Early Childhood Research Quarterly 55 (2021) 35–51

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Relating early care and education quality to preschool outcomes: The same or different models for different outcomes?

Margaret Burchinal[a,*], Kylie Garber[b], Tiffany Foster[b], Mary Bratsch-Hines[c], Ximena Franco[c], Ellen Peisner-Feinberg[d]

[a] Center for Advanced Teaching and Learning, Curry School of Education and Human Development, University of Virginia, 405 Emmet St S, Charlottesville, VA 22904, United States
[b] Department of Psychology & Neuroscience, University of North Carolina at Chapel Hill, 235 E Cameron Ave Chapel Hill, NC 27514, United States
[c] Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill, CB 8180 Chapel Hill, NC 27599-8180, United States
[d] School of Education, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, United States

ARTICLE INFO

Article history:
Received 4 March 2019
Received in revised form 15 October 2020
Accepted 26 October 2020
Available online 21 November 2020

Keywords:
Early care & education
Preschool classroom quality
Classroom Assessment Scoring System
CLASS
Language Interaction Snapshot LISn

ABSTRACT

Although high-quality early care and education (ECE) is widely accepted as one of the most effective means for promoting early learning and development, many ECE programs have limited impact perhaps because of issues with how ECE quality is defined and measured. This study seeks to expand definitions of ECE quality by asking which preschool ECE quality dimensions relate to gains in which developmental outcomes and contrasting measurements using classroom-level ratings and child-level behavior counts. The sample includes 366 children in 63 randomly selected prekindergarten (pre-k) classrooms in six rural counties in a Southeast state. In the fall and spring, children were administered tests of academic achievement, language skills, and executive functions, and teachers rated their social skills. In the winter, the quality of teacher-child interactions was rated at the classroom level and the frequency that the child experienced complex teacher talk, domain-specific instructional activities, and whole group settings were counted at the child level. Two sets of analyses related gains in outcomes to the classroom-level ratings in both analyses and to behavior counts summarized at the classroom level and at the child level in separate analyses. Results indicated that (1) different ECE dimensions related to gains in different outcomes; (2) ECE quality measures based on observing the selected experiences of individual children provide as strong or stronger associations with child outcomes than do ratings of teacher-child interactions; and (3) it may be necessary to measure experiences of individual children if those experiences are likely to vary markedly among children in the same classroom.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

High-quality early care and education (ECE) is thought to be one of the best policy levers to improve school success for children from low-income families (Heckman, 2011; Yoshikawa et al., 2013). "High-quality" ECE programs improve school readiness skills, especially basic reading and math skills, but those effects fade over time (Burchinal, Magnuson, Powell, & Hong, 2015; Phillips et al., 2017). ECE quality is currently defined largely in terms of the extent to which children experience sensitive and responsive interactions with teachers (Hamre, 2014; Mashburn et al., 2008). However, global measures of the quality of ECE teacher-child interactions relate quite modestly, at best, to child outcomes, raising questions about whether ECE quality models should also include more specific types of ECE experiences that may promote different child outcomes (Burchinal, 2018). To examine this question, this study related ECE quality dimensions to child outcomes to test whether ECE models for specific outcomes should include both the ECE experiences thought to promote all outcomes and those experiences thought to enhance a specific outcome. This question was examined among children attending a state-funded pre-kindergarten program in rural areas in a Southeastern state.

2. Different Conceptual Models for Specific Child Outcomes

The current ECE quality model focuses on pathways from "structural quality" to "process quality" to child outcomes (Burchinal, 2018). Process quality is defined as what the child experiences

* Corresponding author.
E-mail address: kqu4rg@virginia.edu (M. Burchinal).

https://doi.org/10.1016/j.ecresq.2020.10.005
0885-2006/© 2020 Elsevier Inc. All rights reserved.

in the ECE setting, especially in terms of the quality of interactions between teachers and the children in the classroom. Typically, process quality is defined as the quality of classroom management practices and the level of emotional and instructional support provided by the teacher (Hamre, 2014). Preschool teachers, however, engage in a variety of other activities that could be also be considered dimensions of quality. Teachers decide how often they provide instructional content (e.g., literacy, math) and how to deliver instructional content in terms of the setting (e.g., whole group, small group, centers). They also determine the frequency and type of language exchanges and emotional tone used overall and with individual children. Each of these quality dimensions appears to promote different child outcomes, suggesting they too should be considered in ECE quality models for specific child outcomes.

Prior studies examined both these general and specific dimensions of ECE quality to understand how they promote different child outcomes. For example, the quality of teacher-child interactions is thought to be foundational for promoting all child outcomes, whereas instructional time in a specific content area is thought to be critical for acquiring basic academic skills (Burchinal et al., 2015). Despite an understanding that cognitive, academic, and social skills are enhanced by different aspects of the ECE experience, most research and monitoring as well as much professional development is based on an ECE quality model that has limited scope and provides very modest prediction of child outcomes (Burchinal, 2018). This study will examine multiple ECE quality indicators in relation to a full range of outcomes to ask whether some quality dimensions are related to all outcomes while other quality dimensions are related to specific outcomes.

As detailed below, we argue that at least two quality dimensions, quality of teacher-child interaction and the proportion time in whole group activities, serve as predictors of all outcomes and at least two quality dimensions, the quality of teacher-child language and the instructional time and quality, serve as predictors for specific outcomes. Although these ECE quality dimensions do not exhaustively represent aspects of the ECE experience that contribute to early learning and development, they were selected based on evidence that they related to the acquisition of skills during early childhood in prior studies. In this study, all quality dimensions are fit to all outcomes to examine the extent to which the hypothesized dimensions emerge as significant predictors for specific outcomes.

2.1. ECE Quality Dimensions Thought to Impact All Child Outcomes

2.1.1. Quality of Teacher-Child Interactions

Currently, ECE quality is typically defined in terms of teacher-child interactions based on developmental theories, and most quality assessments and educator professional development efforts focus on this ECE quality dimension (Burchinal et al., 2015). Ecological or sociocultural developmental models posit that responsive, sensitive caregiving is foundational for cognitive, academic, and social development (Bronfenbrenner & Morris, 2006; Vygotsky, 2001). Responsive, sensitive caregiving is defined in terms of the frequency and quality of teacher-child interactions (Hamre, 2014). Empirical results indicate very modest, but statistically significant, improvements in language, academic, executive function (EF), and social skills when the quality of teacher-child interactions is higher (Burchinal, Vandergrift, Pianta, & Mashburn, 2010; Curby, Rimm-Kaufman, & Ponitz, 2009; Downer, Sabol, & Hamre, 2010; Hamre, 2014; Howes et al., 2008; Mashburn et al., 2008; NICHD Early Child Care Research Network, 2006).

2.1.2. Classroom Setting

In addition to the quality of teacher-child interactions, the amount of time children spend in whole-group activities likely impacts the acquisition of cognitive, academic, EF, and social skills in preschool classrooms. While most preschool classrooms provide some instruction using both whole-group settings and centers, most classrooms favor one of these two setting for much of their instruction (Cabell, DeCoster, LoCasale-Crouch, Hamre, & Pianta, 2013; Chien et al., 2010; Fuligni, Howes, Huang, Hong, & Lara-Cinisomo, 2011). Spending less time in whole-group activities is related to larger gains in early reading and math skills for predominantly low income children in Head Start or pre-k programs (Clements & Sarama, 2007; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011), predominantly middle-income children in community care (Connor, Morrison, & Slominski, 2006), and in prior analyses of data from this study (Bratsch-Hines, Burchinal, Peisner-Feinberg, & Franco, 2019). This likely is because children's engagement is much higher during center time than it is during whole-group activities (Powell, Burchinal, File, & Kontos, 2008). Other research found that more time in whole group activities related to higher rates of off-task and aggressive behaviors among kindergarteners (Rimm-Kaufman, La Paro, Downer, & Pianta, 2005), suggesting that time in whole group may be equally or more detrimental for preschoolers.

2.2. ECE Quality Dimensions that Promote Specific Child Outcomes

In contrast, different ECE practices promote the acquisition of specific skills during early childhood. Language skills are found to be promoted primarily through multi-turn conversations between caregivers and children in which caregivers elaborate on the child's comments, ask open-ended questions, and elicit further comments from the child (Cabell, Justice, McGinty, DeCoster, & Forston, 2015; Pentimonti et al., 2017; Wasik, Bond, & Hindman, 2006; Weizman & Snow, 2001; Whorrall & Cabell, 2016). Executive functioning skills are enhanced when children engage in activities that require cognitive flexibility (Raver & Blair, 2016). Early reading and math skills are acquired when children spend more time in content-specific instruction (Burchinal, Zaslow, & Tarullo, 2016), especially when the instruction introduces skills sequentially and provides many opportunities to master each skill before moving on to teaching the next skills (Bierman et al., 2008; Clements & Sarama, 2007; Powell, Diamond, Burchinal, & Koehler, 2010). Social skills are promoted through supportive interactions that model regulating negative emotions and avoid harsh punitive exchanges (Bierman et al., 2008; Burchinal et al., 2010; Hamre, 2014; Jones, Bouffard, & Weissbourd, 2013; Raver, 2003). Thus, in addition to the general quality of teacher-child interactions and, perhaps, time in whole group settings, ECE quality models for specific outcomes should also reflect the specific types of experiences that promote that outcome. As detailed below, we argue that we should include the quality of teacher-child language exchanges in a model for language development and the instructional time and quality in a model for academic skills.

2.2.1. Teacher-child language exchanges

Measuring the frequency and quality of teacher-child language exchanges may be especially important for understanding whether ECE experiences will improve language skills. Several studies demonstrate that preschoolers gained language when they experienced more frequent responsive or decontextualized language exchanges with ECE teachers (Abt Associates, Inc, 2007; Cabell et al., 2015; Hindman & Wasik, 2012; Justice, Jiang, & Strasser, 2018). Behavior counts of how frequently teachers engaged in responsive and complex language exchanges predicted gains in language skills

in the evaluation of the First 5 LA initiative to improve quality of ECE in Los Angeles (Abt Associates, Inc., 2007). Children in the First 5 LA programs showed larger gains when their teachers facilitated conversations with multiple turns or used more decontextualized language. Ratings of the teachers' linguistic responsiveness, but not of decontextualized language or quality of teacher-child interactions, predicted gains in vocabulary skills in a recent preschool study (Justice et al., 2018).

