participation	Rusby, Taylor, and Milchak, 2001
	Caregiver Interaction Scale (CIS)	4 dimensions: teacher's emotional tone, discipline style, and responsiveness to children	Arnett, 1989
	Classroom Assessment Scoring System (CLASS)	3 domains: emotional support, classroom organization, instructional support	Pianta, La Paro, and Hamre, 2008
	Early Childhood Classroom Observation Measure (ECCOM)	3 subscales of constructivist and didactic practices: instruction, management, social climate	Stipek and Byler, 2004
	Eco-behavioral System for the Complex Assessment of Preschool Environments (ESCAPE)	5 categories of teacher behavior: approval, disapproval, verbal prompting, verbal instruction, no response	Carta, Greenwood, and Atwater, 1992

Table 2

Summary of studies on the association between classroom ability composition and observed classroom quality

Authors/ Year	Study characteristics	Measures	Results	Covariates
Iruka and Morgan (2014)	**Country:** USA **Theoretical framework:** Ecological theory (Bronfenbrenner & Morris, 2007) **Sample** ▪ Program: Head Start, preschool, public school prekindergarten ▪ Teachers/Classroom: 14 children *per* teacher ▪ Children: *n* = 350 preschoolers Mean age 53 months All African-American 48% boys 53% living below (150%) poverty line **Study design:** Cross-sectional **Data set:** ECLS-BC	**Classroom composition index** Proportion of children with IEP *M* = between .12 and .18 **Observation measure of quality** ECERS-R and CIS combined in three quality profiles: 1- Moderately High and Sensitive Interactions (52% of classrooms) 2- Average and Sensitive Interactions (35% of classrooms) 3- Low and Harsh Interactions (13% of classrooms)	No associations	▪ Teacher: education, experience, age, enjoyment of job, intrinsic motivation, professional development opportunities ▪ Classroom: size, % of non-English speakers ▪ Children: age, gender, income-to-needs ratio, maternal education, family structure
Justice, Mashburn, Hamre, and Pianta (2008)	**Country:** USA **Theoretical framework: S**ocial–interactionist theories of language acquisition (e.g., Baumwell, Tamis-LeMonda, & Bornstein, 1997) **Sample** ▪ Program: State-funded preschool, serving at-risk (social/economic) 4-year-old children ▪ Teachers/Classroom: 135 teachers ▪ Children: 350 preschoolers 46% African-American, 29% Caucasian, 12% Hispanic/Latino, 13% other ethnicity 1 out of 5 DLL 50% girls Average income 26.500$ **Study design:** Cross-sectional **Data set:** ECLS-BC	**Classroom composition index** % children with IEP *M* = 9 (*SD* = 16.6) Range 1-100 **Observation measure of quality** CLASS – IS (scale of language modeling) CLASS – IS (scale of literacy focus)	Positive association with literacy focus, no association with language modeling	▪ Teacher: procedural fidelity (routine, teaching), education, participation in language and literacy workshops, years of experience, self-efficacy, teacher-centeredness ▪ Classroom: % with low English proficiency, number of children participating, language lesson

Maxwell, McWilliam, Hemmeter, Ault, and Schusterb (2001)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Public elementary schools ▪ Teachers/Classroom: 69 k-3rd grade (12 K) ▪ Children: 350 preschoolers Mean age 53 months All African-American 48% boys 53% living below (150%) poverty line **Study design:** Cross-sectional	**Classroom composition index** Number of children with disabilities *M* = 3 children per class Range 1-7 **Observation measure of quality** APEEC	No association	▪ Classroom: grade, size
Sontag (1997)	**Country:** USA **Theoretical framework:** Ecological theory (Bronfenbrenner, 1979, 1992) **Sample** ▪ Program: Public school, community-based, Head Start ▪ Teachers/Classroom: 8 teachers, 6 classrooms ▪ Children: Subgroups of 8 preschoolers with disabilities in integrated classrooms and 8 in segregated classrooms Mean age 55.75 and 58.63, respectively 9 White, 6 Black, 1 Hispanic 13 boys 13 from low-income families **Study design:** Longitudinal (multiple measurements of classroom quality over 6 months)	**Classroom composition index** Two subgroups of children with IEP's **Observation measure of quality** ESCAPE	*Positive association in integrated classrooms with more teachers' disapprovals of children's behavior	▪ None

Note. IEP = Individualized Education Plan; ECERS-R = Early Childhood Rating Scale Revised; CIS = Caregiver Interaction Scale; CLASS- IS = Classroom Assessment Scoring System – Instructional Support; APEEC = The Assessment of Practices in Early Elementary Classrooms; ESCAPE = Eco-behavioral System for the Complex Assessment of Preschool Environments; ECLS-BC = Early Childhood Longitudinal Study, Birth Cohort; DLL = Dual Language Learners.
*Correlations calculated based on *M* and *SD* values provided by the authors.

