

A systematic review of studies investigating quality of inclusive preschool classrooms

Lisa Butterworth¹, Salih Rakap², Emrah Gülboy³

Abstract: Inclusion of young children with disabilities in early childhood settings that are primarily designed for typically developing children is a recommended practice. Although several reviews have summarized the literature focused on the quality of preschool classrooms in general, extant literature does not include a study that specifically summarized results of studies focused on investigating quality of inclusive preschool classrooms. The purpose of this systematic literature review was to identify studies focused on examining quality of inclusive preschool classrooms, determine characteristic features of these studies and summarize information about the overall quality of inclusive preschool classrooms. Twenty-five articles met the inclusion criteria and were analyzed using a 27-item coding system developed by the researchers. The findings reveal a lack of consistent reporting of classroom demographics and teacher characteristics across the studies. Furthermore, the review identifies inconsistencies in reporting practices related to the roles and training of individuals conducting classroom quality measures, raising concerns about data reliability and validity. Additionally, the findings suggest areas for improvement in instructional support in inclusive classrooms, calling for strategies to enhance teacher training and professional development. The results underscore differences in preschool classroom quality across countries, emphasizing the necessity for global efforts and tailored interventions to improve early childhood education quality.

Article History

Received: 22 January 2024

Accepted: 22 October 2024

Keywords

Inclusive preschool; Early childhood quality; Early childhood inclusion

Introduction

Inclusion of young children with disabilities is a recommended practice in early childhood (EC) and early childhood special education (ECSE) and legally mandated by the Individuals with Disabilities Education Act in the United States (IDEA) (Individuals with Disabilities Education Act in the United States [IDEA], 2004; Musgrove, 2012). As a result of education policies and research findings supporting inclusion, the number of children with disabilities receiving educational and developmental services within the context of inclusive preschool settings has gradually increased over the years. According to the most recent annual report of the Office of Special Education Programs (OSEP) to the Congress, the percentage of children with disabilities who spend at least part of their week in regular early childhood programs increased from 50% in 2009 to 64.7% in 2019 (Office of Special Education and Rehabilitative Services, 2011; 2021).

Similarly, increasing access to EC and ECSE has been a priority in the European Union (EU) since 1992. Over the years, various policy adjustments have been made at the EU level to achieve this goal. While early EU policies were linked to labor market needs, especially regarding employability and equality, recent efforts have focused on reforms aimed at creating high-quality, fair, and efficient education and training systems. In 2009, the EU Council launched a strategic framework for European cooperation in education and training until 2020, with the objective of increasing participation in EC and ECSE for all

¹ University of North Carolina at Greensboro, School of Education, Department of Specialized Education Services, North Carolina, USA, e-mail: lbutterwor@uncg.edu, ORCID: <https://orcid.org/0000-0002-1608-0805>

² University of North Carolina at Greensboro, School of Education, Department of Specialized Education Services, North Carolina, USA, e-mail: s_rakap@uncg.edu, ORCID: <https://orcid.org/0000-0001-7853-3825>

³ Ondukuz Mayıs University, School of Education, Department of Special Education, Samsun, Türkiye, e-mail: emrah.gulboy@edu.tr, ORCID: <https://orcid.org/0000-0002-7802-6839>

individuals (European Agency for Special Needs and Inclusive Education [EASNIE], 2017). By 2014, 32 million children in Europe were eligible for the EC provisions of the European Commission, with approximately 15 million needing ECSE (European Commission, 2013; EASNIE, 2017). To address this need, the EASNIE (2024) has made it a mission to facilitate access to EC and ECSE, particularly for the most vulnerable groups—children with disabilities, migrants, and children at risk. The Agency's efforts focus on supporting educators, providing access to resources, fostering interdisciplinary cooperation, and promoting parental collaboration. EASNIE also advocates for the accountable use of public funding and leadership models to ensure that EC and ECSE services are both high quality and widely accessible, focusing on the holistic development of every child (EASNIE, 2024).

Although there is not a commonly accepted or agreed upon definition of inclusion in the field of EC and ECSE, the Division of Early Childhood (DEC) and the National Association for the Education of Young Children (NAEYC) described defining feature of early childhood inclusion in their joint position statement as (a) access, (b) participation and (c) supports. *Access* refers to children with and without disabilities being able to access all learning opportunities and environments of the early childhood program. *Participation* refers to teachers making the accommodations needed for individual needs of the children to encourage active engagement in the early childhood program. *Supports* refers to resources and professional development provided to form a strong system of support for children who need additional support (e.g., disabilities or at-risk children) and professional working with these children (DEC/NAEYC, 2009).

Inclusive education in preschool classrooms can be beneficial for children with and without disabilities. For children with disabilities, inclusive programs provide the environment to encourage social engagement with peers, social acceptance by peers and progress toward individualized education plan goals (Green et al., 2014; Odom et al., 2011). Several studies compared the progress children with disabilities made in inclusive preschool classrooms and separate preschool settings. Findings showed that children with disabilities in inclusive preschool classrooms showed greater progress in cognitive and communication development in comparison to those who were placed in segregated special education settings (Holahan & Costenbader, 2000; Lawrence et al., 2016). Children without disabilities also benefit from inclusive education. Research has shown that children without disabilities develop positive attitudes, empathy, and great compassion for individuals with disabilities (Diamond & Huang, 2005).

Success of inclusive practices in early childhood depends on several factors including environmental quality of the classrooms in which children with disabilities are supported along with their typically developing peers. Quality as it applies to early childhood does not have a universally implemented definition. One definition of quality is “a collection of measurable characteristics in the childcare environment that affect children's social and cognitive development (Siraj-Blatchford & Wong, 1999, p. 10).” Researchers such as Pianta et al. (2016) have provided dimensions of quality in preschool classrooms. Structural quality refers to the features and characteristics of a preschool classroom such as the hours of operation and the teacher's education and experience levels. Classroom environment quality refers to the activities and spaces indoor and outdoor in a preschool classroom including the learning experiences, furniture, and equipment. Quality teacher and student interactions occur on a daily basis and foster an environment for development to occur. Clifford et al. (2010) base their view of quality on the needs of the children; maintaining their health and safety, encouraging positive relationships, and stimulating their learning through experiences.

Research shows that children who attend high-quality early childhood programs achieve higher scores on standardized cognitive and language tests than their peers who attended early childhood programs with lower quality ratings (Burchinal et al., 2000; Cunningham, 2010). High quality early childhood programs provide children with developmentally appropriate curriculum and activities and employ early childhood teachers who are responsive to individual children's needs and understand how to meet these individualized needs. Early childhood educators in high quality programs are able to build stronger attachments with children and provide an environment that supports positive experiences (Vandell & Wolfe, 2000). Studies showed that inclusive preschool classrooms did not result in classrooms

with less quality. For example, Hestenes et al. (2008) found that inclusive classrooms had higher global quality than classrooms that did not enroll children with disabilities. Moreover, Odom et al. (2011) reported positive gains in cognitive, communication and motor skills for children with disabilities because of the individualized support provided in inclusive settings.

Numerous studies have examined the quality of early childhood classrooms. Several reviews of literature summarized the findings of these studies. For example, Aguiar and Aguiar (2020) reviewed 25 studies that focused on quality of early childhood programs that enrolled children from low socioeconomic status and at risk. Classroom quality in these studies were assessed using the Early Childhood Environment Rating Scales (ECERS) or Classroom Assessment Scoring System (CLASS) scales. Findings showed that classrooms with lower quality scores did not meet the developmental needs of children from low socioeconomic status or minorities (e.g., Hispanic/Latino). Slot (2018) reviewed 72 studies and reported consensus in lower classroom enrollment providing better literacy and vocabulary skills and emotional regulation. Although several reviews (e.g., Aguiar & Aguiar, 2020; Slot, 2018) have summarized the literature focused on the quality of preschool classrooms in general, extant literature does not include a study that specifically summarized results of studies focused on investigating quality of inclusive preschool classrooms.

Purpose and Research Questions

The purpose of this systematic review is to identify studies focused on examining quality of inclusive preschool classrooms, determine characteristic features of these studies and summarize information about the overall quality of inclusive preschool classrooms. This review is guided by the following research questions:

1. What are the descriptive characteristics of the studies focused on examining the quality of inclusive preschool classrooms?
2. How was quality measured in the studies focused examining the quality of inclusive preschool classrooms?
3. What is the overall quality reported in the studies focused on examining the quality of inclusive preschool classrooms?

Method

Inclusion and Exclusion Criteria

In order for a study to be included in this literature review, it had to meet four criteria. First, the study had to be conducted in preschool classrooms that enrolled both children with and without disabilities who are between the ages of three to five years. Second, the study had to measure classroom quality as the main purpose of the study. Third, the article had to be published in a peer reviewed journal. Fourth, the full text of the article had to be available in English. Articles were excluded if they included classrooms outside of preschool range and separate data were not provided for preschool classrooms. Articles that included childcare homes or regular preschool classrooms with no children with disabilities enrolled were also excluded from the review. There were no date restrictions included in the criteria for inclusion or exclusion.