2.2.2. Instructional time

Observing the amount of time children spent in literacy and math instruction and the content of that instruction is important to understand the acquisition of early academic skills in preschool classrooms. Although the content and quality of instruction almost certainly matters more than the quantity, several studies indicated that measures of instructional time related to gains in literacy and math skills. Measures of the amount of time children spent in reading and math activities were related to gains in those academic skills in a study of pre-K in 11 states (Chien et al., 2010). Behavior counts of time spent in instruction in reading and math related to gains in language, literacy, or math skills in a sample of predominantly middle-income children (Connor et al., 2006), a small meta-analysis of six large ECE studies of predominantly low-income children (Burchinal et al., 2016), and in prior analyses of the data from this study (Bratsch-Hines et al., 2019).

In summary, the primary goal of this study was to test whether gains in different child outcomes were all related to some ECE dimensions and differentially related to other ECE dimensions. We selected ECE quality dimensions that were collected in our longitudinal study because they were shown to relate to child outcomes in prior studies but recognize there are other ECE quality dimensions that our observational tools did not measure. We anticipated that the quality of teacher-child interactions and time in whole group were general ECE quality measures important for the development of all skills. We also hypothesized that teacher-child language exchanges were especially important for developing language skills and instructional time was especially important for acquiring content-specific academic outcomes.

2.3. Approaches to Measuring ECE Quality

How different ECE quality dimensions are measured varies across the selected dimensions, conflating what is measured and how it is assessed. For example, quality of teacher-child interactions is typically measured by ratings of the whole classroom (Hamre, 2014). These ratings reflect the foundational role the teacher(s) play in creating the ECE experience for preschoolers. Counts of specific teacher and child behaviors are typically used to measure the time spent in different settings and instruction in specific content areas (Howes et al., 2008). These behavior counts describe individual differences in ECE experiences thought to play an important role in early learning and development. Thus, our examination of whether different ECE quality dimensions predict different child outcomes involves different approaches to measuring those dimensions.

2.3.1. Ratings

Most quality rating systems measure the sensitivity and responsiveness of caregivers in their interactions with children and access to age appropriate activities. The most widely-used ECE quality tools, the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) and the Early Childhood Environmental Rating Scales (ECERS-R; Harms, Clifford, & Cryer, 1998; ECERS-3 Harms, Clifford, & Cryer, 2014), measure the quality of the ECE classroom by observing an entire classroom, typically during the morning. The observer rates selected indicators of ECE quality at the classroom level. A recent review of the literature (Burchinal, 2018) indicated that gains in children's cognitive, academic, and social skills are significantly, albeit somewhat inconsistently and quite modestly, related to CLASS and ECERS scores.

2.3.2. Behavior counts

Behavior counts involve multiple cycles of iteratively observing individual children for a short period of time each. In time-sampling observations, the short period of observation time per child per cycle is divided into blocks of time (e.g., 30 seconds). The observer typically records whether during each block of time the target child exhibits specific types of behaviors or whether the teachers or peers exhibit specific behaviors toward the target child. They typically record the child's activity setting and instructional focus and can include specific characteristics of how teachers interacted with the target child and how the target child interacted with peers. The number of children observed per classroom, what behaviors are counted, and the amount of time each child is observed varies across coding systems. Often the counts of behavior are tallied across children in a classroom to create classroom-level behavior counts.

One measure, the Emerging Academic Snapshot (EAS; Ritchie, Howes, Kraft-Sayre, & Weiser, 2001), and its adaptation to describe language exchanges, the Language Interaction Snapshot (LISn; Atkins-Burnett, Sprachman, & Caspe, 2010) has been used in multiple studies and evaluations. With the LISn, children are observed for 10 30-second blocks during each cycle, recording the types of language exchanges they experience with teachers and peers, the activity setting, and the instructional content. Typically, at least four cycles are collected per child. Time individual children spent in language activities measured with the EAS was related to larger gains in teacher ratings of emergent literacy skills (Howes et al., 2008). The frequency of child-level language exchanges involving eliciting (decontextualized language) or elaboration (linguistically responsive language) assessed with the LISn were related to moderate to large gains in language skills among predominantly low-income Spanish-English dual language learners (Abt Associates, Inc., 2007).

Perhaps the most widely used behavior count system was developed by Farran and colleagues. They developed the Teacher Observation in Preschool (TOP; Bilbrey, Vorhaus, Farran, & Shufelt, 2010) to measure the teacher's emotional tone in interactions with children and the Child Observation in Preschool (COP; Farran & Son-Yarbrough, 2001) to describe the individual child's engagement and involvement in academic activities and the quality of interactions with teachers and peers. Iteratively for 3 seconds each, the teacher is observed with the TOP and each child in the classroom is observed with the COP. This cycle of observing the teacher and each child repeated 20 times. The mean of children's COP data is computed to represent the classroom. TOP and COP behavior counts are related to short- and long-term child outcomes. For example, ratings of the tone of the interactions and the frequency of engagement and different types of instructional activities as measured by the COP and TOP were associated in a longitudinal study of a pre-K program with both short- and long-term gains in self-regulation (Fuhs, Farran, & Nesbitt, 2013).

2.3.3. Contrasting classroom ratings and behavior counts

At least two recent studies included both classroom-level ratings on the CLASS and behavior counts and ratings collected on individual children or small groups of children averaged to the classroom level. One study contrasted the extent to which two indices of teacher-child language exchanges and the overall quality of teacher-child interactions related to gains in preschoolers' grammar and vocabulary skills in 49 classrooms that served predominantly low-income children (Justice et al., 2018). The first index of language exchanges described teachers' linguistic

responsiveness. It was measured through aggregated ratings of 15 30-second snippets of small-group interactions videotaped in the fall, winter, and spring of the preschool year. Ratings captured the degree to which teachers commented and expanded on children's language to encourage and maintain small-group conversations. The second index of language exchanges described language complexity based on counting the teacher's utterance length, vocabulary size, and grammatical complexity in those small groups. Both the linguistic responsivity and the language complexity variables were coded at the small group level and aggregated to form classroom composites. The overall quality of teacher-child interactions was measured by the CLASS. Findings suggested that growth in vocabulary skills was significantly larger when the teachers displayed higher levels of linguistic responsivity but was not significantly related to the complexity of their language or the overall quality of teacher-child interactions (Justice et al., 2018).

Another study related gains in school readiness skills to both the classroom-level of teacher-child interactions using the CLASS and classroom-level counts of instructional activities and settings of target children using the Behavioral Coding System (BCS) (Pianta et al., 2020). Pianta and colleagues (2020) developed the BCS to measure academic instructional activities and settings. The BCS involves observing classrooms on two days for 40 cycles, iteratively observing each of the participating children for 30-second snapshots, and recording the activity setting and instructional content. The mean of the child-level BSC data was computed for each classroom and used, along with the CLASS, to predict gains in child outcomes in a study of over 120 pre-K classrooms (Pianta et al., 2020). The proportion of time spent in teacher-led activities was significantly, but modestly, related to larger gains in early reading skills, but time spent in academic activities was negatively related to teacher-reported social skills and task orientation and direct assessments of executive functioning (EF). In contrast, classroom-level ratings of the quality of teacher-child interactions were modestly related to gains in math and EF.

In summary, classroom-level ratings of the quality of teacher-child interactions are widely used to measure ECE quality and typically are quite modestly related to at least some child outcomes, but behavior counts of individual child experiences appear to provide additional prediction of gains in preschool child outcomes. A secondary goal of this study is to examine the extent to which behavior counts of specific aspects of the ECE experience relate to gains in specific child outcomes.

3. The Present Study

The primary goal of this study was to ask whether a more differentiated ECE quality model related to the acquisition of specific skills among children attending a state pre-kindergarten program in rural areas in the Southeast. Few studies have examined multiple dimensions of ECE quality simultaneously to consider their relative importance for the development of different school readiness outcomes. We hypothesized that gains in all outcomes would be larger when children experienced higher quality teacher-child interactions and spent less time in whole group activities; gains in language skills would be larger when teachers provided more frequent high quality language exchanges, and gains in academic skills would be larger when children spent more time in academic instruction.

A secondary exploratory goal was to ask whether child-level behavior counts of specific ECE quality dimensions related differently to child outcomes when those experiences vary markedly among children in the same classroom. We hypothesized that classroom-level aggregated behavior counts of selected ECE experiences (i.e. time spent in literacy instruction) would be related to gains in child outcomes as described above. We suspected that child-level behavior counts of the experiences that vary greatly among children in the same classroom would relate to gains in outcomes for the observed children. In particular, we anticipated that individual differences in child-level counts of complex teacher-child language exchanges would relate to individual differences in language skills because we thought teachers would talk quite differently with different children in the same classroom.

4. Methods

4.1. Participants

A sample of 63 classrooms serving children in the North Carolina Pre-Kindergarten (NC Pre-K) program in six rural counties in Central and Eastern North Carolina was randomly selected to participate in this study. The NC Pre-K program serves children in the year prior to kindergarten, and targets children from low-income families and dual language learners, among other criteria. The classrooms were randomly selected proportionate to classroom size from the list of all NC Pre-K classrooms in the six counties, with an acceptance rate of 60%. Of the 63 classrooms, 41 were located in public schools, 4 were in Head Start centers, and 18 were in community-based organizations.

Invitations to participate in the study were sent home to all parents in the selected classrooms, and four to six 4-year-old children with parental consents were randomly selected within each classroom, with oversampling of children who spoke Spanish at home. Overall, consents were received from 52% of the parents. All 366 children were assessed in the fall, and 362 of them were included in the analyses because they had both fall and spring data on at least one outcome. Altogether 311 children had fall and spring outcomes on all outcomes, 12 children had fall and spring direct assessments but were missing the fall or spring teacher ratings of social skills and self-regulation, 11 children had only fall and spring teacher ratings but were missing spring direct assessments, and the remaining 28 children were missing either the fall or spring assessments on between 1 and 4 outcomes. The four children lost to follow-up included 1 Spanish-speaking DLL child, 3 African American children, and 1 male. The analysis sample was about half male, about one-third African American and about one-third Latinx/Hispanic, and about one-third spoke a language other than English at home. Of the 121 dual language learners, 114 spoke Spanish at home. The characteristics of the children and their families are shown in Table 1.

4.2. Procedures

Widely used measures of academic achievement and language skills and a battery of EF skills were administered individually to each child in the fall and spring of their pre-K year. Teachers rated the children's social skills in the fall and spring. Experienced assessment trainers trained and certified data collectors prior to data collection. Table 2 describes the fall and spring assessments of children's academic, language, EF, and social skills.