Table 3

Summary of studies on the association between classroom age composition and observed classroom quality

Authors/ Year	Study characteristics	Measures	Results	Covariates
Ansari and Pianta (2018)	**Country:** USA **Theoretical framework:** Sociocultural theory (Vygotsky, 1978); Social-learning theory (Bandura, 1986) **Sample** ▪ Program: Head Start (61%), public schools (40%) ▪ Teachers/Classroom: 332 preschool teachers (around 50% participated in a 2-year intervention, 50% in the control group) ▪ Children: 8% with disabilities 3-5-year-olds 16% with limited English proficiency 48% girls Average 1.4 income-to-needs ratio **Study design:** Longitudinal (4 measurements of classroom composition and quality over 2 years) **Data set:** NCRECE intervention program	**Classroom composition index** Age diversity Low diversity (73% of 4-year-olds; 8% of 3-year-olds; 19% of 5-year-olds) Moderate diversity (53% of 4-year-olds; 30% of 3-year-olds; 25% of 5-year-olds) High diversity (44% of 4-year-olds; 30% of 3-year-olds; 25% of 5-year-olds) **Observation measure of quality** CLASS - ES CLASS - CO CLASS - IS	Negative associations (all scales) Increase in age diversity associated with lower CO in the beginning of Year 2 and with lower ES at the end of the year.	▪ Teacher: ethnicity, gender, age, years at current program, income-to-needs ratio, works in pre-K vs Head Star, participated in intervention vs control ▪ Classroom: size, average income-to-needs ratio, racial/ethnic diversity, % girls, % with disabilities, % with limited English
Kuger, Kluczniok, Kaplan, and Rossbach (2016)	**Country:** Germany **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Public, private for-profit and not-for profit ▪ Teachers/Classroom: 97 kindergarten classrooms ▪ Children: Mean age 4.7 years 20% with migration background **Study design:** Longitudinal (3 measurements of classroom composition and quality over 3 years)	**Classroom composition index** Classroom mean age Year 1 *M*= 4.6 (*SD* = .4) Year 2 *M*= 4.7 (*SD* = .3) Year 3 *M*= 4.8 (*SD* = .3) **Observation measure of quality** ECERS-R total score	Positive association	▪ Teacher: experience, satisfaction ▪ Classroom: space per child, teacher-child ratio, class size, proportion of children with migration background, number of adults, number of teacher changes

	Data set: BiKs-3-10			
Pakarinen et al. (2010)	**Country:** Finland **Theoretical framework:** Theoretical three-factor model of classroom quality **Sample** ▪ Program: Day care, elementary schools ▪ Teachers/Classroom: 49 teachers and kindergarten classrooms ▪ Children: around 11 *per* class Majority 6-year-olds **Study design:** Cross-sectional **Data Set:** First Steps Study—Interaction and Learning Within the Child–Parent–Teacher Triangle	**Classroom composition index** Number of 6-year-olds *M* = 13.85 (*SD* =5.92) Range 3-24 **Observation measure of quality** CLASS – ES CLASS – CO CLASS - IS	No associations	▪ None
Purtell and Ansari (2018)	**Country:** USA **Theoretical framework:** Bioecological model of human development (Bronfenbrenner & Morris, 2006) **Sample** ▪ Program: Head Start ▪ Teachers/Classroom: 486 classrooms ▪ Children: *n* = 2829 Aged 3 and 4 Proportion of .20 White children 50% girls Proportion of .52 unemployed mothers **Study design:** Cross-sectional	**Classroom composition index** Proportion of 3-year-olds vs 4-year-olds *M* = .59 (*SD* = .31) 3-year-olds *M* = .22 (*SD* = .21) 4-year-olds **Observation measure of quality** CLASS total score	No association	▪ Program: average hours per week ▪ Teacher: depressive symptomology, education, benefits, hourly salary ▪ Classroom: teacher-child ratio, adult-child ratio, size, language (English only vs. English and Spanish) ▪ Children: gender, race/ethnicity, age at the start, months between the fall and spring assessments, mothers' - education, age, employment status, marital status, depressive symptoms; ratio of income to poverty, household size and language
Slot, Bleses, Justice, Markussen-Brown,	**Country:** Denmark **Theoretical framework:** Bioecological model of human development (Bronfenbrenner & Morris, 2006) **Sample**	**Classroom composition index** Mean age *M* = 56.67 (*SD* = 6.91) Range 42-70.2	No associations	▪ None

and Højen (2018)	▪ Program: Preschool (centers chosen to overrepresent high concentrations of children at risk - social disadvantage and non-Danish background) ▪ Teachers/Classroom: 402 teachers, 260 classrooms ▪ Children: *n* = 5359 Aged 4-6 89% monolingual (Danish) 53% girls **Study design:** Part of a randomized control trial	**Observation measure of quality** CLASS – ES CLASS – CO CLASS – IS (pre-intervention scores)

Note. ECERS-R = Early Childhood Rating Scale Revised; CLASS = Classroom Assessment Scoring System; CLASS-ES = Emotional support; CLASS-CO = Classroom organization; CLASS-IS = Instructional support; NCRECE = National Center for Research on Early Childhood Education.