Article Search

The process to search and identify articles to include in this review was completed in two stages. The first stage included an online database search using a set of predetermined search terms. The databases used for the search included Academic Search Complete, APA Psycinfo, Child Development & Adolescent Studies (connecting with EBSCO), ERIC, Education Source, PubMed, Scopus, and SAGE. Combinations of the following search terms were used to conduct the data base search: early childhood, preschool, early child development, early childhood development, child care, childcare, special education, disabilities, disabled, special needs, inclusive, inclusion, classroom quality, program quality, quality

ratings, early childhood environment rating scale, inclusive classroom profile, classroom assessment scoring system, assessing classroom sociocultural equity scale, brief early childhood quality inventory, or developmental environment rating scale. In the second stage, a hand search of the reference and citation lists of articles that qualified for the current review was conducted. In this stage, the curricula vitae of researchers who conduct studies focused on early childhood classroom quality were also reviewed to identify additional studies.

Screening and Coding

As shown in Figure 1 (PRISMA flow diagram; Page et al., 2021), the two-stage search process described above yielded 427 articles. Once the duplicate articles were removed, a total of 184 unique articles remained. These 184 articles were screened using the inclusion criteria by reviewing the title and abstract. When it was not clear whether the study met the inclusion criteria based on the title or abstract, the full text of the article was reviewed. Twenty-five articles met the inclusion criteria. The studies were mainly excluded because they did not include classrooms serving both children with and without disabilities or age range of the children served in the classrooms were out of range. A coding form with 27 items was developed to extract data from qualifying articles. These 27 items were grouped under seven categories including purpose, setting (type of program), adult participants (e.g., role, education, experience), child participants (e.g., age, gender, disability), quality measurement tool(s) used, administration of quality measurement tool (i.e., credentials, training), and results.

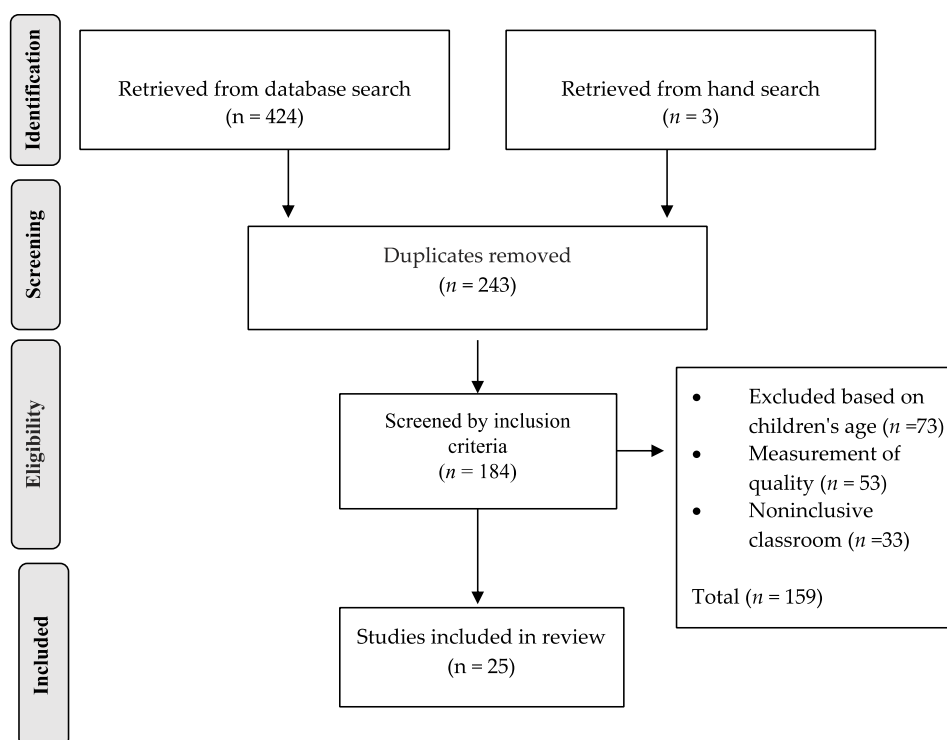


Figure 1. Article search process flowchart

Reliability

Interrater reliability data were collected for screening of the articles based on inclusion and exclusion criteria and coding descriptive information from studies qualified for this review. For this purpose, a second researcher screened and coded 20% of the articles. Interrater reliability percentage scores were calculated using the following formula: $(\text{agreement} / (\text{agreement} + \text{disagreement})) \times 100$ (Kazdin, 2010). The interrater reliability coefficients were 99% for screening and 99% for coding. Cohen’s Kappa coefficient

(Cohen, 1960) was also calculated as an additional measure of interrater reliability for the screening phase, and it was reported to be 0.94, indicating very strong agreement among raters.

Results

The 25 studies that met the inclusion criteria were published from 1993 through 2022. Four studies (16%) were conducted in the 1990s, 3 studies (12%) from 2000 to 2009, 16 studies (64%) from 2010 to 2019 and 2 studies (8%) from 2020 to 2022. Nineteen of the 25 studies (76%) were designed to provide descriptive analysis of inclusive preschool classroom quality while the remaining 6 studies (24%) compared quality of inclusive and noninclusive preschool classrooms. The studies were conducted in various locations around the world including 16 studies in the United States ($n = 64\%$), 2 studies in Greece ($n = 8\%$), 2 studies in Portugal ($n = 8\%$), 2 studies in Spain ($n = 8\%$), 1 study in Botswana ($n = 4\%$), 1 study in Denmark ($n = 4\%$), and 1 study in Türkiye ($n = 4\%$).

Descriptive Characteristics of the Studies

Study Settings

All 25 studies identified the type of preschool settings in which the classroom quality was measured. Twelve studies included more than one type of preschool setting. Across the 25 studies, settings included private child care programs in 14 studies ($n = 603$), public preschool classrooms (located inside a public-school setting) in 11 studies ($n = 744$), preschool classroom in Head Start programs in 10 studies ($n = 161$), religious child care program in 2 studies ($n = 44$), university child care program in 1 study ($n = 6$), and developmental day program in 1 study ($n = 13$). In one study, the setting was identified as preschools ($n = 1357$), 1 study provided the type of preschool setting but did not separate the total number of settings into each type ($n = 36$) and 1 study only provided the type of preschool setting without reporting the total number of settings into each type. Table 1 presents descriptive characteristics of studies included in this review.

Attributes of Study Participants

Teacher Participants. Seventeen of the 25 studies (68%) provided data on the number of preschool teachers included in the studies ($n = 2633$). Six studies (24%) provided the mean ages of the teachers ($M = 41.83$). Ten studies (40%) conducted in the United States reported information about race of adult participants while six studies from the United States and studies that were implemented in other countries did not report such data. Six studies (24%) reported the race of the teachers in numeric form (Black = 251, White = 240, Native American = 2, Hispanic = 7, Latino = 1, Asian Pacific Islander = 3, Mixed Races = 4, Other = 1) while four studies (16%) reported the information in percentages (Black = 1.5%-50%, White = 49%-95%, Native American = 1%-2%, Hispanic = 1%-3%, Latino = 6%, Asian Pacific Islander = 10%, Mixed Races = 1.5%). Thirteen studies (52%) reported the gender of the classroom teachers (range = 93.8% to 100% female).

Six studies (24%) reported teachers' level of education in numeric form (high school = 59, child care credential = 96, associate's = 16, some college = 10, bachelor's = 182, master's = 79, not specified = 20) while nine studies (36%) reported this information in percentages (high school = 9.4%-70%, child care credential = 6%-30.5%, associate's = 11%-27%, bachelor's = 17.2%-100%, master's = 12%-75%, not specified=7%). One study (4%) provided a mean number of years in education for teachers ($M = 15.18$ years). Twelve studies (48%) reported a mean for teachers' years of experience (range = 5.64 - 22.74) while three studies (12%) provided a range for years of teacher's experience (e.g., 2 - 6 years = 18 teachers, 7-11 years = 23 teachers, 5-14 years = 10 teachers, 12-21 years = 6 teachers). One study (4%) provided whether teachers received any training or courses focused on inclusion (yes = 19 teachers, no = 28 teachers).