ECE quality and practices were measured through observations of the classroom during the winter. Table 3 describes the classroom- and child-level assessments of ECE quality dimensions. Classroom visits were scheduled when instructional activities were anticipated, typically in the morning. Data collectors observed the quality of teacher-child interactions for the whole classroom on one day and the teacher-child language exchanges and types of activities for individual study children on a second day.

Table 1
Child and family characteristics.

	N	%	Missing	Mean	SD	Min	Max
Child age at entry to pre-K (years)	362		0.00%	4.54	0.31	3.94	6.13
Time between fall & spring testing	362		0.00%	0.65	0.06	0.15	0.77
Sex			0.00%				
Male	178	49%					
Female	184	51%					
Race/ethnicity			0.28%				
African American/Hispanic	11	3%					
African American/nonHispanic	120	33%					
White/Hispanic	108	33%					
White/nonHispanic	100	28%					
Other[a]	22	11%					
Home language			0.00%				
English	241	67%					
Other[b]	121	33%					
Maternal education (years)	361		0.28%	12.40	2.39	8.00	18.00
Less than high school	77	21%					
High school	104	29%					
Some college	96	27%					
Associate degree	38	11%					
Bachelor's degree or more	46	13%					

[a] The other race category included 9 Native American children, 6 Asian-American children, and 8 biracial children.
[b] The other home language category included 114 children who spoke Spanish at home and 7 children who spoke a language other than Spanish.

Table 2
Child outcomes.

	Fall			Spring			Gain	
	N	Mean	SD	N	Mean	SD	Mean	SD
WJPV: Vocabulary (*w*)	361	453.20	21.97	350	461.10	16.45	7.74	11.08
WJPV: Vocabulary (standard score)	343	94.45	14.70	346	94.48	13.52	1.22	8.14
EOW: Expressive language (raw)[a]	357	47.96	15.73	346	55.26	16.49	7.34	13.32
WJLW: Decoding skills (*w*)	362	318.70	24.66	350	340.00	23.49	21.01	19.78
WJLW: Decoding skills (standard score)	362	91.64	12.59	350	95.71	11.64	3.96	9.97
DIBELS FSF: Initial sounds (raw)	361	2.57	6.79	351	6.49	10.27	3.86	8.97
DIBELS PSF: Phoneme segmentation (raw)	360	1.85	5.75	351	4.78	9.21	2.89	8.40
WJAP: Math (*w*)	362	389.50	28.37	350	408.7	21.72	18.64	17.75
WJAP: Math (standard score)	362	94.65	13.95	349	98.49	11.51	3.55	9.28
NIH Toolbox Flanker: Inhibitory Control (raw)[a]	362	17.43	13.53	341	26.84	11.79	8.74	13.18
NIH Toolbox Flanker: Inhibitory Control (standard score)	284	100.39	14.80	340	100.75	15.43	1.50	17.54
NIH Toolbox DCCS: Cognitive Flexibility (raw)[a]	362	12.32	12.99	332	19.94	13.20	7.00	15.89
NIH Toolbox DCCS: Cognitive Flexibility (standard score)	245	100.01	11.76	331	98.16	14.53	−0.75	16.61
Teacher rating: Self-regulation	354	3.90	0.56	348	3.92	0.61	0.03	0.35
STRS: Conflict	353	1.41	0.56	348	1.45	0.70	0.04	0.55
TCRS: Acting out	354	1.53	0.69	347	1.59	0.79	0.06	0.48
TCRS: Frustration tolerance	354	3.38	0.92	347	3.47	1.00	0.10	0.73
TCRS: Task orientation	354	3.48	0.99	347	3.70	1.01	0.22	0.75
LBS: Attitude towards learning	354	1.83	0.29	347	1.81	0.33	−0.02	0.25
LBS: Attention/persistence	354	1.62	0.48	347	1.62	0.47	0.01	0.33
LBS: Competence/motivation	354	1.66	0.40	347	1.68	0.42	0.02	0.31
LBS: Strategy/flexibility	354	1.76	0.32	347	1.73	0.34	−0.03	0.28
Teacher rating: Social skills	354	4.09	0.60	348	4.22	0.56	0.14	0.43
TCRS: Assertiveness	354	3.36	1.00	347	3.60	0.98	0.24	0.78
STRS: Closeness	354	4.42	0.62	348	4.56	0.52	0.14	0.63
TCRS: Shyness/anxiety	354	1.42	0.56	347	1.40	0.60	−0.03	0.47
TCRS: Peer social skills	354	3.99	0.81	347	4.13	0.83	0.16	0.67

Note: WJPV = Woodcock-Johnson III Picture Vocabulary (Woodcock et al., 2001). EOW = Expressive One-Word Picture Vocabulary Test (Brownell, 2000). WJLW = Woodcock-Johnson III Letter-Word Identification (Woodcock et al., 2001). DIBELS FSF and PSF = Dynamic Indicators of Basic Early Literacy Skills First Sound Fluency and Phoneme Segmentation Fluency (Good, Gruba, & Kaminski, 2002). NIH Toolbox Flanker and DCCS = Flanker Inhibitory Control and Attention and Dimensional Change Card Sort (NIH Toolbox, 2015). WJAP = Woodcock-Johnson III Applied Problems (Woodcock et al., 2001). Teacher-rated self-regulation and social skills from factor analysis of scale scores from the Teacher-Child Relationship Scale (Hightower et al., 1986), Learning Behavior Scales (McDermott, 1999), and Student-Teacher Relationship Scale (Pianta, 2001).
[a] EOW and Toolbox tasks were administered in Spanish for Spanish-English dual language learners.

Table 3
ECE quality dimensions.

	Classroom level			Child level		
	N	Mean	SD	N	Mean	SD
Teacher-child interactions (CLASS)						
Total score	62	4.341	0.592			
Emotional Support	62	5.318	0.669			
Classroom Management	62	5.050	0.642			
Instructional Support	62	2.656	0.659			
Teacher-child language exchanges (LISn)	61	0.042	0.033	309	0.041	0.048
Decontextualized language	61	0.024	0.024	309	0.023	0.035
Elaborated language	61	0.021	0.019	309	0.021	0.030
3+ turns	61	0.002	0.003	309	0.001	0.008
Classroom settings (LISn)						
Small group	61	0.084	0.114	309	0.081	0.141
Whole group	61	0.365	0.183	309	0.362	0.233
Free choice/center	61	0.473	0.198	309	0.467	0.227
Instructional activities (LISn)						
Literacy activities	61	0.235	0.125	309	0.237	0.211
Sounds activities	61	0.048	0.051	309	0.046	0.105
Math activities	61	0.163	0.134	309	0.152	0.199

Note: CLASS = Classroom Assessment Scoring System (Pianta et al., 2008). LISn = Language Interaction Snapshot (Sprachman, Caspe, & Atkins-Burnett, 2010).

5. Measures

5.1. Child Outcomes

5.1.1. Woodcock Johnson III Tests of Achievement (WJIII)

WJIII (Woodcock, McGrew, & Mather, 2001) scales were administered to measure language, literacy, and math skills. The Picture Vocabulary (WJPV) subtest measures receptive and expressive vocabulary; the Letter-Word Identification (WJLW) subtest measures early reading skills such as recognizing letters and reading words; and the Applied Problems (WJAP) subtest measures numeracy math skills such as simple counting, adding, subtracting, and making comparisons. WJIII is calibrated and normed for ages 2 to 90 years, with reported reliabilities of .69 to .99. The scores used in analyses were *w* scores, which are calibrated growth scores.

5.1.2. Expressive One-Word Picture Vocabulary Test (EOW)

The EOW (Brownell, 2000) was administered to measure expressive language with an individually administered, norm-referenced test. Children are asked to name objects, actions, and concepts when viewing color images. EOW assesses the child's ability to use spoken words to name depictions of images without context. EOW takes approximately 15–20 minutes to administer and measures English and Spanish/bilingual skills for 2- to 70-year-olds. Reported reliabilities range from .94 to .98. Children were administered the EOW in English or using a bilingual Spanish/English version depending on what parents and teachers reported was the primary home language. The scores used in analyses were the average standard scores across both the English and bilingual versions.

5.1.3. Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS (Good & Kaminski, 2002) was administered to measure letter-sound knowledge using First Sound Fluency (FSF) and phonemic awareness skills using Phoneme Segmentation Fluency (PSF). The number of initial sounds of words correctly identified within one minute is measured on the FSF, with reported reliability for a pre-k sample of 0.86 (Cummings, Kaminski, Good III, & O'Neil, 2011). The number of correctly orally segmented sounds within words in one minute is assessed with the PSF, with reported reliability for a pre-k sample of 0.88 (Kaminski & Good, 1996). Raw scores on these scales were used in analyses because we used the subtests benchmarked for kindergarten, as pre-K versions were not available at the time of administration.

5.1.4. NIH Toolbox - Executive Functioning

Tasks from the NIH Toolbox (Gershon et al., 2013) were administered to capture EF skills using the Flanker Inhibitory Control and Attention Test (Flanker) and Dimensional Change Card Sort (DCCS) tests. Each task is designed to take 3-4 minutes for children three years and older. On the Flanker, a child is shown a row of fish on an iPad and asked to focus on the middle fish (target stimulus) while inhibiting attention to other stimuli. The child is asked to press the arrow that matches the direction the middle fish is pointing. Sometimes all the fish point in the same direction (congruent) and sometimes in the opposite direction from the target stimuli (incongruent). This measure had test-retest reliability estimates of 0.92 and external validity of 0.84 for 5-year-olds (Zelazo et al., 2013). On the DCCS, which measures cognitive flexibility, a child is shown two pictures on an iPad that can differ from the target picture based on two characteristics: shape and color. The child is asked to match related test pictures to the target pictures, first according to one characteristic and then, after a number of trials, to the other characteristic. This measure had a test-retest reliability estimate of 0.92 and external validity of 0.72 for 5-year-olds (Zelazo et al., 2013). The tasks were administered in English when English was the primary language and Spanish when Spanish was the primary language according to teachers and children. As part of the NIH Toolbox, raw and norm-reference scores are available, but raw scores were used in analyses to describe growth.

5.1.5. Teacher ratings of social skills

Teachers were contacted via email with a link to an online survey about individual study children in their classroom. The survey included the *Teacher-Child Rating Scale* (TCRS, Hightower, et al., 1986), the *Learning Behavior Scale* (LBS, McDermott, Green, Francis, & Stott, 1999), and the short form of the *Student-Teacher Relationship Scale* (STRS; Pianta, 2001). In this sample, scale scores from all three measures showed good internal consistency (0.91–0.95).