Table 4

Summary of studies on the association between classroom gender composition and observed classroom quality

Authors/ Year	Study characteristics	Measures	Results	Covariates
Broekhuizen, Slot, van Aken, & Dubas (2017)	**Country:** Netherlands **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Preschool, child care ▪ Teachers/Classroom: 37 classrooms ▪ Children: *n* = 113 Mean age 37 months 70% monolingual (Dutch) **Study design:** Cross-sectional **Data set:** pre-COOL study	**Classroom composition index** % girls *M* = 47.8% **Observation measure of quality** CLASS (Toddler) – Emotional and Behavioral Support	No association	▪ None
Debnam, Pas, Bottiani, and Cash (2015)	**Country:** USA **Theoretical framework:** Culturally responsive teaching (Gay, 2000) **Sample** ▪ Program: Elementary schools ▪ Teachers/Classroom: 142 K-8th grade teachers ▪ Children: 58% from ethnic minorities 50% boys **Study design:** Cross-sectional **Data set:** CSRP and FOL	**Classroom composition index** % male No information on % at the classroom level **Observation measure of quality** ASSIST Cultural Responsiveness Teaching	No association	▪ Classroom: total number of children, % White children
Friedman-Krauss et al. (2014)	**Country:** USA **Theoretical framework:** Transactional model of coercive cycles of adult–child conflict; (Snyder, Cramer, Afrank, & Patterson, 2005); Transactional model of stress and coping (Lazarus, 1991) **Sample** ▪ Program: Head Start (in high-poverty neighborhoods)	**Classroom composition index** % male No information on % at the classroom level **Observation measure of quality** CLASS - EC	No association	▪ Teacher: gender, primary income earner for family or not, psychological distress, job stress in spring ▪ Classroom: number of children present in fall observations, race/ethnicity, average income-to-needs ratio, teacher-reported externalizing behavior

- Teachers/Classroom: 42 teachers, 17 classrooms in control group
- Children: *n* = 262
  63% Black children
  48% boys
  Average 1.17 income-to-needs ratio

Study design: Longitudinal (2 measurements of classroom quality over 1 school year)
Data set: CSRP and FOL

Note. CLASS = Classroom Assessment Scoring System; CLASS-ES = Emotional support/CLASS-EC= Emotional climate; CLASS-CO = Classroom organization; CLASS-IS = Instructional support; ASSIST= Assessing School Settings: Interactions of Students and Teachers; CSRP = Chicago School Readiness Project; FOL = Foundations of Learning Demonstration in Newark.

Table 5

Summary of studies on the association between classroom minority status composition and observed classroom quality

Authors/ Year	Study characteristics	Measures	Results	Covariates
Bassok and Galdo (2016)	**Country:** USA **Theoretical framework:** Input effects on bilingual language development (Unsworth, 2016) **Sample** ▪ Program: Pre-K, serving around 55% at-risk children ▪ Teachers/Classroom: 3883 classrooms 19% in poverty (based on zip code) 6 % Hispanic, 27% Black (based on zip code) **Study design:** Cross-sectional	**Classroom composition index** 1 – % Hispanic 3 Quartiles: Low *M* = 2.1; Middle *M* = no value given; High *M* = 6.9 2 - % Black 3 Quartiles: Low *M* = 9.5; Middle *M* = no value given; High *M* = 40.2 **Observation measure of quality** CLASS – ES CLASS – CO CLASS – IS	1- No associations 2- Lower quality in all scales in higher quartile compared with lower quartile	▪ None
Bihler, Agache, Shneller, Willard, and Leyendecker (2018)	**Country:** Germany **Theoretical framework:** Input effects on bilingual language development (Unsworth, 2016) **Sample** ▪ Program: Preschool ▪ Teachers/Classroom: 169 classrooms ▪ Children: *n* = 903 Mean age 40.37 months 22% from low-income families **Study design:** Cross-sectional	**Classroom composition index** % DLLs (German – Other language) *M* = 27.78 (*SD* = 17.9) **Observation measure of quality** CLASS - total	*No association	▪ None
Broekhuizen, Slot, van Aken, & Dubas (2017)	**Country:** Netherlands **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Preschool, child care	**Classroom composition index** Proportion of non-Dutch children	Negative association	▪ None