Table 1. Study Participants and Setting

Reference/ Country	Setting	Educators					Students				
		Total	Age (in years)	Race - Gender	Education/ PD	Years of Experience	Total	Age (in months)	Race - Gender	Disabilities	
Aguiar et al. (2019) Portugal	39 HS 5 PRI	44	48	N/A – 1 M	NA	22.74	222 (42 w/d)	63.75	N/A – INC: 90 M, 132 F, NI: 29 M	12 DD, 9 ASD, 7 OTR, 4 UD, 3 RD, 3 SLD, 2 DS, 2 MIN	
Aguiar et al. (2010) Portugal	64 PUB	N/A	46.81	N/A – N/A	N/A	INC:12.33, NI:22.67	1121	62.74	N/A – 594 M	403 DD, 280 ASD, 107 CP, 102 OTR, 71 MUL, 53 ADHD, 35 DS, 35 HEA, 35 SLD	
Bakkaloglu et al. (2019) Türkiye	13 PRI	47	25+	N/A – N/A	Training in Inclusion: 19 Yes, 28 No	2+	N/A	39+	N/A – N/A	12 ID, 11 SLP, 10 ASD, 9 PHY, 5 HEA	
Buyse et al. (1999) United States	115 PRI 38 REL 20 HS 7 PUB	180	N/A	N/A – N/A	INC: 9 HS, 33 CC, 11 BAC, 9 NA NI: 26 HS, 62 CC, 19 BAC, 11 NA	INC 7.65, NI 7.06	N/A	N//A	N/A – N/A	N/A	
Cadima et al. (2018) Spain	130 PUB 48 PRI	N/A	N/A	N/A – 99% F	12% MAS	21.18	N/A	N/A	N/A – N/A	N/A	
Campbell et al. (2005) United States	228 PUB	228	39.86	177 B, 50 W, 1 L – 224 F	56% HS, 11% ASC, 22% BAC, 6% CC	10.57	N/A	N/A	N/A – N/A	N/A	
Chhabra et al. (2018) Botswana	34 PRI	N/A	N/A	N/A – 93.8% F	36% EDU, 31% CC, 17% BAC, 9% HS, 7% OTH	Range: 1 - 10	N/A	N/A	N/A – N/A	10 SLD, 10 TS, 3 CP, 3 DS, 2 ASD, 2 HEA	
Classen and Westbrook. (2022) United States	26 PRI, 2 HS, 6 REL, 6 UNV	83	N/A	45 B, 36 W, 2 MD – 83 F	7 MAS, 27 BAC, 15 ASC, 10 Some College 24 HS	N/A	N/A	N/A	N/A – N/A	N/A	
Clawson and Luz. (2008) United States	7 HS 4 PRI	11	N/A	8 W, 1 H, 1 B, 1 O – N/A	7 MAS, 4 BAC	N/A	60 (30 w/d)	56.52	49 W – 24 M	40 SLD, 28 BEH, 16 PHY, 6 COG, 4 VIS	
Coelho et al., (2019)	39 PRI	39	49.49	N/A – 100% F	100% BAC	10.86	184 (104 w/d)	47.06	N/A – N/A	40 RD, 20 DD, 17 ASD, 4 CP, 3 DS, 2 SLD, 2 ADHD, 1 PHY, 1 HEA	

A Systematic review of studies investigating quality...

Spain											
File and Kontos (1993) United States	PRI, UNV	36	N/A	N/A – 35 F, 1 M	30% BAC	N/A	28	56	N/A – 24 M	14 COG/SLD	
Fyssa and Vlachou (2015) Greece	52 PUB	96	N/A	N/A – N/A	95- BAC	N/A	93 w/d	72	N/A – 65 M	38 ASD, 26 DD, 15 ID, 13 NEU, 1 HEA	
Grisham-Brown et al.(2010) United States	64 PRI 2 HS	59	N/A	INC: 16% B, 84% W, NI: 9% B, 2% NAT, 88 % W, 2% MR, – N/A	INC: 67% HS, 27% BAC NI: 50% HS, 54% BAC	NI 5.64, INC 6.38	N/A	N/A	N/A – N/A	42% SLD+MD, 24% PHY, 18% SEL, 3% COG	
Hestenes et al. (2008) United States	1357 Pre-school	1357	N/A	Study 1: N/A, Study 2: 49% W, 50% B, 1% H – 72 F	Study 1: N/A, Study 2: 70% HS 30% BAC	N/A	N/A	N/A	N/A – N/A	Study 1: 658 SLD, 240 DD, 78 PHY, 52 OTH, 46 COG, 40 ASD, 27 BEH, 26 ADHD, 16 DS, Study 2: 11 ADHD, 11 SPD, 10 DD, 10 PHY 7 ASD, 3 BEH, 3 DS, 2 OTH, 1 COG, Study 2: N/A	
Jeon et al. (2010) United States	54 HS 52 PRI	106	N/A	64 W, 22 B, 6 H, 2 API – 106 F	15.18	10.17	138	62.16	78 W, 34 B, 16 H, 10 O – 69 M	N/A	
Keesbury (2015) United States	5 PUB	N/A	N/A	N/A – N/A	N/A	N/A	34 (20 w/d)	N/A	N/A – 19 M	N/A	
Kontos et al. (1998) United States	PUB, HS, PRI, UNV	N/A	N/A	N/A – N/A	N/A	N/A	48	53.16	N/A – N/A	14 COG, 14 SLD, 8 PHY, 4 ASD	
La Paro et al. (1998) United States	16 PRI 7 HS 33 PUB	N/A	NI: 36.38 INC: 39.66	INC: 79% W, 17% B, 3% H, NI: 69% W, 28% B, 3% H – N/A	INC: 62% BAC NI: 100% BAC	INC: 11.39 NI:10.30	N/A	N/A	N/A – N/A	N/A	

Muccio et al. (2014) United States	9 HS	71	N/A	N/A – N/A	N/A	N/A	N/A	N/A	N/A	32%-56% B, 4%-42% W, 2%-4% AIP 3%-13% NAT, 29-36% MD, 5%-10%MR 7% O - N/A	N/A
Næsby (2020) Denmark	88 PRI	N/A	N/A	N/A – N/A	N/A	N/A	N/A	N/A	N/A	N/A – N/A	N/A
Pelatti et al. (2016) United States	85 PUB 79 PRI	N/A	N/A	INC: 95% W, 2% NAT, 1.5% B, 1.5% AIP, NI: 62% W 21% B, 10% AIP, 6% L 1% NAT - INC: 99% F, NI: 95% F,	INC: 75% MAS, 25% BAC, NI: 37% BAC, 27% ASC, 18% MAS, 18% HS	INC: 11.6, NI: 8.44	1192 (96 w/d)	52		723 W, 272 B, 46 MR, 73 L, 78 O – 681 M	N/A
Soukakou et al. (2014) United States	13 HS 20 PRI 5 PUB 13 DEV	150	N/A	N/A – N/A	HS: 54% BAC, PUB: 100% BAC, DD: 92% BAC, CC: 25% BAC	HS 13.62 PRI 12.85 PUB 5.60 DD 8.10	150 w/d	4.43		N/A – N/A	SLD; 38%, DD; 37%, ASD: 12%, OTR: 8%, HEA: 2%, MD: 1%, PHY: 1%, VIS: 1%,
Stanton-Chapman et al. (2016) United States	10 HS	10	28.8	4 W, 4 B, 2 MR – 9 F, 1 M	4 MAS, 4 BAC, 1 CC, 1 ASC	Range: 5-14.	179	N/A		88 B, 56 W, 35 H – 91 M	14 SLD, 11 DD, 11 ESL, 2 ASD
Vlachou and Fyssa (2016) Greece	52 PUB	96	N/A	N/A – 94 F	N/A	N/A	93 w/d	72.68		N/A – 65 M	38 ASD, 26 DD, 15 ID, 13 NEU, 1 HEA
Yeomans-Maldonado et al. (2019) US	83 PUB	83	N/A	78 W, 2 B, 2 NAT, 1 AIP – 82 F	22 BAC, 61 MAS	12.66	670	52		438 W, 99 MD, 62 B, 30 L, 41 O, 436 M	N/A

Notes. *Types of Programs*: INC: Inclusive; NI: Noninclusive; HS: Head Start; PUB: Public School PreK; PRI: Private or Community Child Care; REL: Religious; UNV: University; DEV: Developmental Day Program. *Education Levels/Professional Development*: HS: High School; CC: Child Care Credential; ASC: Associates; BAC: Bachelors; MAS: Masters. *Total Children*: w/d: With Disabilities; *Race*: B: Black/African American; W: White; L: Latino; H: Hispanic; NAT: Native American; AIP: Asian Pacific Islander; MR: Multi-Racial; O: Other; MD: Missing Data. *Gender*: F: Female; M: Male. *Type of Disability*: ASD: Autism Spectrum Disorder; BEH: Behavior; COG: Cognitive Delays; CP: Cerebral Palsy; DD: Developmental Delays; ESL: English As Second Language; FMD: Fine Motor Delay; GMD: Gross Motor Delay; HEA: Hearing Impairment; ID: Intellectual Delays; MD: Multiple Disabilities; MIN: Missing Information; NEU: Neurological Delay; OTR: Other; PHY: Physical Delay; RD: Rare Disorder; SEN: Sensory Delay; SLD: Speech/Language Delays; TS: Turner Syndrome; VIS: Vision Impairment.