The TCRS is a teacher rating scale that assesses a child's social skills in seven subscales: acting out (6 items; e.g., *disruptive in class*); shyness/anxiety (6 items; e.g., *shy, timid*), learning problems (6 items; e.g., *poor work habits*), assertive social skills (5 items; e.g., *defends own views under group pressure*), task orientation (5 items; e.g., *functions well even with distraction*), frustration tolerance (5 items; e.g., *accepts imposed limits*), and peer social skills (5 items; e.g., *makes friends easily*). Teachers use a 1- to 5-point Likert scale to indicate how well each of 38 statements describe the child.

The LBS is a teacher rating scale that assesses a child's classroom learning behaviors in four subscales: competence motivation (8 items; e.g., *easily gives up on tasks*), attitude toward learning (9 items; e.g., "*don't care*" *attitude to success or failure*), attention/persistence (7 items; e.g., *doesn't stick to tasks*), and strategy/flexibility (7 items; e.g., *invents silly ways to do tasks*). Teachers use a 1- to 3-point Likert scale to indicate how well each of 29 items describe the child.

The STRS is a teacher rating scale that assesses teacher's perceptions of his/her relationship with an individual child on two subscales: conflict (8 items; e.g., *this child easily becomes angry with me*) and closeness (7 items; e.g., *this child values his/her relationship with me*). Teachers use a 1- to 5-point Likert scale to indicate how well each of 15 items describe their relationship with the child. Teacher ratings on STRS have been validated through associations with children's academic and social skills within and across years in an economic and racially mixed sample (Pianta & Stuhlman, 2004).

Many of the TCRS, LBP, and STRS scales scores were highly correlated, so a principal component analysis (PCA) was conducted to create fewer composites that are more precise than the individual scale scores. The PCA with varimax rotation yielded two factors with eigenvalues greater than one. The first factor, labeled *self-regulation*, included TCRS frustration tolerance, TCRS assertive social skills, LBS competence motivation, LBS strategy/flexibility, and reverse scores of STRS conflict and TCRS acting out, and accounted for 43-44% of the total variance in the fall and spring ratings (alpha = .91 fall and spring). The second factor, *social skills*, included STRS closeness, TCRS peer social skills, and reverse scores of TCRS shyness/anxiety, and accounted for 31-32% of the total variance in the fall and spring ratings (alpha = .79, fall and .74, spring). The scale scores of the LBS were transformed to be on the same scale as the TCRS and STRS and a mean of the scales loading on each factor was computed.

5.2. ECE quality measures

5.2.1. Classroom Assessment Scoring System

The CLASS (Pianta et al., 2008) measures the quality of teacher-child interactions in preschool and early elementary school. Ten dimensions are observed and scored into three domains: Emotional Support, Classroom Organization, and Instructional Support. Trained data collectors observe classrooms for 4 to 6 20-minute cycles followed by 10 minutes of scoring. Each dimension is rated from 1 (low quality) to 7 (high quality). All data collectors were certified as reliable by the developer and interim reliability was also computed. Twenty percent of classrooms were visited by two data collectors for reliability checks, yielding high levels of reliability for each dimension calculated as agreement within one point on the scale (> .95) and weighted kappas that ranged from acceptable to good (.48–.76; M = .65) according to Landis and Koch (1977). On each domain, intra-class correlations ranged from good to excellent (.83–.97; M = .90; Koo & Li, 2016). The total CLASS score, or the average of the ratings on Emotional Support, Classroom Organization, and Instructional Support, was used in analyses due to the high correlations among the domain scores ($.66 \leq r \leq .91$).

5.2.2. Language Interaction Snapshot (LISn)

LISn (Sprachman, Caspe, Atkins-Burnett, & Kennen, 2009) describes language interactions between the teacher and target child in English and Spanish and the classroom settings and activities in which the target child engages. The data collector (who was bilingual when observing classrooms with dual language learners) observes a target child for ten 30-second cycles (5 minutes), recording who spoke to and to whom the target child spoke, the language used, the type of verbal communication (e.g., directives, simple questions, elaborated questions), and the use of contextualized or decontextualized language during exchanges. At the end of the 5-minute "snapshot," the data collector records the settings (e.g., individual, small group, large group) and type of activities (e.g., reading, math, arts, transition) in which the target child was engaged. Each selected child is observed in turn for 5 minutes, before any child is observed for a second cycle.

Building from the protocol for the Snapshot used in the National Center for Education, Development and Learning Pre-Kindergarten Study (Early et al., 2010), we aimed to observe six children per classroom. Four 5-minute snapshots per child were collected, for a total of 20 minutes of observation per child and up to 120 minutes of observation per classroom. All data collectors obtained 80% exact agreement with a LISn trainer to be certified. In addition, 20% of classrooms were visited by two data collectors for reliability checks. Reliability was computed for 15% of each data collector's observations, with exact agreement ranging from .70 to .99 (M = .91) and weighted kappas of .50 to 1.00 (M = .71) for the selected LISn variables.

We focused on a small set of the LISn variables. The selected variables were included because they reflected the ECE experiences that we hypothesized would be related either to all outcomes or to specific outcomes. We hypothesized that experiencing more frequent complex language exchanges with the teacher would promote language skills. We created a composite measure of complex teacher talk by computing the proportion of 30-second cycles in which the teacher elicited more information from the target child individually or from a group including the target child using decontextualized language, the teacher elaborated on the child's prior comment, or the teacher engaged the target child in a conversation with three or more turns. In addition, for follow-up analyses, complex teacher talk was conceptualized as two separate variables: (1) decontextualized language, measured as its own construct; and (2) linguistic responsivity, measured from teacher elaborated talk or engagement in sustained conversation.

We also hypothesized that spending more time in whole-group settings would be negatively related to all outcomes, so we included the counts of 5-minute snapshots in which whole-group settings were observed. Finally, we hypothesized that spending more time in content-specific instruction would relate positively to gains in skills in that content area, so we included the counts of 5-minute snapshots in which literacy activities were observed (including activities involving writing, print, and sounds), in which sound-related activities were observed, and in which math activities were observed (including activities involving counting, number line, and number sense).

A mean for each child was computed to describe the ratings of the 30-second cycles within each 5-minute snapshot. Indices of internal consistency indicated very little consistency in repeated measures of the same child but much higher consistency in measures across children in the same classroom: complex teacher talk (α = .07, .74), literacy activities (α .03, .54), math activities (α = .26, .69), and whole group activities (α = .00, .83). Nevertheless, there were small, but statistically significant differences among children within the same classroom in the between- and within-classroom variability estimated by unconditional hierarchical linear models.

Our original focus was solely on classroom level data, so we maximized the number of classrooms to be observed rather than the number of children per classroom. Classroom-level summary scores were computed as the mean of the child-level language exchanges, settings, and types of activities for all children observed in the classroom. These classroom-level variables have been used in prior studies (Bratsch-Hines et al., 2019; Howes et al., 2008) to provide measures of classroom-level practices. After the data were collected, we noted that child variables were from 20% to 50% more variable than the class LISn variables, so we decided to examine both the child- and classroom-level measures of the LISn variables.

Overall, 61 classrooms were observed with both the CLASS and LISn, one classroom was observed only with the CLASS because we were not able to schedule the LISn observation, and one classroom was not observed with either measure because the teacher declined to participate in classroom observations. Of the 362 children in the analysis sample, 359 had CLASS scores, 354 had classroom-level LISn data, and, due to absences on the observation day, 309 had the child-level LISn variables.

5.2.3. Covariates

When parents signed consent forms, they were asked to provide information about the child's date of birth, sex, race, language(s) used in the home, and the mother's (or primary caregiver's) education. All parents provided this information, but about a quarter of them answered the race question by indicating they were Latinx. Covariates included a continuous measure of child's age at entry into pre-K; a continuous measure of time between fall and spring testing; sex, 0 = female, 1 = male; race, 0 = non-African American, 1 = African American; home language, 0 = English 1 = another language, typically Spanish; and a continuous measure of the mother's years of education.

5.3. Data Analysis

5.3.1. Descriptive analyses

Descriptive statistics included the sample demographics (Table 1), pre-K gain scores from fall and spring outcomes (Table 2), and the classroom quality measures (Table 3). Correlations were computed among the ECE quality dimensions measured at the classroom level and child level (Appendix Tables A1 and Table A2), and between the ECE quality dimensions and children's gains scores (Table 4).

5.3.2. Inferential analyses

Inferential analyses fit hierarchical linear models (HLM) to determine the extent to which the classroom-level and child-level ECE quality dimensions related to gains in child outcomes. A random intercept was estimated for each classroom to account for nesting of children within classrooms. Clustering of the classrooms within school districts was considered when estimating residuals. Fall-to-spring gain scores were computed using the *w* scores for the Woodcock-Johnson outcomes and the raw scores for the other measures. This approach was used because gain scores have been found to reduce selection bias more effectively than using the fall score as a covariate (NICHD Early Child Care Research Network & Duncan, 2003). Two two-level HLM models were estimated. Both models included the classroom-level rating on the CLASS total score and covariates. Behavior counts of complex teacher talk, content-specific instruction, and whole group were included in Model 1 as aggregated classroom-level LISn variables and in Model 2 as child-level LISn variables.

All continuous variables were standardized to have a mean of zero and standard deviation of one, so model coefficients can be interpreted as effect sizes. The effect sizes are interpreted as the difference in standard deviation units between the two levels of the categorical predictor variable or in a one standard deviation change in the continuous predictor variable.