	▪ Teachers/Classroom: 37 classrooms ▪ Children: *n* = 113 Mean age 37 months 70% monolingual (Dutch) 51% boys **Study design:** Cross-sectional **Data set:** pre-COOL study	45.9% majority Dutch; 21.6% close mix of Dutch and non-Dutch; 32.4% majority non-Dutch **Observation measure of quality** CLASS (Toddler) – Emotional and Behavioral Support		
Debnam, Pas, Bottiani, and Cash (2015)	**Country:** USA **Theoretical framework:** Culturally responsive teaching (Gay, 2000) **Sample** ▪ Program: Elementary schools ▪ Teachers/Classroom: 142 K-8th grade teachers ▪ Children: 58% from ethnic minorities 50% boys **Study design:** Cross-sectional **Data set:** CSRP and FOL	**Classroom composition index** % White children No information on % at the classroom level **Observation measure of quality** ASSIST Cultural Responsiveness Teaching	No association	▪ Classroom: total number of children, % boys
Dotterer et al. (2014)	**Country:** USA **Theoretical framework:** Ecological theory (Bronfenbrenner & Ceci, 1994) **Sample** ▪ Program: Pre-K, target and universal ▪ Teachers/Classroom: 716 classrooms, 76% in targeted programs (64% poor classrooms), 24% in universal programs (41% poor classrooms) ▪ Children: 4-years-old Around 50% boys **Study design:** Longitudinal (2 measurements of classroom quality over 1 school year) **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** % White children 51% in Universal 38% in Targeted **Observation measure of quality** ECERS-R LI ECERS-R PL CLASS - EC CLASS - IC	Lower ECERS scores in Universal compared with Targeted classrooms No association with CLASS – EC Lower CLASS - IC in Universal compared with Targeted classrooms	▪ Program: hours per day ▪ Teacher: education ▪ Classroom: % children living below (150%) poverty line, teacher-child ratio

Downer et al. (2012)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K, 16% Head Start ▪ Teachers/Classroom: 721 classrooms ▪ Children: *n* = 2983 49% boys 60% non-Caucasian 59% living below (150%) poverty line **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** % DLLs (English-Spanish) No DLL = 0 (48.3 % of classrooms) Mid DLL = proportion below 50% (35.5% of classrooms) Hi-DLL = proportion above 50% (14.2% of classrooms) **Observation measure of quality** CLASS - total	No association	▪ None
Fram and Kim (2012)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Kindergarten ▪ Teachers/Classroom: 1500 classrooms ▪ Children: *n* = 2983 51% girls *M* = 57 months 55% White 25% living below (150%) poverty line **Study design:** Cross-sectional **Data set:** ECLS-BC	**Classroom composition index** 1- Predominantly Latino/Hispanic children (around 70% in class) 2- Predominantly Black children (around 80% in class) 3- "Other" predominant (around 70% in class) 4- Predominantly White children (around 80% in class) **Observation measure of quality** CIS	1, 4- No association 2, 3- Negative association	▪ Children: gender, birthweight, age in months, age at entry, hours per week in ECE; parent marital status, work status, number of siblings, number of adults in house, parent education, poverty status, annual household income and location
Friedman-Krauss et al. (2014)	**Country:** USA **Theoretical framework:** Transactional model of coercive cycles of adult–child conflict (Snyder, Cramer, Afrank, & Patterson, 2005); Transactional model of stress and coping (Lazarus, 1991) **Sample** ▪ Program: Head Start (in high-poverty neighborhoods)	**Classroom composition index** % Black children No information on averages at the classroom level **Observation measure of quality** CLASS - EC	Negative association	▪ None

	▪ Teachers/Classroom: 42 teachers, 17 classrooms in control group ▪ Children: *n* = 262 63% Black children 48% boys Average 1.17 income-to-needs ratio **Study design:** Longitudinal (2 measurements of classroom quality over 1 school year) **Data set:** CSRP and FOL			
Iruka and Morgan (2014)	**Country:** USA **Theoretical framework:** Ecological theory (Bronfenbrenner & Morris, 2007) **Sample** ▪ Program: Head Start, preschool, public school prekindergarten ▪ Teachers/Classroom: *M* = 14.18 (*SD* = 5.11) children per teacher ▪ Children: *n* = 350 preschoolers Mean age 53 months All African-American 48% boys 53% living below (150%) poverty line **Study design:** Cross-sectional **Data set:** ECLS-BC	**Classroom composition index** % non-English-speaking children (i.e., Spanish speakers) **Observation measure of quality** ECERS-R with CIS to create classroom quality profiles: 1- Moderately High and Sensitive Interactions (52%) 2- Average and Sensitive Interactions (35%) 3- Low and Harsh Interactions (13%)	Higher percentage in classrooms that fit both the lower and higher quality profiles	▪ Teacher: education, experience, age, enjoyment of job, intrinsic motivation, professional development opportunities ▪ Classroom: size, % of non-English speakers ▪ Children: age, gender, income-to-needs ratio, maternal education, family structure
Justice, Mashburn, Hamre, and Pianta (2008)	**Country:** USA **Theoretical framework: S**ocial–interactionist theories of language acquisition (e.g., Baumwell, Tamis-LeMonda, & Bornstein, 1997) **Sample** ▪ Program: State-funded preschool, serving at-risk (social/economic) 4-year-old children ▪ Teachers/Classroom: 135 teachers ▪ Children: *n* = 350 preschoolers	**Classroom composition index** % with low English proficiency *M* = 12.7 (*SD* = 25.8) Range 1-100 **Observation measure of quality** CLASS – IS (scale of language modeling)	No associations	▪ Teacher: procedural fidelity (routine, teaching), education, participation in language and literacy workshops, years of experience, self-efficacy, teacher-centeredness ▪ Classroom: % with low English proficiency, number of participating children, language lesson