Table 2. Measurement of classroom quality

Reference/Country	Measure	Observer/Qualification/Training	Reliability (Frequency/IOA Score)
Aguiar et al. (2019)/Portugal	CLASS	4 certified independent observers; 2-day training; Trained to a reliability standard of 80%	
Aguiar et al. (2010)/Portugal	QuIEM	N/A	11% of classrooms; Mean IOA range across subscales = 91%-96%
Bakkaloglu et al. (2019)/Türkiye	ICP	N/A	IOA = 92%.
Buysse et al. (1999)/United States	ECERS	N/A	Mean IOA = 86% (range = 75% - 92%)
Cadima et al. (2018)/Spain	CLASS	2-day training; certification test; Reached reliability criterion (80%)	N/A
Campbell et al. (2005)/ United States	ECERS	14 observers; Trained to reliability using videotapes; Live observation with a trained rater for a minimum of three observations or until IOA of greater than 85% on each subscale was achieved.	N/A
Chhabra et al. (2018)/Botswana	ICP	N/A	N/A
Classen and Westbrook (2022)/United States	ICP	Formal level 2 reliability training with a certified ICP training team; Trained to a reliability standard of 80%	N/A
Clawson and Luz. (2008)/United States	ECERS-R	Two research assistants; 24 hours of training; Trained to a reliability standard of 85%	17% of observations; Mean IOA = 90% (range = 75% - 95%), mean kappa = .96 (range .82-.96)
Coelho et al. (2019)/Spain	CLASS	2 independent observers; Trained before data collection; Trained to a reliability standard of 80%	25% of the observations; Exact IOA = 67%, IOA within one point = 99.50%
File and Kontos (1993)/United States	ECERS	IOA prior to data collection = 94%	20% of the classrooms, Mean IOA = 94%
Fyssa and Vlachou (2015)/Greece	ICP	Formal ICP reliability training; 2-day training program	Mean IOA = 92.5% (range = 87.3% - 98.1%), mean kappa = .87
Grisham-Brown et al. (2010)/United States	ECERS-R	Observers were trained in ECERS-R	
Hestenes et al. (2008)/United States	ECERS-R	Trained evaluators; Extensive training; Trained to a reliability standard of 85%	Study 1: Met a minimum of 85% reliability. Study 2: Mean interrater reliability = 92% (range = 86%-98%)
Jeon et al. (2010)/United States	ECERS-R	Trained research assistants: IOA within one-point was at least 85%	N/A
Keesbury (2015)/United States	ECERS-R	Smart Start Evaluators	N/A
Kontos et al. (1998)/United States	ECERS	2 trained graduate students	Mean IOA = 93%, mean kappa = .86

Lisa BUTTERWORTH et al.

La Paro et al. (1998)/United States	ECERS	3 observers were trained by watching videotapes and visiting classrooms; Trained to a reliability standard	N/A
Muccio et al. (2014)/United States	ICP	N/A	N/A
Næsby (2020)/Denmark	ECERS-3	2 certified observers were trained	N/A
Pelatti et al. (2016)/ United States	CLASS	2-day CLASS training; IOA of 90 % agreement with 6 master-coded videos.	Mean IOA = 92 %
Soukakou et al. (2014)/United States	ECERS-R ICP	4 Trained evaluators; ICP 3 hr. training and 4 hr. reliability observations; Met reliability standard of 85% IOA	Mean IOA = 98% (range = 91% - 100%)
Stanton-Chapman et al. (2016)/United States	ECERS-R CLASS	2 research associates and 2 master's students; Trained to a reliability standard of 90%	N/A
Vlachou and Fyssa (2016)/Greece	ICP	Research completed a reliability training	9.6% of the classrooms; Mean IOA= 92.5% (range = 87.3% - 98.1%)
Yeomans-Maldonado et al. (2019)/United States	CLASS	2-day training; Trained to a reliability standard of 90%	N/A

Note. INC: Inclusive; NI: Noninclusive; M: Mean Scores; PC: Personal care; FR: Furnishings; LA: Language reasoning; FG: Fine and gross motor; CA: Creative activities; SD: Social development; AN: Adult needs; SF: Space and furnishings; LE: Learning activities; IN: Interactions, PS: Program structure; PA: Parents and staff; ECERS: Early Childhood Environment Rating Scale

Child Participants. Fourteen studies (56%) included data for the number of preschool children enrolled in preschool classrooms. Eight of the fourteen studies (57%) included data that differentiated between children with and without disabilities. A total of 4258 preschool children participated in these studies and 628 had a disability. Child gender was reported in 11 of the 25 studies (44%). A total of 2187 males and 2025 females participated in these studies. Thirteen studies (52%) reported the age of the children. One of the thirteen studies provided the range of ages, and twelve studies provided the mean. Range of mean age reported across studies was 52 months – 72.68 months. Fourteen studies (56%) included data on the types of disabilities children had. Eight studies (28%) included children that were characterized as developmentally delayed or at risk for developmental delay. Other disability categories reported were speech and language delay ($n = 12$ studies), autism spectrum disorder ($n = 11$ studies), physical disability ($n = 7$ studies), cognitive or neurological delays ($n = 7$ studies), hearing impairment ($n = 7$ studies), Down syndrome ($n = 6$ studies), cerebral palsy ($n = 3$ studies), attention deficit hyperactivity disorder ($n = 3$ studies), multiple disabilities ($n = 2$ studies), other disabilities ($n = 10$; for example, vision impairment, social emotional delay). Only 6 of the 25 studies (24%) reported data regarding the race of preschool children.

Measurement of Quality

Measures Used

Across the 25 studies, 23 (92%) used one classroom quality measure while 2 (8%) used two measures of classroom quality. Across the studies, four classroom quality measures were used: Early Childhood Environment Ratings Scale (ECERS; $n = 13$), Inclusive Classroom Profile (ICP; $n = 7$), Classroom Assessment Scoring System (CLASS, $n = 6$), and Quality of Inclusive Experiences Measure (QuIEM; $n = 1$). Table 2 presents information about measurement of classroom quality across studies.

The original version of ECERS (Harms & Clifford, 1980) was introduced in 1980. As one of the initial measures of early childhood classroom quality, ECERS became a commonly used measure of quality during the 1980's. In 2005, a revised edition, ECERS-R (Harms et al., 2005) was introduced. The revised edition expanded the notes for clarification and combined or deleted redundant indicators. Examples of how to meet quality standards for diversity and inclusion were added to the notes for clarification. In 2015, a revised edition, ECERS-3 (Harms et al., 2015) was introduced. ECERS-3 revised the beginning age from 2 ½ to 3 years old. The teacher interview portion has been removed and an increased emphasis on language and math has been incorporated. Indicator 37, provisions for children with disabilities, has also been removed in ECERS-3. ECERS and ECERS-R have 7 subscales with a total of 37 and 43 items, respectively while original ECERS-3 have 35 items under 6 subscales. Each item is individually scored on a scale between 1 (inadequate) and 7 (excellent), with higher scores indicating better quality.

The QuIEM was developed by Wolery et al. (2000) to measure the quality of inclusive experiences in conjunction with other quality assessment tools. The QuIEM was specifically designed to address the quality of experiences for children with disabilities and comprised of 7 areas. The measure utilizes a 5-point scoring system, with higher scores denoting better quality. It is intended to assist in identifying areas where improvement is needed and to help guide program planning and implementation to enhance the quality of inclusive experiences for children with disabilities.

The CLASS is developed by Pianta et al. (2008) as an observation instrument that evaluate the quality of teacher-child interactions in center-based preschool classrooms. CLASS contains 10 dimensions under three domains: Emotional Support, Classroom Organization, and Instructional Support. Each dimension is rated based on a 7-point scale with a score of 1-2 meaning low levels of teacher-child interactions and 6-7 meaning the teacher and child interactions occurred consistently.

The ICP is a structured observation rating scale designed to evaluate the quality of provisions and daily practices that support the developmental needs of children with disabilities in early childhood settings serving children ages 2–5. The scale was developed to complement other scales that measure classroom quality (Soukakou, 2012). The ICP contains 12 key practices/items. Each item on the ICP is rated using a 7-point rating scale ranging from 1 (practices considered highly inadequate for promoting

children's active participation in the group and meeting their individual needs) to 7 (practices that promote to the highest degree children's active participation in the group through individualized supports).

Observer Characteristics

Seven studies (28%) of the 25 reported who and how many observers were included. These included 5 research assistants in 3 studies, 4 trained evaluators in 1 study, 4 graduate students in 2 studies, and 2 authors in 2 studies. In 5 studies (20%), researchers reported employing 25 independent observers (e.g., doctoral students, researchers, trained observers) for data collection. However, they did not specify the role or credentials of these observers. Three of the 25 studies (12%) included who the observers were but did not include how many. Ten studies (40%) did not specify who or how many observers were employed.

Observer Training

Nineteen of the 25 studies (76%) reported that the observers were trained prior to collecting data for the study. Of these 19 studies, 18 studies (72%) reported that the observers met a required reliability criteria before collecting data. Twelve of the studies (48%) reported the reliability criterion required for observers to become a reliable observer during the training (range = 80% - 90%).

Reliability of Classroom Quality Ratings

Interobserver agreement (IOA) reliability data was reported in 14 of 25 studies (56%). The mean IOA ranged between 86% - 99.5% across the 14 studies. Five of the 25 studies (20%) reported the percentage of classrooms in which the IOA reliability data were collected (range = 9.6% - 25%).

Classroom Quality Reported

As noted earlier, the quality of inclusive classrooms was evaluated using four measures including ECERS measures, ICP, CLASS, and QuIEM. Results obtained from each measure were presented separately in the following section.