The first model included both classroom-level CLASS and LISn variables at Level 2. Level 1 described the gains of the i^{th} child in the j^{th} classroom, including the child and family covariates (child age at entry to pre-K, sex, race, home language, and maternal education), and the residual for that child, ε_i. Level 2 related the Level 1 intercept to the CLASS total score and classroom-level LISn variables, and includes an error term for that classroom, $\lambda_{j0.}$

$$\text{Level1 }: Y_{ij} = \delta_{0i} + \delta_{1i} < \text{ChildCovariates}_{ij} > + \varepsilon_{ij}$$

$$\text{Level2 }: \delta_{0j} = B_{00} + B_{01}\text{CLASSTotal}_j + B_{02}$$

$$\text{Teacher Complex Language}_i + B_{03}\text{Content}$$

$$\text{Activities}_j + B_{04}\text{WholeGroup}_j + \lambda_{j0}$$

$$\delta_{1i} = B_{10}$$

The second model included the child-level LISN variables at Level 1 and the CLASS total scores at Level 2 model for the Level 1 intercept.

$$\text{Level1 }: Y_{ij} = \delta_{0j} + \delta_{1j}\text{Teacher Complex Language}_{ij}$$

$$+ \delta_{2j}\text{Content Activities}_{ij} +$$

$$\delta_{3j}\text{WholeGroup}_{ij} + \delta_{4j} < \text{ChildCovariates}_{ij} > + \varepsilon_{ij}$$

$$\text{Level2 }: \delta_{0j} = B_{00} + B_{01}\text{CLASSTotal}_j + \lambda_{j0}$$

$$\delta_{1j} = \text{B}_{10}; \delta_{2j} = B_{20}; \delta_{3j} = B_{30}; \delta_{4j} = B_{40}$$

5.3.3. Multiple imputation

Almost all children were assessed again in the spring (> 95%), so missing data on most outcome measures were minimal. All but one classroom was observed so there were minimal missing data on the measures of classroom practices. However, not all children were present when LISn data were collected and not all parents provided demographic information. Multiple imputations were conducted using the data from the 366 recruited children and the 62 observed classrooms. The EM algorithm was used to create 40 imputed data sets. All analyses were conducted with each of these data sets that included the imputed demographic and classroom quality data. Analyses were conducted only for children without missing outcome data (> 95%; Angrist, Imbens, & Rubin, 1996). Parameter estimates were combined, taking within- and between-dataset variations into account.

5.3.4. Sensitivity analyses

Four sets of sensitivity analyses were conducted. First, each of the three CLASS domain scores was examined as a separate variable in the Model 1 analyses, replacing the CLASS total score. Second, due to moderately high correlations between LISn complex teacher talk and the CLASS, we conducted separate sets of analyses to include the CLASS in the first analysis, the classroom-level LISn variables in the second analysis, and the child-level LISn variables in the final analysis. Third, to examine whether linguistic responsiveness or language complexity related to gains in language skills as contrasted by Justice and colleagues (2018), follow-up analyses were conducted. Our teacher language composite involved the frequency with which teacher elaborated or sustained conversations (linguistic responsiveness) or elicited (language complexity), so follow-up analyses examined each of these variables separately. Fourth, we examined the distributions of all variables involved in significant associations to address concerns about distributional issues. Highly skewed distributions on two of the child-level LISn variables led to testing models with square root transformations of those variables.

Table 4
Correlations between gains in child outcomes and classroom- and child-level assessments of ECE quality dimensions.

	WJ PV	EOW	WJ LW	DIBELS FSF	DIBELS PSF	WJ AP	NIH Toolbox Flanker	NIH Toolbox DCCS	Self-regulation	Social skills
Classroom-level measures										
CLASS Total	0.09	0.05	0.17***	0.01	−0.07	0.06	−0.19***	−0.07	0.10	0.11*
Emotional support	0.11*	0.04	0.14**	0.02	−0.05	0.06	−0.18***	−0.03	0.11*	0.10*
Classroom organization	0.04	0.02	0.16**	−0.05	−0.09	0.03	−0.18**	−0.10	0.09	0.08
Instruction support	0.10	0.08	0.17**	0.06	−0.02	0.08	−0.15**	−0.06	0.06	0.12*
LISn teacher complex talk	0.14**	0.09	0.16*	0.09	0.05	0.03	−0.06	−0.08	0.04	0.15**
Decontextualized language	0.12**	0.10	0.16*	0.09	0.05	0.05	−0.08	−0.08	0.01	0.10
Linguistic responsivity	0.15**	0.06	0.12*	0.07	0.05	0.00	−0.02	−0.08	0.07	0.17**
LISn literacy activities	0.07	0.16**	0.05	0.07	0.06	−0.03	0.03	−0.04	−0.10	−0.03
LISn sounds activities	−0.02	0.18**	0.09	0.15*	0.14*	0.00	0.09	0.02	0.06	0.04
LISn math activities	0.13**	0.00	0.22***	0.09	0.12*	0.10	−0.01	−0.07	0.07	0.10*
LISn whole group	−0.04	−0.14**	0.09	−0.08	−0.09	−0.14*	−0.13*	−0.09	−0.02	−0.07
Child-level measures										
LISn teacher complex talk	0.15**	0.10	00.12*	0.07	0.06	0.03	−0.07	−0.04	0.03	0.10
Decontextualized language	0.08	0.07	0.11*	0.04	0.07	0.01	−0.07	−0.05	0.01	0.05
Linguistic responsivity	0.18**	0.09	0.10	0.06	0.02	0.05	−0.07	−0.03	0.07	0.13*
LISn literacy activities	0.01	0.09	0.08	0.06	0.05	−0.05	0.00	0.02	−0.05	−0.10
LISn sounds activities	−0.04	0.06	0.05	0.14*	0.14*	−0.03	0.03	0.03	0.07	−0.01
LISn math activities	0.03	−0.00	0.03	0.04	0.04	0.08	−0.05	−0.02	0.01	0.11
LISn whole group	−0.12*	−0.16**	0.11	−0.08	−0.06	−0.20***	−0.09	−0.01	−0.02	−0.04

Note: * p < .05; ** p < .01; *** p < .001. WJPV = Woodcock-Johnson III Picture Vocabulary (Woodcock et al., 2001). EOW = Expressive One-Word Picture Vocabulary Test (Brownell, 2000). WJLW = Woodcock-Johnson III Letter-Word Identification (Woodcock et al., 2001). DIBELS FSF and PSF = Dynamic Indicators of Basic Early Literacy Skills First Sound Fluency and Phoneme Segmentation Fluency (Good, Gruba, & Kaminski, 2002). NIH Toolbox Flanker and DCCS = Flanker Inhibitory Control and Attention and Dimensional Change Card Sort (NIH Toolbox, 2015). WJAP = Woodcock-Johnson III Applied Problems (Woodcock et al., 2001). Teacher-rated self-regulation and social skills from factor analysis of scale scores from the Teacher-Child Relationship Scale (Hightower et al., 1986), Learning Behavior Scales (McDermott, 1999), and Student-Teacher Relationship Scale (Pianta, 2001).

6. Results

6.1. Descriptive Analyses

Tables 1 and 2 describe the sample, child outcomes, and ECE quality measures. As shown in Table 1, the sample is predominantly from low-income families and diverse, with about one-third of the participants being African American, Latinx, and White. Most of the Latinx children spoke Spanish at home (about 82%). Table 2 shows that children made statistically significant gains from fall to spring on language, academic, and EF assessments, and teacher ratings of social skills, but gains were small for language and EF skills. Using the CLASS, the pre-K classrooms were rated on average as having moderately high levels of emotional support and classroom organization and moderate levels of instructional support (see Table 3). Teachers were observed engaging in complex language exchanges with individual children only an average of 4% of the time, which included decontextualized language an average of 2% of the time, elaborations that maintained or extended child comments an average of 2% of the time, and conversations with three or more turns an average of 0.1% of the time. Children were observed in literacy activities for an average of 24% of the time and in math activities for 15% of the time. Their instructional time was split, on average, between whole group (36%), center/free choice (48%), and small group (8%), but individual classrooms tended to spend considerable time in ether whole group or centers (r = -.65).

Correlations among the ECE quality dimensions are reported in Tables A1 and A2 in Appendix A. Not surprisingly, large correlations were observed among the three CLASS domain scores, among the three teacher language variables, between literacy and sound-related activities and between classroom- and child-level measures of the same ECE experiences (ranging from .48 for sound-related activities to .77 for whole group). The only large correlation among ECE quality variables included in the same model involved the frequency of teacher complex language and of math activities (r = .47). These correlations support compositing the CLASS and language variables and including the child- and classroom-level LISn variables in separate models.

Table 4 shows the correlations between gains in child outcomes and ECE quality measures. The correlations tended to be quite modest and to reflect the anticipated associations between more frequent teacher complex talk and gains in vocabulary skills, more time in literacy activities and gains in language and literacy skills, and less time in whole group activities and gains in language, math, and inhibitory control.

6.2. Hierarchical Linear Model Analyses

The two sets of HLM analyses with random classroom intercepts examined children's fall to spring gains on each of the outcomes. Both models included the CLASS total score and the covariates. The LISn scores reflecting the proportion of time the child experienced complex teacher talk, content-specific instructional activities, and whole group settings were summarized at the classroom level in Model 1 and at the child level in Model 2.

Table 5 shows the variance components for the unconditional models. Results indicate between-classroom variability on all outcomes except the NIH Toolbox DCCS measure of cognitive flexibility. Accordingly, no further analyses were conducted to examine the DCCS.

Table 6 shows the results from Model 1 that included both classroom-level CLASS ratings and LISn behavior counts. Higher scores on the CLASS related to significantly smaller gains on the NIH Toolbox Flanker measure of inhibitory control (B = -0.17, SE = .06, p < .01). Children showed significantly larger gains in language skills on the EOW when observed spending more time in literacy-related activities (B = 0.19, SE = .08, p < .05) or less time in whole group activities (B = -0.19, SE = .08, p < .01). They made significantly larger gains in recognizing the first letters of words (B = 0.15, SE = .06, p < .05) and in phonemic skills (B = 0.13, SE = .06, p < .05) when observed spending more time spent in sound-related activities. Finally, they made smaller gain in math skills when they spent more time in whole group activities (B = -0.14, SE = .06, p < .05). Thus, this analysis using classroom-level ECE quality measures indicated that more time in instructional activities related to larger gains in language and literacy skills and less time in whole group related to larger gains in language and math skills.

Table 5
Hierarchical linear model results: random effects.