	46% African-American, 29% Caucasian, 12% Hispanic/Latino, 13% other ethnicity 1 out of 5 DLL 50% girls Average income 26.500$ **Study design:** Cross-sectional **Data set:** ECLS-BC	CLASS – IS (scale of literacy focus)		
Kuger, Kluczniok, Kaplan, and Rossbach (2016)	**Country:** Germany **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Public and private for-profit and not-for profit ▪ Teachers/Classroom: 97 kindergarten classrooms ▪ Children: Mean age 4.7 years 20% of children with migration background **Study design:** Longitudinal (3 measurements of classroom composition and quality over 3 years) **Data set:** BiKs-3-10	**Classroom composition index** Proportion of children with low proficiency in the German language Year 1 *M*= 18.3 (*SD* = 18.5) Year 2 *M*= 20.1 (*SD* = 24.6) Year 3 *M*= 21.9 (*SD* = 23.1) **Observation measure of quality** ECERS-R total	Negative association Increase in the proportion of migrant children from year 1 to year 2 was associated with a decrease in ECERS scores	▪ Teacher: experience, satisfaction ▪ Classroom: space per child, teacher-child ratio, class size, proportion of children with migration background, number of adults, number of teacher changes
LoCasale-Crouch et al. (2007)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K, 15% Head Start ▪ Teachers/Classroom: 692 classrooms ▪ Children: *n* = 2800 Majority of 4-year-olds 58% non-Caucasian Around 50% boys 58% living below (150%) poverty line **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** Proportion of non-Caucasian *M* = between .49 and .73 **Observation measure of quality** CLASS divided into 5 quality profiles: 1- highest quality 2- positive EC, high IC 3- positive EC, mediocre IC 4- mediocre EC, low IC 5- poorest quality	Higher proportion in poorest quality classrooms	▪ None

Sanders and Downer (2012)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K ▪ Teachers/Classroom: 692 classrooms ▪ Children: around 17 per class 48% boys Proportion of .58 living below (150%) poverty line **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** Bilingual (vs. non-bilingual) classrooms (English-Spanish) *M*= .33 bilingual classrooms **Observation measure of quality** ECERS-R AD CLASS – EC CLASS – IC	Positive association with ECERS-R AD No association with CLASS	▪ None
Sawyer et al. (2016)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Federally and state-funded preschool serving children from low-income families ▪ Teachers/Classroom: 62 teachers ▪ Children: around 18 *per* class **Study design:** Part of a randomized control trial	**Classroom composition index** % DLL enrollment (English-Spanish) *M* = 50% (*SD* = 22%) **Observation measure of quality** ELLCO-DLL	No association	▪ None
Slot, Bleses, Justice, Markussen-Brown, and Højen (2018)	**Country:** Denmark **Theoretical framework:** Bioecological model of human development (Bronfenbrenner & Morris, 2006) **Sample** ▪ Program: Preschool (centers chosen to overrepresent high concentrations of children at risk - social disadvantage and non-Danish background) ▪ Teachers/Classroom: 402 teachers, 260 classrooms with overrepresentation of children ▪ Children: *n* = 5359 Aged 4-6	**Classroom composition index** Proportion of non-Danish children *M* = .11 (*SD* = .22) Range 0-1 **Observation measure of quality** CLASS – ES CLASS – CO CLASS – IS (pre-intervention scores)	Negative associations (all scales)	▪ None