ECERS, ECER-R and ECERS-3

As seen in Table 3, 13 studies used a version of the ECERS as the measurement tool. This includes five studies using ECERS (Harms & Clifford, 1980), seven studies using ECERS-R (Harms et al., 2005) and one study using ECERS-3 (Harms et al., 2015). Of the five studies using ECERS (Harms & Clifford, 1980), three studies reported an overall mean score for all ECERS subscales combined (range = 4.51-4.8) and two studies reported individual mean subscale scores (range of mean scores across studies; Personal Care = 3.25 - 4.72, Furnishings = 3.28 - 4.41, Language Reasoning = 2.9 - 4.38, Fine/Gross Motor = 3.44 - 4.64, Creative Activities = 2.98 - 4.45, Social Development = 2.83 - 3.99, and Adult Needs = 3.5 - 4.49)

Of the seven studies using ECERS-R (Harms et al., 2005), four reported overall mean scores (range = 4.52 - 5.45) as well as individual subscale scores (range across studies; Space and Furnishings = 4.68 - 5.02, Personal Care Routines = 2.88 - 4.58, Language and Literacy = 4.66 - 5.81, Learning Activities = 4.18 - 5.29, Interactions = 5.29 - 6.18, Program Structure = 4.94 - 5.99, Parents and Staff = 5.44 - 6.13). One study only reported the overall mean score ($M = 5.16$). One study reported individual subscale means only and provided them for four subscales (Language and Literacy = 5.50, Learning Activities = 4.19, Interactions = 5.38, Program Structure = 5.55). The last study using ECERS-R (Harms et al., 2005) reported overall mean scores obtained from ECERS-R based on the type of child care programs ($M_{\text{Head Start}} = 4.95$, $M_{\text{Child Care}} = 4.58$, $M_{\text{Public PreK}} = 5.14$, $M_{\text{Developmental Day}} = 5.31$, and $M_{\text{Overall}} = 4.92$). One study using ECERS-3 (Harms et al., 2015) reported findings as individual mean subscale scores. This included Space and Furnishings = 2.97, Personal Care Routines = 2.56, Language and Literacy = 3.3, Learning Activities = 2.14, Interactions = 4.46, and Program Structure = 4.04).

Table 3. Early childhood environment rating scale mean scale and subscale/indicator scores (SD)

Reference/Country/n	ECERS								
	M	PC	FR	LA	FG	CA	SD	AN	
Buyse et al. (1999)/United States/ n = 180	N/A	INC = 4.72 (.87), NI = 4.38 (.80)	INC = 4.41 (.78), NI = 4.13 (.79)	INC = 4.38 (1.02), NI = 3.99 (.93)	INC = 4.64 (.62), NI = 4.45 (.49)	INC = 4.45 (.82), NI = 4.24 (.62)	INC = 3.99 (1.01), NI = 3.61 (.78)	INC = 4.49 (1.06), NI = 4.08 (.94)	
Campbell et al. (2005)/United States/ n = 228	N/A	3.25 (range = 1.8-7.0)	3.28 (range 1.0-7.0)	2.9 (range 1.0-7.0)	3.44 (range 1.4-5.0)	2.98 (range 1.0-6.14)	2.83 (range 1.0-6.67)	3.50 (range 1.5-6.5)	
File and Kontos (1993)/United States/ n = 36	4.51 (.70)								
Kontos et al. (1998)/United States; n = N/A	4.8								
La Paro et al. (1998) United States; n = N/A	INC = 4.77 (.83) NI = 4.68 (.60)								
ECERS-R									
Reference/Country/n	M	SF	PC	LA	LE	IN	PS	PA	
Clawson and Luz. (2008) United States; n = 11	N/A	N/A	N/A	5.50 (1.30)	4.19 (0.88)	5.38 (1.51)	5.55 (1.34)	N/A	
Grisham-Brown et al. (2010), United States; n = 59	INC = 4.74 (.9), NI = 3.93 (.95)	INC = 4.79 (1.03), NI = 4.05 (.99)	INC = 2.88 (1.29), NI = 2.62 (1.09)	INC = 5.00 (1.13), NI = 4.40 (1.3)	INC = 4.70 (1.67), NI = 3.33 (1.19)	INC = 5.64 (1.15), NI = 5 (1.4)	INC = 5.68 (1.3), NI = 3.73 (1.17)	INC = 5.44 (.96), NI = 4.96 (1.2)	
Hestenes et al. (2008) United States; n = 1357	Study 1; INC: 5.19, NI: 4.90, Study 2; INC: 4.68, NI: 4.19	Study 1; INC: 5.02, NI: 4.91, Study 2; INC: 4.88, NI: 4.38	Study 1; INC: 3.97, NI: 3.8, Study 2; INC: 4.58, NI: 3.71	Study 1; INC: 5.62, NI: 5.30, Study 2; INC: 4.66, NI: 4.11	Study 1; INC: 5.29, NI: 5.02, Study 2; INC: 4.18, NI: 4.06	Study 1; INC: 5.86, NI: 5.45, Study 2; INC: 5.29, NI: 4.38	Study 1; INC: 5.88, NI: 5.56, Study 2; INC: 4.94, NI: 4.79	Study 1; N/A, Study 2; N/A	
Jeon et al. (2010) United States; n = 106	INC = 4.52 (.78), NI = 3.71 (1.22)	N/A	N/A	INC = 5.81 (.82), NI = 4.49 (1.51)	INC = 4.41 (.83), NI = 3.83 (1.10)	INC = 5.92 (1.22), NI = 5.18 (1.49)	INC = 5.66 (1.66), NI = 4.83 (1.88)	N/A	
Keesbury (2015) United States; n = N/A	5.45	4.68	4.45	5.69	5.16	6.18	5.99	6.13	
ECERS-3									
Næsby (2020) Denmark; n = N/A	N/A	2.97	2.56	3.3	2.14	4.46	4.04	N/A	
Soukakou et al. (2014) United States; n = 148	4.92								
Stanton-Chapman et al. (2016) United States; n = 10	5.16								

Note. INC: Inclusive; NI: Noninclusive; M: Mean Scores; PC: Personal care; FR: Furnishings; LA: Language reasoning; FG: Fine and gross motor; CA: Creative activities; SD: Social development; AN: Adult needs; SF: Space and furnishings; LE: Learning activities; IN: Interactions; PS: Program structure; PA: Parents and staff; ECERS: Early Childhood Environment Rating Scale

Table 4. Inclusive classroom profile mean scale and item scores (SD)

Reference/ Country/n	M	Key practice/Item											
		AS	AI	AG	CR	ME	RE	SC	AD	TR	FE	FO	MO
Bakkaloglu et al. (2019) Türkiye; n = 47	N/A	3.38 (.67)	1.80 (.61)	2.02 (.76)	.08 (.35)	2.19 (.77)	2.46 (.85)	1.72 (.53)	2.02 (1.52)	2.02 (.14)	1.87 (.49)	1.00 (.00)	1.21 (.41)
Chhabra et al. (2018) Botswana; n = N/A	N/A	4.38 (1.0)	4.48 (1.34)	3.71 (1.0)	3.81 (1.3)	4.19 (.91)	4.46 (.67)	4.62 (.80)	5.04 (.93)	4.46 (.80)	4.33 (1.09)	2.91 (.70)	4.29 (.67)
Classen and Westbrook (2022) United States; n = 83	N/A	4.93 (1.67)	3.14 (1.67)	3.36 (1.32)	N/A	3.19 (1.56)	3.39 (1.48)	2.87 (.947)	3.72 (1.58)	2.23 (1.33)	3.08 (1.24)	1.60 (1.33)	2.17 (1.15)
Fyssa and Vlachou (2015) Greece; n = 96	N/A	3.69 (1.50)	2.73 (1.50)	2.96 (1.73)	1.33 (0.51)	2.58 (1.58)	3.37 (1.14)	2.88 (1.42)	2.73 (1.54)	2.73 (1.54)	3.17 (1.25)	N/A	1.17 (.55)
Muccio et al. (2014) United States; n = 71	N/A	4.89 (.6)	4 (.87)	4.38 (.92)	4.44 (.73)	4.67 (.87)	4.78 (1.09)	4.44 (1.13)	5 (1.5)	5.22 (.97)	4.78 (.97)	N/A	N/A
Vlachou and Fyssa (2016) Greece; n = 96	NA	3.69 (1.5)	2.73 (1.4)	2.96 (1.73)	N/A	.58 (1.58)	3.37 (1.14)	2.88 (1.42)	2.73 (1.54)	2.92 (1.45)	3.17 (1.25)	N/A	N/A
Soukakou et al. (2014) United States; n = 148	Mean: Overall (4.39), Head Start (4.64), Child Care (3.67), Public PreK (4.76), Developmental Day (5.12)												

Note. M: Mean Scores; AS: Adapt space; AI: Adult involved-Interaction; AG: Adult guide play; CR: Conflict resolution; ME: Membership; RE: Relation between adults/ children; SC: Support communication; AD: Adaptation of group activities; TR: Transition between activities; FE: Feedback; FP: Family-professional partnership; MO: Monitor child learning.

Table 5. Classroom assessment scoring system mean scale and domain/dimension scores (SD)

Reference/ Country/n	M	Emotional Support				Classroom Organization				Instructional Support		
		POS	NEG	TSE	RFS	BEH	PRO	ILF	COD	QOF	LMO	
Aguiar et al. (2019) Portugal; n = 44	N/A	medium quality				medium quality				low quality		
Cadima et al. (2018) Spain; n = N/A	N/A	4.84 (1.01)	1.52 (0.60)	4.36 (1.12)	4.24 (1.12)	4.69 (1.08)	5.19 (0.84)	4.13 (0.99)	2.02 (0.87)	2.28 (0.91)	2.51 (1.01)	
Coelho et al. (2019) Spain; n = 39	N/A	4.83 (.65)				3.97 (.65)				2.48 (.70)		
Pelatti et al. (2016) United States; n = N/A	N/A	Inclusive: 5.20 (.64) Non-Inclusive: 4.91 (.88)				Inclusive: 4.62 (.64) Non-Inclusive: 4.38 (.86)				Inclusive: 2.19 (.54) Non-Inclusive: 3.86 (.97)		
Stanton-Chapman et al. (2016) Unites States; n = 10	5.16	N/A				4.67				2.95		
Yeomans-Maldonado et al. (2019) United States; n = 83	5.23 (1)	5.23 (1)	1.35 (.46)	4.68 (.99)	4.11 (.99)	N/A	N/A	N/A	2.18 (.76)	2.23 (.81)	2.11 (.80)	

Note. M: Mean; POS: Positive Climate; NEG: Negative Climate; TSE: Teacher Sensitivity; RFS: Regard for Student Perspective; BEH: Behavior Management; PRO: Productivity; ILF: Instructional Learning Formats; COD: Concept Development; QOF: Quality of Feedback; LMO: Language Modeling.