	WJPV	EOW	WJLW	DIBELS FSF	DIBELS PSF	WJAP	NIH Toolbox Flanker	NIH Toolbox DCCS	Self-regulation	Social skills
	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)	σ^2(se)
Unconditional model										
Class intercept	0.05***	0.23***	0.15***	0.06***	0.07***	0.13***	0.04***	0.00	0.09***	0.16***
	(0.04)	(0.08)	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)		(0.05)	(0.06)
Residual	0.95***	0.77***	0.86***	0.94***	0.93***	0.87***	0.96***	1.00***	0.91***	0.85***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.09)	(0.09)	(0.08)	(0.07)
Model 1										
Class intercept	0.00	0.19**	0.07	0.05	0.06	0.09	0.02		0.10*	0.13*
		(0.06)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)		(0.05)	(0.05)
Residual	0.83***	0.77***	0.86***	0.92***	0.93***	0.80***	0.93***		0.90***	0.84***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)		(0.08)	(0.07)
Model 2										
Class intercept	0.00	0.20**	0.07	0.05	0.07	0.09*	0.02		0.10*	0.13*
		(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)		(0.05)	(0.05)
Residual	0.80***	0.76***	0.85***	0.93***	0.92***	0.79***	0.93***		0.90***	0.85***
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)	(0.08)		(0.08)	(0.07)

Note: * p < .05; ** p < .01; *** p < .001. WJPV = Woodcock-Johnson III Picture Vocabulary (Woodcock et al., 2001). EOW = Expressive One-Word Picture Vocabulary Test (Brownell, 2000). WJLW = Woodcock-Johnson III Letter-Word Identification (Woodcock et al., 2001). DIBELS FSF and PSF = Dynamic Indicators of Basic Early Literacy Skills First Sound Fluency and Phoneme Segmentation Fluency (Good, Gruba, & Kaminski, 2002). NIH Toolbox Flanker and DCCS = Flanker Inhibitory Control and Attention and Dimensional Change Card Sort (NIH Toolbox, 2015). WJAP = Woodcock-Johnson III Applied Problems (Woodcock et al., 2001). Teacher-rated self-regulation and social skills from factor analysis of scale scores from the Teacher-Child Relationship Scale (Hightower et al., 1986), Learning Behavior Scales (McDermott, 1999), and Student-Teacher Relationship Scale (Pianta, 2001).

Table 6
HLM results: relating CLASS and Classroom-Level LISn quality indices to gains in child outcomes.

	WJPV	EOW	WJLW	DIBELS FSF	DIBELS PSF	WJAP	NIH Toolbox Flanker	Self-regulation	Social skills
	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)
Model 1									
Intercept	−1.23*	0.27	−0.66	0.75	0.15	−0.48	0.08	−0.11	0.12
	(0.58)	(0.68)	(0.65)	(0.65)	(0.66)	(0.64)	(0.65)	(0.70)	(0.68)
Age at entry to pre-K	0.03	−0.08	0.04	0.12*	0.11*	0.05	−0.09	0.00	−0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Time between fall & spring testing	1.88*	−0.41	0.99	−1.16	−0.23	0.71	−0.12	0.17	−0.18
	(0.89)	(1.03)	(0.99)	(0.99)	(1.01)	(0.97)	(0.99)	(1.07)	(1.04)
Sex (male = 1)	−0.07	−0.02	−0.07	−0.08	−0.03	−0.01	−0.08	−0.02	0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
African American	−0.03	−0.03	−0.06	−0.02	0.02	−0.06	0.10	−0.12	−0.06
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
DLL	0.36***	−0.15*	0.14*	0.02	−0.02	0.27***	−0.02	0.00	0.11
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Maternal education	−0.05	−0.07	−0.05	0.05	−0.03	−0.04	−0.05	−0.11	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
CLASS total	0.05	0.07	0.11	0.03	−0.03	0.08	−0.17**	0.09	0.08
	(0.05)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)
Classroom-Level LISn									
Complex talk	0.04	0.03	0.09	0.06	0.05	−0.06	−0.01	−0.03	0.09
	(0.06)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)
Literacy activities	0.02	0.19*	−0.03						
	(0.06)	(0.08)	(0.07)						
Sounds activities				0.15*	0.13*				
				(0.06)	(0.06)				
Math activities						0.04			
						(0.07)			
Whole group	−0.04	−0.19*	0.07	−0.09	−0.09	−0.14*	−0.09	−0.02	−0.08
	(0.05)	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)

Note: * p < .05; ** p < .01; *** p < .001. WJPV = WJIII Picture Vocabulary; EOW = Expressive One-Word Picture Vocabulary Test; WJLW = WJIII Letter-Word Identification; DIBELS FSF and PSF = DIBELS First Sound Fluency and Phoneme Segmentation Fluency; NIH Toolbox Flanker and DCCS = Flanker Inhibitory Control and Attention and Dimensional Change Card Sort; WJAP = WJIII Applied Problems. Teacher-rated self-regulation and social skills = composites from Teacher-Child Scale, Learning Behavior Scales and Student-Teacher Relationship Scale. CLASS = Classroom Assessment Systems, LISn = Language Interactions Snapshot. Under LISn, only the activity type relevant to the outcome was included (e.g. the math activities variable was examined for WJAP, a math outcome). Some outcomes did not have a relevant activity type (i.e. NIH Toolbox Flanker, self-regulation, and social skills).

Table 7 show the results from Model 2 that included classroom-level CLASS ratings and child-level behavior counts on the LISn. Some findings were consistent with those from Model 1 using the classroom-level LISn variables. Children in classrooms with higher CLASS scores showed smaller gains on the Flanker measure of inhibitory control (B = -0.17, SE as = .06, p < .01). Children made significantly larger gains in recognizing the first letters of words (B = 0.13, SE = 0.06, p < .05) and in phonemic skills (B = 0.13, SE = 0.06, p < .05) when they were individually observed spending more time in sound-related activities. More time in whole group activi-

Table 7
HLM results: relating CLASS and child-level LISn quality indices to gains in child outcomes.

	WJPV	EOW	WJLW	DIBELS FSF	DIBELS PSF	WJAP	NIH Toolbox Flanker	Self-regulation	Social skills
	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)	B(SE)
Model 2									
Intercept	−1.24*	0.20	−0.62	0.69	0.13	−0.57	0.02	−0.14	0.10
	(0.58)	(0.68)	(0.65)	(0.66)	(0.67)	(0.63)	(0.65)	(0.7)	(0.68)
Age at entry to pre-K	0.02	−0.09	0.03	0.10	0.09+	0.05	−0.10	0.00	−0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Time between fall & spring testing	1.89*	−0.30	0.93	−1.06	−0.20	0.86	−0.03	0.22	−0.15
	(0.88)	(1.03)	(0.99)	(1.00)	(1.02)	(0.97)	(0.99)	(1.07)	(1.04)
Sex (male = 1)	−0.05	0.00	−0.07	−0.06	−0.01	0.01	−0.08	−0.01	0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
African American	−0.02	−0.04	−0.06	−0.04	0.01	−0.05	0.10	−0.11	−0.06
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
DLL	0.37***	−0.14*	0.15*	0.03	−0.01	0.26***	−0.03	0.00	0.13
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Maternal education	−0.05	−0.07	−0.05	0.04	−0.03	−0.04	−0.05	−0.11	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
CLASS Total	0.04	0.06	0.12*	0.02	−0.05	0.06	−0.17**	0.09	0.09
	(0.05)	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Child-Level LISn									
Complex talk	0.13*	0.11	0.05	0.04	0.04	0.02	−0.04	0.00	0.06
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
Literacy activities	−0.01	0.06	0.05						
	(0.05)	(0.06)	(0.06)						
Sounds activities				0.13*	0.13*				
				(0.06)	(0.06)				
Math activities						0.05			
						(0.05)			
Whole group	−0.10*	−0.19**	0.09	−0.09	−0.05	−0.18**	−0.08	−0.06	−0.06
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)

Note: * p < .05; ** p < .01; *** p < .001. WJPV = WJIII Picture Vocabulary; EOW = Expressive One-Word Picture Vocabulary Test; WJLW = WJIII Letter-Word Identification; DIBELS FSF and PSF = DIBELS First Sound Fluency and Phoneme Segmentation Fluency; NIH Toolbox Flanker and DCCS = Flanker Inhibitory Control and Attention and Dimensional Change Card Sort; WJAP = WJIII Applied Problems. Teacher-rated self-regulation and social skills = composites from Teacher-Child Scale, Learning Behavior Scales and Student-Teacher Relationship Scale. CLASS = Classroom Assessment Systems, LISn = Language Interactions Snapshot. Under LISn, only the activity type relevant to the outcome was included (e.g. the math activities variable was examined for WJAP, a math outcome). Some outcomes did not have a relevant activity type (i.e. NIH Toolbox Flanker, self-regulation, and social skills).

ties was related to smaller gains in language skills according to the EOW measure of expressive skills (B = -0.19, SE = .07, p < .01) and the WJ measure of math skills (B = -0.18, SE = .06, p < .05).

Other associations emerged in Model 2 that were not detected in Model 1. Children in classrooms with higher CLASS scores showed larger gains on the WJ measure of basic reading skills (B = 0.12, SE = .06, p < .05). Children showed significantly larger gains on the WJPV measure of vocabulary when the target child experienced more frequent complex language exchanges with the teachers (B = 0.13, SE = .05, p < .05). More time spent in whole-group instruction was related to smaller gains on the WJ measure of receptive language (B = -0.10, SE = .056, p < .05). Thus, this analysis using child-level LISn quality measures indicated: higher CLASS scores were related to larger gains in literacy skills, but smaller gains in EF; more frequent complex language exchanges with the teacher were related to larger gains in receptive language; more time in instructional activities related to larger gains in language and literacy skills; and less time in whole group related to larger gains in language and math skills.

6.3. Sensitivity Analyses

Four sets of follow-up analyses were conducted. First, separate analyses included the CLASS, classroom-level LISn, and child-level LISn variables to see whether moderately large correlations between the CLASS and LISn teacher complex talk suppressed findings. Whereas most findings replicated, one finding emerged that was not statistically significant in either of the prior analyses. Results in Appendix Table A3 indicated that the classroom-level estimates of time in whole-group activities were related to smaller gains on the Flanker measure of inhibitory control (B = -0.13, SE = 0.06, p < .05) in analyses that excluded the CLASS. It should be noted that this change reflected small changes in effect sizes that moved the estimates from p-values just above p < .05 to just below p < .05.

Second, individual CLASS domain scores replaced the CLASS total scores to ensure that analyses were not ignoring findings for specific domains. Given their high intercorrelations (r's > .69), it is not surprising that the same findings were obtained with the individual CLASS domains as with the CLASS total score.

The third set of set of follow-up analyses examined indices of the teacher's linguistic responsivity and language complexity when the composite variable was significantly related to gains in child outcomes. Follow-up analyses indicated that more frequent elaborations by the teacher toward the target child were related to larger gains on both the WJ measure of expressive language, (B = 0.14, SE = .05, p < .01) and EOW measure of receptive language (B = 0.14, SE = .05, p < .01). In contrast, more frequent use of decontextualized language was not related to larger gains on both the WJ measure of expressive language, (B = 0.07, SE = .05, p > .05) or EOW measure of receptive language (B = 0.06, SE = .06, p > .05).