	89% monolingual (Danish) 53% girls **Study design:** Part of a randomized control trial			
Stipek (2004)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Private (*n*= 12) and public schools (*n*= 142) ▪ Teachers/Classroom: 314 K-2nd classrooms, 109 kindergarten ▪ Children: 46% African-American and Latino/Hispanic 51% living in poverty **Study design:** Cross-sectional	**Classroom composition index** 1- % African-American children 2- % Latino/Hispanic No information on % at the classroom level **Observation measure of quality** ECCOM - CT ECCOM - DT	1, 2- Negative associations with CT and positive with DT	▪ Teacher: goals (basic skills, higher-order thinking, social skills), perceptions of family (challenges, barriers)
Valentino (2018)	**Country:** USA **Theoretical framework:** Link between state policy, classroom structure, process, and child outcomes **Sample** ▪ Program: Pre-K serving mostly children from low-income families ▪ Teachers/Classroom: 647 classrooms ▪ Children: *n* = 12334 Proportion of .54 living in poverty **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** 1-Group of DLL children (vs. non-DLL) (English-Spanish) Proportion *M* = .21 (*SD* = .32) DLL 2- Group of Black children (vs. White children) Proportion *M* = .19 (*SD* = .30) Black (vs *M* = .40 [*SD* = .37] White) 3- Group of Hispanic children (vs White children) Proportion *M* = .27 (*SD* = .35) Hispanic **Observation measure of quality** ECERS-R total ECERS-R LI ECERS-R PL CLASS - total CLASS - EC	1, 2, 3 - Negative associations with all ECERS scales 1 – No associations with CLASS 2, 3 - Negative associations with all CLASS scales	▪ None

CLASS - IC

Note. DLL = Dual Language Learners; CIS = Caregiver Interaction Scale; ECERS-R = Early Childhood Rating Scale Revised; ECERS-R LI = Language/interactions; ECERS-R PV = Provisions for learning; ECERS-R AD = Acceptance of diversity; CLASS = Classroom Assessment Scoring System; CLASS-ES = Emotional support/ EC= Emotional climate; CLASS-CO = Classroom organization; CLASS-IS = Instructional support/ IC = Instructional climate; ELLCO-DLL= Early Language and Literacy Classroom Observation; ASSIST= Assessing School Settings: Interactions of Students and Teachers; ECCOM = Early Childhood Classroom Observation Measure; ECCOM-CT = Constructivist teaching; ECCOM-DT = Didactic teaching; ECLS-BC = Early Childhood Longitudinal Study, Birth Cohort; NCEDL Multi-State = National Center for Early Development and Learning's Multi-State Study of Pre-Kindergarten; SWEEP = Study of State-Wide Early Education Programs; CSRP = Chicago School Readiness Project; FOL = Foundations of Learning Demonstration in Newark; ECLS-BC = Early Childhood Longitudinal Study, Birth Cohort.; BiKS study = Early childhood cohort of the German longitudinal BiKS study.
*Correlation coefficients were provided by the authors of the study upon request.

Table 6

Summary of studies on the association between classroom socioeconomic composition and observed classroom quality

Authors/ Year	Study characteristics	Measures	Results	Covariates
Bassok and Galdo (2016)	**Country:** USA **Theoretical framework:** Input effects on bilingual language development (Unsworth, 2016) **Sample** ▪ Program: Pre-K, serving around 55% at-risk children ▪ Teachers/Classroom: 3,883 classrooms, 19% poor (based on zip code), ▪ Children: 6 % Hispanic, 27% Black (based on zip code) **Study design:** Cross-sectional	**Classroom composition index** % children living in poverty by zip code 3 Quartiles: Low *M* = 11.4; Middle *M* = no value given; High *M* = 24.9 **Observation measure of quality** CLASS – ES CLASS – CO CLASS – IS	Lower CLASS – ES and IS in higher quartile compared with lower quartile No association with CLASS - CO	▪ None
Bihler, Agache, Shneller, Willard, and Leyendecker (2018)	**Country:** Germany **Theoretical framework:** Input effects on bilingual language development (Unsworth, 2016) **Sample** ▪ Program: Preschool ▪ Teachers/Classroom: 169 classrooms ▪ Children: *n* = 903 Mean age 40.37 months 22% from low-income families **Study design:** Cross-sectional	**Classroom composition index** % low SES children No information on % at the classroom level **Observation measure of quality** CLASS total score	*No association	▪ Classroom: % children from low income families, teacher-child ratio ▪ Children: age, duration of attendance
Dotterer et al. (2014)	**Country:** USA **Theoretical framework:** Ecological theory (Bronfenbrenner & Ceci, 1994) **Sample** ▪ Program: Pre-K, target and universal, serving 4-year-olds	**Classroom composition index** % children living below the 150% poverty line 41% in Universal 64% in Targeted programs	Lower ECERS scores in Universal compared with Targeted classrooms	▪ Program: hours per day ▪ Teacher: education ▪ Classroom: % White children, teacher-child ratio