ICP

As illustrated in Table 4, seven studies used the **ICP** as the measurement tool to assess quality of inclusive preschool classrooms. Of the seven studies, six reported individual mean subscale scores. The range of mean subscale scores reported across studies were Adaptation of Space and Materials = 3.38 – 4.93, Adult Involvement in Peer Interactions = 1.8 – 4.48, Adult's Guidance of Children's Play = 2.02 – 4.38, Conflict Resolution = .08 – 4.44, Membership = 2.19 – 4.67, Relationships Between Adults and Children = 2.46 – 4.78, Support for Communication = 1.72 – 4.62, Adaptation of Group Activities = 2.02 – 5.04, Transitions Between Activities = 2.02 – 5.22, Feedback = 1.87 – 4.78, Family Professional Partnership = 1.0 – 2.91, Monitoring Children's Learning = 1.17 – 4.29. Of the seven studies using the ICP, one study reported overall mean scores obtained from ICP based on the type of child care programs ($M_{\text{Head Start}} = 4.64$, $M_{\text{Child Care}} = 3.67$, $M_{\text{Public PreK}} = 4.76$, $M_{\text{Developmental Day}} = 5.12$, and $M_{\text{Overall}} = 4.39$).

CLASS

As presented in Table 5, six studies used **CLASS** as the measurement tool to evaluate preschool classroom quality. Across the six studies, three studies reported mean domain scores (range of mean scores across studies; Emotional Support = 4.83-4.91, Classroom Organization = 3.97- 4.67, and Instructional Support = 2.48-3.86). Two studies reported mean dimension scores for 7 of the 10 dimensions (range of mean scores across studies; Positive Climate = 4.8 – 5.23, Negative Climate = 1.35 - 1.5, Teacher Sensitivity = 4.36 – 4.68, Regard for Student Perspectives = 4.11 - 4.24, Concept Development = 2.02 – 2.18, Quality of Feedback = 2.23 – 2.28, and Language Modeling = 2.11 – 2.51). One of these studies provided mean dimension scores for the three remaining dimensions while the other did not (Behavior Management = 4.69, Productivity = 5.19, Instructional Learning Formats = 4.13). One study did not report accrual mean scores obtained from **CLASS** but categorized scores as range (Classroom Organization – medium range, Emotional Support – medium range, Instructional Support – low range).

QuIEM

One study used the **QuIEM** and reported mean scores for each area. The mean area scores were 65.94 for Program Goals and Purpose, 86.99 for Staff Supports and Perceptions, 16.86 for Accessibility/Physical Movement, 28.66 for Individualization, 90.72 for Children's Participation and Engagement, 39.06 for Adult to Child Contacts, and 53.16 for Child to Child Contacts.

Discussion

The purpose of this systematic review was to identify studies focused on examining quality of inclusive preschool classrooms, determine characteristic features of these studies and summarize information about the overall quality of inclusive preschool classrooms. Twenty-five articles met the inclusion criteria and were analyzed using a 27-item coding system developed by the researcher. In the following sections, I discuss the findings of the present study with respect to each research question, describe limitations of the present review, and offer implications for future research.

Descriptive Characteristics

Across the 25 studies, there was not a consistent method of reporting classroom demographics including characteristics of early care providers or preschool teachers. The majority of the studies reported the type of preschool classroom. However, some of the studies did not specify the type of preschool classroom and only referred to the data as occurring in preschool classrooms. Limited data in relation to early care providers' or preschool teachers' demographics were reported across the studies including age, years of experience or training on inclusion. Racial demographics were only provided in studies conducted in the United States. The majority of the teachers in the studies identified as white or Black females which is consistent with the demographics of early child care providers in the United States (Whitebook et al., 2018). The limited data on the classroom characteristics and demographics of early care providers or preschool teachers, is a significant limitation in understanding the characteristics of teachers who are

successful in promoting high-quality inclusive classrooms. Moreover, it limits the generalizability of the findings to larger populations. The EASNIE (2014) highlighted the importance of addressing the diverse needs of teachers and ensuring equitable representation in early childhood education settings to improve inclusive practices. Future studies should provide detailed information on classroom characteristics, such as the type of the classroom (inclusive versus non-inclusive), the size of the classroom and consider including more detailed information on the demographics and qualifications of teachers, which could provide valuable insights for developing effective professional development programs for teachers in relation to improving classroom quality.

Measurement of Classroom Quality

Earlier studies investigating preschool classroom quality used CLASS, ECERS, and ECERS-R, measures designed to evaluate general preschool classroom quality. The provision of children with disabilities in these measures is limited as they were not specifically designed to assess the quality of inclusive classrooms. While these measures do include some items related to the provision of services for children with disabilities, they may not capture the full range of practices needed to create high quality inclusive environments. For example, the ECERS-R includes items related to the provision of specialized equipment and materials for children with disabilities but does not assess the extent to which children with disabilities are actively engaged in classroom activities and routines or involved in peer interactions. As a result of these limitations, researchers have developed measures that are specifically designed to evaluate the quality of inclusive preschool classrooms. One such measure is the Inclusive Classroom Profile which has been used in seven more recent studies included in this review to assess the quality of inclusive early childhood settings.

Fifteen studies reported some information about individuals who conducted classroom quality measures. This includes roles/credentials and number of observers. Nineteen studies specified that observers were trained prior to data collection for the study but training procedures or trainers' credentials were rarely described. These inconsistent reporting practices make it difficult to evaluate whether observers were trained to use these observational measures using valid training procedures by trainers who were qualified to provide training. For example, if ECERS is being used as a high stakes assessment, the training must come from an authorized trainer, which is currently the authors. Environment Rating Scale Institute (ERSI) does acknowledge that agencies may be using unauthorized trainers which can result in unrealistic scores for programs (Environment Rating Scale Institute [ERSI], n. d.), which may also be true for research studies.

Interobserver agreement (IOA) reliability data was only reported in 14 studies. The studies had a broad range of acceptable IOA percentages ranging between 86% to 99.5%. Eleven studies did not report that they met any IOA standards in their studies. The lack of reporting of IOA in observational research is problematic as it raises questions about the reliability and validity of the data collected. IOA is a measure of the extent to which multiple observers agree on the occurrence or non-occurrence of a behavior or event and is crucial for ensuring that the data collected is accurate and consistent (Ledford et al., 2018). The fact that only 14 studies reported IOA in this review suggests that many studies may not have conducted adequate reliability checks, which can undermine the credibility and usefulness of their findings. Future studies should collect and report IOA data more systematically to improve the rigor and credibility of observations. Moreover, future studies should report how observers were trained, whether they were certified, or reached to a reliability standard before collecting data for the study.

Classroom Quality

Six studies compared the quality of preschool classrooms with and without children with disabilities in the United States. Of six studies, three used ECERS-R, two used ECERS, and one used CLASS. The findings of the studies using the ECERS or ECERS-R tools indicate that inclusive classrooms had higher overall and mean subscale scores than classrooms that only served typically developing children. However, the mean subscale scores for both classroom types generally fell within the good quality range (between 4 and 6). This suggests that both inclusive classrooms and classrooms that only served typically developing

children generally provided good quality care as measured by ECERS or ECERS-R. One study that utilized the CLASS tool found that inclusive classrooms had higher mean scores for the Emotional Support and Classroom Organization domains, while classrooms that only served typically developing children had a higher score for the Instructional Support domain. The mean domain scores for Emotional Support and Classroom Organization were in the mid-range quality across classroom types. However, the mean quality score for Instructional Support was in the mid-range for classrooms that only served typically developing children and low range for inclusive classrooms. These findings suggest that while both types of classrooms generally provided good quality care, there may be areas for improvement in terms of instructional support, particularly in inclusive classrooms. Further research is needed to explore the factors that contribute to these differences and to identify strategies for improving the quality of early childhood education for all children.

Of the thirteen studies that utilized one of the ECERS measures, twelve were conducted in the United States, while one was conducted in Denmark. The studies conducted in the United States reported overall subscale scores that were in the good quality range (between 4 and 6), while the study conducted in Denmark reported overall subscale scores that were in the minimal quality range (between 2 and 4). Across the six studies that utilized the CLASS, three were conducted in the US, two were conducted in Spain, and one was conducted in Portugal. The reported mean domain scores for Emotional Support and Classroom Organization were in the medium quality range across the countries, while the mean score for Instructional Support domain was in the low-quality range.