The final set of follow-up analyses address potential distributional issues with the change scores or ECE quality scores. The distributions were relatively normal for the fall to spring outcome change scores, CLASS scores, and both the child- and classroom-level measures of time spent in literacy and math activities and whole-group settings. However, the distribution of child-level teacher complex language and the amount of time in sound-related activities were triangular. A square root transformation and a binary recoding (whether the child ever experienced or not) were applied to these variables. Both sets of analyses indicated that the

associations between time spent in sound-related activities and gains on both DIBELS measures of decoding skills continued to be statistically significant. The association between time exposed to teacher complex talk and EOW was statistically significant, and the association between teacher complex talk with WJ PV remained positive but was no longer statistically significant.

7. Discussion

This study investigated whether ECE quality models should be more differentiated to include both the types of ECE experiences that may be foundational for developing multiple skills and the ECE experiences that promote the acquisition of specific skills. Overall, results indicated that selected quality dimensions were differentially associated with gains in cognitive and academic skills in anticipated ways. Furthermore, the quality dimensions that significantly related to child outcomes were measured using behavior counts of individual children, not classroom-level ratings of teacher-child interactions. If replicated in future studies, these findings suggest classroom-level behavior counts should be considered to measure ECE quality under most circumstances. As such, this study makes several psychometric contributions to the study of ECE quality by suggesting: (1) the "one-size-fits-all" approach to studying ECE quality should be expanded to examine different ECE indices for different child outcomes; (2) measures of the experiences of individual children should be considered in addition to ratings of the entire classroom when assessing ECE quality; and (3) perhaps child-specific counts will be most useful in predicting individual differences in child outcomes when they measure experiences that vary markedly among children.

7.1. Differentiated ECE Quality Models for Specific Child Outcomes

Results indicated that different child skills appeared to be promoted by different ECE quality dimensions. Evidence did not support the hypotheses that either higher quality teacher-child interactions measured globally or less time in whole group settings would predict gains in all domains, although time in whole group settings was related to multiple outcomes in the anticipated direction. In contrast, many of the anticipated associations between specific ECE quality dimensions were observed for some of the language, academic, and EF outcomes as discussed below for each of the child outcomes.

7.1.1. Language skills

This study indicated that children showed gains in language skills when they experienced more frequent complex language exchanges with their teacher, spent more time in literacy activities, and spent less time in whole group activities. These findings largely replicate prior studies and may have important implications for preschool programs, especially pre-kindergarten programs.

Results indicate that the acquisition of receptive language skills in ECE may be promoted by frequent complex language exchanges between the teacher and the target child. This is aligned with the results of experimental studies demonstrating that linguistic responsivity and complexity in teacher-child language exchanges improves children's language skills (Cabell et al., 2015; Hindman & Wasik, 2012) and of a recent study that suggested teacher responsiveness was more important than the complexity or the overall quality of teacher-child interactions in increasing vocabulary skills (Justice et al., 2018). Despite differences in how language interactions were measured, both studies suggest that linguistic responsiveness is related slightly more strongly to vocabulary growth than is language complexity or the overall quality of teacher-child interactions. Nevertheless, most preschoolers infrequently experience complex teacher-child language exchanges according to this study and prior studies (Cabell et al., 2011; Sawyer et al., 2018).

In addition, children showed larger gains in expressive language skills when they spent more time in literacy activities and less time in whole group activities. It is not surprising that spending more time in literacy activities, aggregated to the classroom level, related to gains in language given the documented importance of literacy activities, such as shared book reading, for oral language development in early childhood (Callaghan & Madelaine, 2012). Nor is it surprising that spending less time in whole group activities, measured both at the child- and classroom-level, predicted larger gains in language skills. Whole group activities likely provide fewer opportunities for high quality language exchanges between individual children and teachers, as was demonstrated in an experimental study of a preschool tiered intervention (Buysse et al., 2016).

7.1.2. Academic skills

Both instructional time and quality of teacher-child interactions related to gains in reading skills. Gains in decoding skills were larger when children spent time in decoding activities as in prior studies (Bratsch-Hines et al., 2019; Burchinal et al., 2016; Connor et al., 2006; Pianta et al., 2020). As in prior work (Burchinal et al., 2010, 2016; Howes et al., 2008), gains in literacy skills in this study were related to higher CLASS scores in at least one analysis.

Consistent with prior studies, children showed significantly larger gains in math skills when they spent less time in whole group activities and non-significantly larger gains when CLASS scores were higher. Intervention studies indicated that mathematical skills are promoted when children experience instruction involving a combination of settings (Clements & Sarama, 2007). Previous studies reported associations between the quality of teacher-child interactions and math skills as measured by the WJ Applied Problems scale (Hong et al., 2019; Mashburn et al., 2008; Pianta et al., 2020) that were similar or smaller in magnitude to those observed in this study, but were significant due to substantially larger samples.

7.1.3. EF and social skills

This study contributes little to developing an ECE model for promoting EF or social skills. There was no between classroom variability on assessed cognitive flexibility and modest classroom variability on assessed inhibitory control. Gains in inhibitory control were negatively related to CLASS scores in all analyses and time spent in whole group activities in the follow-up analyses. Whereas it is not surprising that more time in whole group activities would result in smaller gains in EF skills in this study in a similar manner as reported in other studies, it is difficult to explain why higher quality teacher-child interactions relate to smaller gains in inhibitory control. As in this study, Pianta and colleagues (2020) reported smaller gains in EF skills when more time was spent in academic activities, much of which occurred during whole group settings. Unlike this study, however, they observed larger gains in EF when CLASS scores were higher (Pianta et al., 2020).

None of the quality measures related to gains in teacher ratings of social skills or self-regulation. Power may have been an issue. Prior studies demonstrate associations between CLASS Emotional Support and social skills (Burchinal et al., 2010; Hong et al., 2019), and reported similar effect sizes as reported in this study. This study also may not have observed the most relevant teacher and child behaviors. For example, self-regulation in preschool and even in first grade was related to classroom-aggregated ratings of the tone of the interactions between teachers and individual children and cooperation in peer interactions (Fuhs et al., 2013).

7.1.4. Implications

This study may contribute largely to focusing attention on a more differentiated ECE quality model especially for promoting language skills in the preschool classroom. This may be particularly important because language skills are foundational to subsequent school success (Lonigan & Shanahan, 2009), and some evidence suggests that school-entry language, EF, and social skills may be more important than school-entry academic skills in acquiring academic and social skills during the elementary school (Burchinal et al., 2020; Pace, Alper, Burchinal, Golinkoff, & Hirsh-Pasek, 2019). Preschool programs, especially pre-kindergarten programs, however, are more successful in promoting basic reading and math skills than language skills (Bailey, Duncan, Odgers, & Yu, 2017; Burchinal et al., 2015). This preschool advantage in reading and math skills fades and can become negative during the elementary years (Lipsey, Farran, & Durkin, 2018; Phillips et al., 2017). This may explain why long-term achievement gaps have not diminished (Reardon, Valentino, & Shores, 2012).

7.2. Measuring ECE Quality

Findings from multiple ECE studies, including this one, raise questions about relying solely on general classroom-level ratings of teacher-child interactions to describe ECE quality. In this study, classroom ratings of the quality of teacher-child interactions were not consistently significantly related to any child outcome in the hypothesized direction, whereas behavior counts of ECE experiences related to gains in language and academic skills as hypothesized. Prior studies also suggested that gains in more child outcomes were related to behavior counts or ratings than to classroom-level ratings of the quality of teacher-child interactions (Justice et al., 2018; Pianta et al., 2020).

Whether behavior counts of individual children should be considered as child-level or aggregated classroom-level measures probably depends on the degree to which those experiences vary within a classroom and how the ECE quality measure is being used. In this study, remarkably similar associations emerged between child outcomes and the classroom- and child-level measures of instructional time and use of whole-group settings. This is likely because different children's experiences were similar during the observational period. These are activities that teachers probably organize similarly for most to all children in the classroom, even if not all children are observed doing them at the same time. In that case, observing even a subset of the classroom probably provides a good index of experiences for all children in the classroom.

In contrast, other ECE experiences may vary dramatically among children. How and how much teachers talk with individual children and the level of affect they show during interactions with a given child probably vary depending on child and teacher characteristics. The degree to which teachers engage children in higher level conversations markedly varied among selected children in this study and at least one prior study (Sawyer et al., 2018). This might explain why child-level measures of teacher complex talk based on 20 minutes of observation related to gains in children's vocabulary skills whereas the classroom-level measure based on 2 hours of observation did not. While very few children experienced frequent complex language exchanges with teachers, about two-thirds of the children experienced them at least once during our observations. These individual differences in the degree to which experiences teacher was significantly related to individual differences in language skills, suggesting that child-level observations may be useful in predicting gains in outcomes for the observed children when there are substantial individual differences in those behaviors. It is likely, nevertheless, that a classroom-level measure of complex teacher talk is needed for other purposes such as monitoring and describing classroom quality due to the much higher reliability of the classroom-level measures.

Unfortunately, we did not design this study to focus on within-child assessments of quality, and thus we acknowledge it is surprising that 20 minutes of observation of individual children significantly related to gains in language, academic, and EF skills as shown in Model 2 in Table 7. The short observation period, however, clearly limits the precision of measurement. Accordingly, it is possible that short observation periods attenuated observed associations but should not account for them. If so, more observations, collected across either multiple days or more settings would probably provide even better prediction of learning and development in preschool classrooms, but further investigation of child-level quality assessment is needed.

7.2.1. Implications

Measures of ECE quality influence how "high-quality" ECE programs are conceived and funded by policymakers, implemented by administrators and practitioners, and chosen by families. Including the specific ECE experiences measured by behavioral counts in definitions of high quality ECE has implications for monitoring and professional development. Classroom-level ratings of teacher-child interactions currently dominate how quality is defined, programs are evaluated, and professional development is designed to improve quality (Burchinal et al., 2015). The most widely-used measure of teacher-child interaction quality, the CLASS, is used to evaluate the quality of Head Start programs (e.g., Office of Head Start, 2011) and state and local pre-K programs (Peisner-Feinberg et al., 2014; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013) and, along with the ECERS, serves as the basis for preschool ECE quality improvement efforts (Build Initiative & Child Trends, 2014). Both the CLASS and ECERS have been used widely in professional development, with considerable evidence that such professional development improved ratings on those scales but little rigorous evidence to indicate that the professional development improved child outcomes (for a review, see Burchinal et al., 2015).

In contrast, some evidence suggests professional development based on behavior counts may improve both quality and child outcomes. Professional development based on the Snapshot portion of the LISn was implemented in several districts to improve and align pre-kindergarten to third-grade instruction, providing non-experimental evidence that the professional development improved pre-kindergarten quality and child outcomes (Manship, Farber, Smith, & Drummond, 2016). Similarly, a non-experimental study indicated that professional development on selected TOP and COP measures was associated with both improved ECE quality and child outcomes (Farran, Meador, Christopher, Nesbitt, & Bilbrey, 2017).