	▪ Teachers/Classroom: 716 classrooms, 76% in targeted programs (64% poor), 24% in universal programs (41% poor) ▪ Children: 50% boys **Study design:** Longitudinal (2 measurements of classroom quality over 1 school year) **Data set:** NCEDL Multi-State and SWEEP studies	**Observation measure of quality** ECERS-R LI ECERS-R PL CLASS - EC CLASS - IC	No association with CLASS – EC Lower CLASS - IC in Universal compared with Targeted classrooms	
Friedman-Krauss et al. (2014)	**Country:** USA **Theoretical framework:** Transactional model of coercive cycles of adult–child conflict; (Snyder, Cramer, Afrank, & Patterson, 2005); Transactional model of stress and coping (Lazarus, 1991) **Sample** ▪ Program: Head Start (in high-poverty neighborhoods) ▪ Teachers/Classroom: 42 teachers, 17 classrooms in control group ▪ Children: *n* = 262 63% Black 48% boys Average 1.17 income-to-needs ratio **Study design:** Longitudinal (2 measurements of classroom quality over 1 school year) **Data set:** CSRP and FOL	**Classroom composition index** Average income-to-needs ratio (based on number of people in household + federal poverty level) No information on averages at the classroom level **Observation measure of quality** CLASS - EC	Negative association	▪ None
LoCasale-Crouch et al. (2007)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K, 15% Head Start ▪ Teachers/Classroom: 692 classrooms ▪ Children: *n* = 2800	**Classroom composition index** 1- Mean level of maternal education *M* = between 12.3 and 13.5 years	1-Positive association with highest quality	▪ None

	Majority of 4-year-olds 58% non-Caucasian 50% boys 58% living below the 150% poverty line **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	2 - Proportion of children living below the poverty line *M* = between .59 and .65 **Observation measure of quality** CLASS divided into 5 quality profiles: 1- highest quality 2- positive EC, high IC 3- positive EC, mediocre IC 4- mediocre EC, low IC 5- poorest quality	2- Positive association with poorest quality	
Phillips, Gormley and Lowenstein (2009)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K, Head Start, serving 4-year-olds ▪ Teachers/Classroom: 77 Pre-K + 28 Head Start classrooms **Study design:** Cross-sectional	**Classroom composition index** % children in poverty below (130%) poverty line 63% in Pre-K; 95% in Head Start **Observation measure of quality** CLASS – ES CLASS – CO CLASS – IS	No associations	▪ Program: full-day (or half-day) ▪ Teacher: education, years of experience, type of curriculum used
Pianta et al. (2005)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K ▪ Teachers/Classroom: 238 teachers ▪ Children: around 7 3- and 4-year-olds per teacher **Study design:** Cross-sectional	**Classroom composition index** % poor children in classroom 54% classrooms with > 60% children living in poverty **Observation measure of quality** ECERS-R LI ECERS-R PL CLASS – ES CLASS – IS	Negative associations with all ECERS and CLASS scales	▪ Program: in-school, full-day (or half-day), state (location) ▪ Teacher: education, experience, traditional attitudes, depressive symptoms, wage ▪ Classroom: child-staff ratio

Reid and Ready (2013)	**Country:** USA **Theoretical framework:** Bioecological theory of human development (Bronfenbrenner, 2004) **Sample** ▪ Program: Pre-K, public schools, Head Start ▪ Teachers/Classroom: 704 classrooms, 51% with 2/3 of poor children ▪ Children: *n* = 2966 50% boys **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** Mean SES (average between family income and maternal education) 1- low 2- middle 3- high SES classrooms **Observation measure of quality** ECERS-R total CLASS - ES CLASS – IS	1 - Negative associations with both ECERS and CLASS – ES 2 - No associations 3 – Positive associations with both ECERS and CLASS – EC No associations with CLASS-IC	▪ None
Sanders and Downer (2012)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample** ▪ Program: Pre-K ▪ Teachers/Classroom: 692 classrooms ▪ Children: around 17 per class 48% boys **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** 1 - % children living below (150%) poverty line *M* = .58 (*SD* = .32) 2 - Mean of maternal education level *M* = 12.8 (*SD* = 1.39) **Observation measure of quality** ECERS-R AD CLASS – EC CLASS – IC	1-Negative association with ECERS AD and CLASS – EC 2-Positive association with ECERS AD and CLASS - EC No associations with CLASS - IC	▪ None
Stipek (2004)	**Country:** USA **Theoretical framework:** Not clearly stated **Sample**	**Classroom composition index** % eligible for free lunch	Negative association with	▪ Teacher: goals (basic skills, higher-order thinking, social skills),