The examination of the seven studies utilizing the ICP reveals significant geographical disparities in preschool classroom quality, which prompts a deeper inquiry into the factors influencing these outcomes. Among the studies, three were conducted in the United States, two in Greece, one in Türkiye, and one in Botswana. Findings from the United States show that average quality scores range from Poor to Good, while in Greece and Türkiye, they predominantly fall within the Poor to Minimal range. In Botswana, mean quality scores ranged from Minimal to Good. Although comparing the progress of different countries in promoting inclusion is challenging due to varying interpretations of inclusion in national policies and among researchers, this variability highlights the influence of different socio-cultural contexts and educational policies. Pianta et al. (2009) noted that systemic factors, such as teacher training and curriculum standards, play a crucial role in determining classroom quality. The impact of poor classroom quality in early childhood settings is significant, potentially affecting children's developmental trajectories and long-term educational outcomes (Burchinal, 2016). Therefore, further research is needed to explore the unique educational strategies employed in various countries.

These findings suggest that there are differences in preschool classroom quality across different countries, as measured by the ECERS, CLASS, and the ICP. While preschool classrooms in the United States generally scored well on the ECERS measures, there were differences in quality between the United States and Denmark on this measure. On the CLASS measure, preschool classrooms across countries scored similarly across all domains. For the ICP measure, the United States had higher scores overall than Botswana, Greece, and Türkiye, but there were variations in quality within the United States and within the other countries. These findings suggest that efforts are needed to improve preschool classroom quality globally and that further research is needed to understand the factors that contribute to differences in quality across countries. The EASNIE (2014) also stressed that teacher collaboration and professional development are key in raising these quality levels. Such approaches can help improve classroom environments, making them more inclusive and responsive to diverse needs.

Limitations

First, this study only included articles published in English, which may have limited the range of studies reviewed. There may be studies conducted in the international context that could provide additional evidence for the quality of inclusive preschool classrooms. Second, some relevant literature may have been excluded because the researchers did not clearly report whether the classroom environment was inclusive or not. This could have resulted in the exclusion of studies that met the other inclusion criteria

but did not explicitly state that the classroom included both children with and without disabilities. Finally, the study did not provide a detailed analysis of the quality of the articles included in the review.

Implications for Future Research and Practice

The current systematic review provides valuable insights into the characteristics of studies focused on examining quality of inclusive preschool classrooms. The lack of consistent reporting of classroom and teacher demographics in studies focused on inclusive preschool classrooms is a significant limitation in understanding the characteristics of successful inclusive classrooms and limits the generalizability of the findings. Therefore, researchers should prioritize collecting and reporting detailed information on classroom characteristics and teacher demographics, including age, years of experience, and training on inclusion. This information could provide valuable insights for developing effective professional development programs for teachers in relation to improving classroom quality. Future studies should also consider including more diverse samples of teachers and classrooms to increase the generalizability of the findings. Additionally, the development of a standardized method for reporting classroom and teacher demographics in studies focused on inclusive preschool classrooms could improve the consistency and accuracy of future research in this area. Inconsistent reporting practices regarding the roles and training of individuals who conduct classroom quality measures raise concerns about the reliability and validity of the data collected. Future studies should provide more detailed information about the training and credentials of observers to ensure that data is collected using valid and reliable procedures. The limited reporting of the interobserver agreement reliability data across studies is another limitation that must be addressed in future research. Researchers should collect and report IOA data systematically to improve the rigor and credibility of their findings. The findings of the current review highlight differences in preschool classroom quality across different countries. It would be valuable to investigate country-specific factors that may be affecting preschool quality, such as funding, teacher training, and cultural differences, to develop tailored interventions to improve quality. Finally, future research should focus on developing a quality appraisal tool specifically designed for observational studies. This would fill a critical gap in the field and provide researchers with a standardized method to assess and compare the quality of observational research more effectively.

The results also suggest that inclusive classrooms may need improvement in the area of instructional support. Educators and policymakers should focus on identifying strategies to improve instructional support in inclusive classrooms to ensure that all children receive high-quality early childhood education. To ensure successful implementation of inclusive practices, professional development programs should be designed to equip educators with the necessary skills to support diverse learners. Training should focus on evidence-based strategies for fostering social interactions, adapting curricula, and implementing individualized support for children with disabilities. By investing in ongoing training and support, educators will be better prepared to create high quality inclusive environments that meet the needs of all children. The EASNIE (2014) further emphasizes that continuous professional development should incorporate collaboration between educators and families to create personalized support systems for each child.

The use of established quality assessment tools, such as the Inclusive Classroom Profile, can provide valuable insights into the effectiveness of inclusive practices. Regular evaluations of classroom quality can help educators identify areas for improvement and track progress over time. Schools and programs should adopt a culture of continuous assessment to enhance the overall quality of inclusive education. The EASNIE (2014) supports the idea of embedding ongoing assessments into daily practices, ensuring that inclusivity and quality are monitored at every stage.

Collaboration among educators, families, and support services is crucial for creating effective inclusive environments. Schools should promote open communication with families and involve them in the decision-making process regarding their child's education. Building strong partnerships with families can enhance support systems for children with disabilities, ultimately leading to better educational outcomes. Policymakers should advocate for the allocation of resources to support high-quality inclusive

programs. This includes funding for professional development, classroom materials, and resources that facilitate inclusive practices. By creating policies that prioritize inclusivity and classroom quality, stakeholders can ensure that all children have access to high-quality early childhood education.

Declarations

Authors' Declarations

Acknowledgements: Not applicable.

Authors' contributions: The study was conceptualized and the paper was written by the first and second authors. The third author contributed to the completion of interrater reliability coding.

Competing interests: The authors declare that they have no competing interests.

Funding: Not applicable.

Ethics approval and consent to participate: Not applicable.

Publisher's Declarations

Editorial Acknowledgement: The editorial process of this article was completed under the editorship of Dr. Ibrahim H. Acar through a double-blind peer review with external reviewers.

Publisher's Note: Journal of Childhood, Education & Society remains neutral with regard to jurisdictional claims in published maps and institutional affiliation.

References

* Studies that met all inclusion criteria.

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

*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(2019), 81–92. <https://doi.org/10.1016/j.ecresq.2019.05.005>

*Aguiar, C., Moiteiro, A. R., & Pimentel, J. S. (2010). Classroom quality and social acceptance of preschoolers with disabilities. *Infants and Young Children*, 23(1), 34–41. <https://doi.org/10.1097/TYC.0b013e3181c9766e>

*Bakkaloglu, H., Sucuoglu, B., & Yilmaz, B. (2019). Quality of inclusive preschool classrooms: Predictive variables. *Education and Science*, 44(199), 223–238. <https://doi.org/10.15390/EB.2019.7839>

Burchinal, M. (2016). Large-scale evaluations of child care as a context for development: Implications for research and practice. In K. E. Sanders & A. W. Guerra (Eds.), *The culture of child care: Attachment, peers, and quality in diverse communities* (pp. 205–228). Oxford University Press.

Burchinal, M. R., Roberts, J. E., Riggins, R., & Zeisel, S. A. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development*, 71(2), 339–357. <https://doi.org/10.1111/1467-8624.00149>

*Buysse, V., Wesley, P. W., Bryant, D., & Gardner, D. (1999). Quality of early childhood programs in inclusive and noninclusive settings. *Exceptional Children*, 65(3), 301–314. <https://doi.org/10.1177/001440299906500302>

*Cadima, J., Aguiar, C., & Barata, M. C. (2018). Process quality in Portuguese preschool classrooms serving children at-risk of poverty and social exclusion and children with disabilities. *Early Childhood Research Quarterly*, 45(2018), 93–105. <https://doi.org/10.1016/j.ecresq.2018.06.007>

*Campbell, P., Milbourne, S., Silverman, C., & Feller, N. (2005). Promoting inclusion by improving child care quality in inner-city programs. *Journal of Early Intervention*, 28(1), 65–79. <https://doi.org/10.1177/105381510502800107>

*Chhabra, S., Bose, K., & Chadha, N. (2018). Early childhood educators' perspectives and practices about inclusion of children with special needs in Botswana. *Journal of Research in Childhood Education*, 32(2), 234–249. <https://doi.org/10.1080/02568543.2018.1425223>

*Classen, A. I., & Westbrook, A. (2022). Professional credential program: Impacting early childhood inclusive learning environments. *International Journal of Inclusive Education*, 26(7), 719–736. <https://doi.org/10.1080/13603116.2020.1717652>

*Clawson, C., & Luze, G. (2008). Individual experiences of children with and without disabilities in early childhood settings. *Topics in Early Childhood Special Education*, 28(3), 132–147. <https://doi.org/10.1177/0271121407311482>

Clifford, R. M., Reszka, S. S., & Rossbach, H.-G. (2010). *Reliability and validity of the early childhood environment rating scale*. FPG Child Development Institute.

- *Coelho, V., Cadima, J., & Pinto, A. I. (2019). Child engagement in inclusive preschools: Contributions of classroom quality and activity setting. *Early Education and Development*, 30(6), 800–816. <https://doi.org/10.1080/10409289.2019.1591046>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cunningham, D. D. (2010). Relating preschool quality to children's literacy development. *Early Childhood Education Journal*, 37(6), 501–507. <https://doi.org/10.1007/s10643-009-0370-8>
- DEC/NAEYC. (2009). *Early childhood inclusion: A joint position statement of the Division for Early Childhood (DEC) and the National Association for the Education of Young Children (NAEYC)*. Chapel Hill: The University of North Carolina, FPG Child Development Institute. <https://doi-org.libproxy.uncg.edu/10.1177/1096250609344935>
- Diamond, K. E., & Huang, H.-H. (2005). Preschoolers' ideas about disabilities. *Infants and Young Children*, 18(1), 37–46.
- European Commission. (2013). *Support for children with special educational needs (SEN)*. Employment, Social Affairs & Inclusion.
- European Agency for Special Needs and Inclusive Education (EASNIE). (2014). Five key messages for inclusive education. <https://www.european-agency.org/sites/default/files/Five%20Key%20Messages%20for%20Inclusive%20Education.pdf>
- European Agency for Special Needs and Inclusive Education (EASNIE). (2017). *Inclusive Early Childhood Education: Literature Review*. Denmark. <https://www.european-agency.org/sites/default/files/IECE%20Literature%20Review.pdf>
- European Agency for Special Needs and Inclusive Education (EASNIE). (2024, October). *Inclusive early childhood education*. <https://www.european-agency.org/activities/inclusive-early-childhood-education>
- Environment Rating Scales Institute (ESRI), Inc. (n.d.) *Information on Authorized Trainers for the Environment Rating Scales (ERS)*. Retrieved from <https://www.ersi.info/PDF/Info%20on%20Authorized%20ERS%20trainers.pdf>
- *File, N., & Kontos, S. (1993). The Relationship of program quality to children's play in integrated early intervention settings. *Topics in Early Childhood Special Education*, 13(1), 1–18. <https://doi.org/10.1177/027112149301300104>
- *Fyssa, A., & Vlachou, A. (2015). Assessment of quality for inclusive programs in Greek preschool classrooms. *Journal of Early Intervention*, 37(3), 190–207. <https://doi.org/10.1177/1053815115606908>
- Green, K., Terry, N., & Gallagher, P. (2014). Progress in language and literacy skills among children with disabilities in inclusive early reading first classrooms. *Topics in Early Childhood Special Education*, 33(4), 249–259. <https://doi.org.libproxy.uncg.edu/10.1177/0271121413477498>
- *Grisham-Brown, J., Cox, M., Gravil, M., & Missall, K. (2010). Differences in child care quality for children with and without disabilities. *Early Education and Development*, 21(1), 21–37. <https://doi.org/10.1080/10409280902783491>
- Harms, T., & Clifford, R. M. (1980). *Early Childhood Environment Rating Scale*. Teachers College Press.
- Harms, T., Clifford, R., & Cryer, D. (2005). *Early Childhood Environment Rating Scale—Revised edition*. Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (2015). *Early Childhood Environment Rating Scale, Third Edition (ECERS-3)*. Teachers College Press.
- *Hestenes, L., Cassidy, D., Shim, J., & Hegde, A. (2008). Quality in inclusive preschool classrooms. *Early Education and Development*, 19(4), 519–540. <https://doi.org/10.1080/10409280802230973>
- Holahan, A., & Costenbader, V. (2000). A comparison of developmental gains for preschool children with disabilities in inclusive and self-contained classrooms. *Topics in Early Childhood Special Education*, 20(4), 224–235. <https://doi.org.libproxy.uncg.edu/10.1177/027112140002000403>
- Individuals with Disabilities Education Act (IDEA) Regulations, 34 CFR § 300.1, (2004).
- *Jeon, H.-J., Langill, C. C., Peterson, C. A., Luze, G. J., Carta, J. J., & Atwater, J. B. (2010). Children's individual experiences in early care and education: Relations with overall classroom quality and children's school readiness. *Early Education and Development*, 21(6), 912–939. <https://doi.org/10.1080/10409280903292500>
- Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press
- *Keesbury, S. A. (2015). Student outcomes in a blended preschool program. *Journal of the American Academy of Special Education Professionals*, 2015(1), 147–173.
- *Kontos, S., Moore, D., & Giorgetti, K. (1998). The ecology of inclusion. *Topics in Early Childhood Special Education*, 18(1), 38–48. <https://doi.org/10.1177/027112149801800107>
- *La Paro, K. M., Sexton, D., & Snyder, P. (1998). Program quality characteristics in segregated and inclusive early childhood settings. *Early Childhood Research Quarterly*, 13(1), 151–167. [https://doi.org/10.1016/S0885-2006\(99\)80030-1](https://doi.org/10.1016/S0885-2006(99)80030-1)
- Lawrence, S., Smith, S., & Banerjee, R. (2016). *Preschool inclusion: Key findings from research and implications for policy*. <https://academiccommons.columbia.edu/doi/10.7916/D8571C1C>

- Ledford, J., Lane, J., & Severini, K. (2018). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment*, 19(1), 4-17. <https://doi.org/10.1017/Brlmp.2017.16>
- *Muccio, L. S., Kidd, J. K., White, C. S., & Burns, M. S. (2014). Head Start instructional professionals' inclusion perceptions and practices. *Topics in Early Childhood Special Education*, 34(1), 40-48. <https://doi.org/10.1177/0271121413502398>
- Musgrove. (2012). *United States Department of Education, Office of Special Education Programs (OSEP)* [Dear Colleague Letter]. Preschool LRE. Retrieved from <http://www2.ed.gov/policy/speced/guid/idea/memosdcltrs/preschoollre22912.pdf>
- *Næsby, T. (2020). The quality of inclusion in Danish preschools. *Early Child Development and Care*, 190(10), 1561-1573. <https://doi.org/10.1080/03004430.2018.1541320>
- Odom, S., Buysse, V., & Soukakou, E. (2011). Inclusion for young children with disabilities: a quarter century of research perspectives. *Journal of Early Intervention*, 33(4), 344-356. <https://doi-org.libproxy.uncg.edu/10.1177/1053815111430094>
- Office of Special Education and Rehabilitative Services (ED). (2011). *33rd Annual Report to Congress on the Implementation of the "Individuals with Disabilities Education Act," 2011*.
- Office of Special Education and Rehabilitative Services (ED). (2021). *43rd Annual Report to Congress on the Implementation of the "Individuals with Disabilities Education Act," 2021*.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, 372(71), 1-9. <https://doi.org/10.1136/bmj.n71>
- *Pelatti, C. Y., Dynia, J. M., Logan, J. A. R., Justice, L. M., & Kaderavek, J. (2016). Examining quality in two preschool settings: Publicly funded early childhood education and inclusive early childhood education classrooms. *Child and Youth Care Forum*, 45(6), 829-849. <https://doi.org/10.1007/s10566-016-9359-9>
- Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest*, 10(2), 49-88. <https://doi.org/10.1177/1529100610381908>
- Pianta, R., Downer, J., & Hamre, B. (2016). Quality in early education classrooms: Definitions, gaps, and systems. *The Future of Children*, 26(2), 119-137. <https://doi.org/10.1353/foc.2016.0015>
- Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Paul H. Brookes.
- Siraj-Blatchford, I., & Wong, Y. (1999). Defining and evaluating 'quality' early childhood education in an international context: dilemmas and possibilities. *Early Years*, 20(1), 7-18. <https://doi.org/10.1080/0957514990200102>
- Slot, P. (2018). *Structural characteristics and process quality in early childhood education and care: a literature review* (Ser. OECD education working papers, no.176). OECD Publishing. <https://doi.org/10.1787/edaf3793-en>
- Soukakou, E. P. (2012). Measuring quality in inclusive preschool classrooms: Development and validation of the Inclusive Classroom Profile (ICP). *Early Childhood Research Quarterly*, 27(3), 478-488. <https://doi.org/10.1016/j.ecresq.2011.12.003>
- *Soukakou, E. P., Winton, P. J., West, T. A., Sideris, J. H., & Rucker, L. M. (2014). Measuring the quality of inclusive practices: Findings from the Inclusive Classroom Profile pilot. *Journal of Early Intervention*, 36(3), 223-240. <https://doi.org/10.1177/1053815115569732>
- *Stanton-Chapman, T. L., Walker, V. L., Voorhees, M. D., & Snell, M. E. (2016). The evaluation of a three-tier model of positive behavior interventions and supports for preschoolers in Head Start. *Remedial and Special Education*, 37(6), 333-344. <https://doi.org/10.1177/0741932516629650>
- Vandell, D., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Madison: Institute for Research on Child Poverty, University of Wisconsin-Madison.
- *Vlachou, A., & Fyssa, A. (2016). Inclusion in practice: Programme practices in mainstream preschool classrooms and associations with context and teacher characteristics. *International Journal of Disability, Development & Education*, 63(5), 529-544. <https://doi.org/10.1080/1034912X.2016.1145629>
- Whitebook, M., McLean, C., Austin, L.J.E., & Edwards, B. (2018). *Early Childhood Workforce Index-2018*. Berkeley, CA: Center for the Study of Child Care Employment, University of California, Berkeley. <http://csce.berkeley.edu/topic/early-childhood-work-force-index/2018/>
- Wolery, M., Pauca, T., Brashers, M. S., & Grant S. (2000). *Quality of inclusive experiences measure*. Frank Porter Graham Child Development Center, University of North Carolina.
- *Yeomans-Maldonado, G., Justice, L. M., & Logan, J. A. R. (2019). The mediating role of classroom quality on peer effects and language gain in pre-kindergarten ECSE classrooms. *Applied Developmental Science*, 23(1), 90-103. <https://doi.org/10.1080/10888691.2017.1321484>