	▪ Program: Private (*n*= 12) and public schools (*n*= 142) ▪ Teachers/Classroom: 314 K-2nd classrooms, 109 kindergarten ▪ Children: 46% African-American and Latino/Hispanic 51% living in poverty **Study design:** Cross-sectional	No information on % at the classroom level **Observation measure of quality** ECCOM - CT ECCOM - DT	CT and positive with DT	perceptions of family (challenges, barriers)
Valentino (2018)	**Country:** USA **Theoretical framework:** Link between state policy, classroom structure, process, and child outcomes **Sample** ▪ Program: Pre-K serving mostly children from low-income families ▪ Teachers/Classroom: 647 classrooms Children: *n* = 12334 Proportion of .19 Black, .40 White, .27 Hispanic Proportion of .21 DLL **Study design:** Cross-sectional **Data set:** NCEDL Multi-State and SWEEP studies	**Classroom composition index** Poor children (i.e., living below the 150% poverty line) vs. non-poor Proportion *M* = .54 (*SD* = .32) poor **Observation measure of quality** ECERS-R total ECERS-R LI ECERS-R PL CLASS - total CLASS - EC CLASS - IC	Negative associations (all scales)	▪ None

Note. ECERS-R = Early Childhood Rating Scale Revised; ECERS-R LI = Teaching/interactions; PV = Provisions for learning; CLASS = Classroom Assessment Scoring System; CLASS-ES = Emotional support/ CLASS-EC= Emotional climate; CLASS-CO = Classroom organization; CLASS-IS = Instructional support/ CLASS-IC = Instructional climate; ECCOM = Early Childhood Classroom Observation Measure; ECCOM-CT = Constructivist teaching; ECCOM-DT = Didactic teaching; NCEDL Multi-State = National Center for Early Development and Learning's Multi-State Study of Pre-Kindergarten; SWEEP = Study of State-Wide Early Education Programs; CSRP = Chicago School Readiness Project; FOL = Foundations of Learning Demonstration in Newark.

*Correlation coefficients provided by the authors of the study upon request.

Table 7

Summary of findings on the association between classroom composition and observed classroom quality by type of characteristic, index, and classroom quality observation measure

		APEEC	ESCAPE	ECERS-R total score	ECERS-R LI	ECERS-R PL	ECERS AD	ELLCO-DLL	ASSIST	CIS	CLASS total score	CLASS ES/EC	CLASS CO	CLASS IS/IC	CLASS LM	CLASS LF	ECCOM CT	ECCOM DT	ECERS X CIS
Ability	% with IEP's														●	□			●
	Number with disabilities	●																	
	Integrated classroom[1]		◘																
Age	Nº of 6-year-olds											●	●	●					
	Proportion of 3 (vs. 4-year-olds)										●								
	Age diversity											◘	◘	◘					
	Mean age			□								●	●	●					
Minority Status	%/Proportion DLLs			◘	◘	◘	□	●			●●●	●		●●					
	% migrant background (only speak home-language or low proficiency in the host country's language)			◘								◘	◘	◘	●◘				◘□
	%/Proportion Black/African-American			◘	◘	◘				◘			◘				◘	□	
	%/Proportion Latino/Hispanic			◘	◘	◘				●	◘◘◘●	●◘◘◘	◘	◘◘◘●			◘	□	
	%/Proportion non-Caucasian											◘		◘					
	%/Proportion White				◘	◘			●	●		●		●					
	%/Proportion "other" (than African-American, Hispanic, Caucasian)									◘									
Gender	% boys								●			●							
	% girls											●							
SES	%/Proportion living in poverty			◘	◘◘□	◘◘□	◘				◘●	◘◘◘◘ ◘●●	◘●●	◘◘◘□ ●●			◘	□	
	Mean SES			□◘●								◘□		●●●					
	Mean maternal education						□					□□		●					

Note.[1] The association was significant with teachers' disapproval behaviors but not with other teacher behaviors.
● = association was not significant; ◘ = association was negative; □ = association was positive.
APEEC = The Assessment of Practices in Early Elementary Classrooms; ECERS-R = Early Childhood Rating Scale Revised; ECERS-R LI = Language/interactions; ECERS-R PV = Provisions for learning; ECERS-R AD = Acceptance of diversity; ELLCO-DLL= Early Language and Literacy Classroom Observation; ASSIST= Assessing School Settings: Interactions of Students and Teachers; CIS = Caregiver Interaction Scale; CLASS = Classroom Assessment Scoring System; CLASS-ES/EC = Emotional support/climate; CLASS-CO = Classroom organization; CLASS-IS/IC = Instructional support/climate; CLASS-LM = Language Modeling; CLASS-LF = Literacy Focus; ECCOM = Early Childhood Classroom Observation Measure; ECCOM-CT = Constructivist teaching; ECCOM-DT = Didactic teaching; ESCAPE = Eco-behavioral System for the Complex Assessment of Preschool Environments.