7.3. Limitations

First, this study cannot demonstrate causal links between the examined ECE quality dimensions and children's learning and development. Using gain scores accounts, to some degree, for pre-existing differences in children's skills at entry to pre-K, but other factors could still account for the observed associations (NICHD Early Child Care Research Network & Duncan, 2003).

Second, the findings tend to be modest and somewhat mixed. For example, the child-level teacher complex language significantly related to gains in one of two language outcomes, time in content-specific instruction related to one of two language and two of three literacy outcomes but not to math outcomes, and time in whole group settings related to both language outcomes, the math outcome, and the EF outcome in at least one of the analyses. It should be noted that associations between complex teacher-child language and both language outcomes were similar, but only one was sta-

tistically significant, likely due to limited power in this study of 63 classrooms. Nevertheless, stronger, more consistent findings would have been more convincing.

Third, no child-level ratings of teacher-child interactions were collected, so it was only possible to contrast classroom-level ratings and both classroom- and child-level behavior counts. The contrast between the CLASS and LISn reflects both differences in methodology and the unit of observation. Different conclusions might have been drawn if child-level ratings of teacher-child sensitivity or ECE quality ratings or behavior count systems had been collected.

Fourth, the sample is limited to children attending state pre-K classrooms in rural areas in the Southeast region of the United States. The NC Pre-K program has higher training and education requirements and lower ratios and group size regulations than private community-based preschools, so it is not clear whether these findings would generalize to other types of programs. Whether results generalize beyond rural or pre-K settings cannot be determined in this study, but analyses of nationally representative data did not reveal differences in associations between ECE quality and child outcomes in rural areas and the country in general (McCoy, Morris, Connors, Gomez, & Yoshikawa, 2016) or between pre-K and other types of center-based programs (Burchinal et al., 2015). Nonetheless, caution should be used when generalizing these findings to children in other types of programs or from other regions of the country.

Fifth, limited between-classroom variability in one of the EF outcomes meant this study was hampered in relating classroom-level ECE experiences to cognitive flexibility. Limiting classroom observations to mornings when instruction was most likely to occur may have excluded observations of classroom activities that promote EF or social skills.

Sixth, observing only four to six children per classroom limited distinctions between the classroom- and child-level LISn variables since the classroom-level variables were created from the child-level variables. Thus, caution is warranted when drawing conclusions regarding whether the classroom- and child-level LISn variables yielded similar findings.

Seventh, it is also likely that teacher ratings of social skills may be less sensitive than direct assessments of cognitive and language skills because they reflect both observations of the child and the teacher's views about what is appropriate social behavior. It is possible such biases limited the ability of this study to relate ECE experiences to children's social skills. In addition, this study observed classrooms when teachers indicated they were focusing on instruction. It is possible that the ECE quality might have related to children's social skills if observations included activities such as meals, nap, and outdoor time.

Finally, it is not clear to what extent teachers or children elicited the observed interactions among them. It is highly likely that more engaging children elicited more frequent complex talk, and it is possible those children show larger gains regardless of what happens in their classrooms.

8. Conclusion

In summary, findings from this study indicated that different child skills appear to be promoted by different ECE quality dimensions as anticipated, with gains in vocabulary related to the frequency of complex language exchanges between that child and the teacher, gains in language and literacy skills related to time spent on literacy instruction, and gains in language, math, and perhaps inhibitory control related negatively to time spent in whole group instruction. Results suggest that behavior counts of individual children's experiences provide good measurement of these experiences when they are aggregated to the classroom level for most uses of ECE quality assessments. Exploratory analyses suggest that aspects of ECE quality that likely vary markedly among classmates may require individual-level assessments of those experiences when the goal is predicting individual differences in child outcomes.

Acknowledgements

We are grateful to each of the NC pre-kindergarten teachers and the children and their parents who participated in this study. We thank Sarah Wackerhagen for assistance in preparing this paper and the entire UNC Early Learning Network team for their many contributions. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305N160022 to the University of North Carolina at Chapel Hill. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Appendix A.

Table A1
Correlations among classroom assessments of ECE quality.

	Class ES	Class CO	Class IS	Teacher complex talk	Decontextualized language	Linguistic responsivity	Literacy activities	Sounds activities	Math activities	Whole group
Classroom-level measures										
CLASS Total	0.93***	0.87***	0.91***	0.32***	0.33*	0.23	0.08	−0.07	0.15	0.19
Emotional support (ES)		0.70***	0.80***	0.25***	0.23	0.21	−0.01	−0.14	0.19*	0.13*
Classroom organization (CO)			0.66***	0.30***	0.30*	0.22	0.06	−0.06	0.02	0.26***
Instructional support (IS)				0.31***	0.34**	0.21	0.18	0.00	0.18	0.13
LISn teacher complex talk					0.90***	0.90***	0.31*	0.23*	0.47***	0.07
Decontextualized language						0.65***	0.38**	0.24	0.45***	0.15
Linguistic responsivity							0.12	0.13	0.02	−0.05
LISn literacy activities								0.54***	0.03	0.25*
LISn sounds activities									0.12	0.04
LISn math activities										0.05
LISn whole group										

Note: CLASS = Classroom Assessment Scoring System (Pianta et al., 2008). LISn = Language Interaction Snapshot (Sprachman, Caspe, & Atkins-Burnett, 2009).

Table A2
Correlations among classroom- and child-level assessments of ECE quality.

	Child-Level Measures (n = 339)						
	Teacher complex language	Decontext-ualized language	Linguistic responsivity	Literacy activities	Sounds activities	Math activities	Whole group
Classroom-level measures							
CLASS Total	0.19**	0.20**	0.13*	0.05	−0.04	0.07	0.18***
Emotional Support (ES)	0.14*	0.14*	0.10	0.01	−0.07	0.09	0.13*
Classroom Organization (CO)	0.19***	0.19***	0.12*	0.05	−0.02	−0.01	0.24***
Instructional Support (IS)	0.18**	0.22***	0.13*	0.09	−0.01	0.12*	0.12*
LISn teacher complex talk	0.66***	0.60***	0.55***	0.23***	0.13*	0.28***	0.09
Decontextualized language	0.61***	0.65***	0.42***	0.23***	0.11*	0.29***	0.12*
Linguistic responsivity	0.60***	0.44***	0.62***	0.24***	0.09	0.29**	0.12*
LISn literacy activities	0.26***	0.27***	0.14*	0.54***	0.26***	0.22**	0.12*
LISn sounds activities	0.17**	0.16**	0.12*	0.29***	0.48***	0.08	0.04
LISn math activities	0.29***	0.30***	0.23***	0.05	0.06	0.64***	0.04
LISn whole group	0.07	0.10	0.01	0.13*	−0.02	0.04	0.77***
Child-level measures							
LISn teacher complex talk		0.85***	0.83***	0.20***	0.13*	0.21***	0.09
Decontextualized language			0.46***	0.19***	0.12*	0.25***	0.13*
Linguistic responsivity				0.13*	0.08	0.11*	0.02
LISn literacy activities					0.44***	0.04	0.17**
LISn sounds activities						0.04	0.06
LISn math activities							0.07
LISn whole group							

Note: CLASS = Classroom Assessment Scoring System (Pianta et al., 2008). LISn = Language Interaction Snapshot (Sprachman, Caspe, & Atkins-Burnett, 2010).

Table A3
HLM results relating ECE quality to gains in child outcomes in separate models for the CLASS, classroom-level LISn, and child-level LISn.

	WJPV	EOW	WJLW	DIBELS FSF	DIBELS PSF	WJAP	NIH Toolbox Flanker	Self-regulation	Social skills
Model 1 CLASS									
Intercept	−1.23*	0.33	−0.70 (0.65)	0.66	0.08	−0.39	0.11	−0.10	0.12
	(0.58)	(0.69)		(0.67)	(0.67)	(0.64)	(0.64)	(0.70)	(0.68)
Age at entry to	0.03	−0.09+	0.03	0.10*	0.10+	0.05	−0.10+	0.00	−0.04
pre-K	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Time between	1.88*	−0.5	1.06	−1.02	−0.13	0.57	−0.17	0.16	−0.18
F & S tests	(0.88)	(1.05)	(0.98)	(1.02)	(1.02)	(0.98)	(0.98)	(1.06)	(1.04)
Sex (male = 1)	−0.07	−0.02 (0.05)	−0.07	−0.08	−0.02	−0.01	−0.08	−0.02	0.01
	(0.05)		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
African	−0.03	−0.05	−0.05	−0.04	0.00	−0.07	0.09	−0.12+	−0.07
American	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
DLL	0.37***	−0.15*	0.14*	0.03	−0.01	0.27***	−0.02	0.00	0.12+
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Maternal	−0.05	−0.07	−0.05	0.04	−0.03	−0.04	−0.05	−0.10+	0.02
Education	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
CLASS total	0.05	0.06	0.15*	0.01	−0.05	0.03	−0.19***	0.08	0.09
	(0.05)	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Model 2 Classroom-level LISn[a]									
T complex talk	0.05	0.05	0.12	0.07	0.04	−0.04	−0.05	0.00	0.12
	(0.06)	(0.08)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Literacy	0.01	0.19*	−0.04						
activities	(0.06)	(0.08)	(0.07)						
Sounds				0.15*	0.14*				
activities				(0.06)	(0.06)				
Math activities						0.04			
						(0.07)			
Whole group	−0.03	−0.18*	0.09	−0.09	−0.10	−0.12	−0.13*	0.00	−0.07
	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Model 3 Child-level LISn[a]									
T complex talk	0.14**	0.11	0.07	0.04	0.04	0.03	−0.07	0.02	0.07
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)
Literacy	−0.01	0.06	0.05						
activities	(0.05)	(0.06)	(0.06)						
Sounds				0.12*	0.13*				
activities				(0.06)	(0.06)				
Math activities						0.05			
						(0.05)			
Whole group	−0.10	−0.18**	0.10	−0.08	−0.06	−0.17**	−0.11	0.05	−0.05
	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)

Note: * p < .05; ** p < .01; *** p < .001. WJPV = Woodcock-Johnson III Picture Vocabulary (Woodcock et al., 2001). EOW = Expressive One-Word Picture Vocabulary Test (Brownell, 2000). WJLW = Woodcock-Johnson III Letter-Word Identification (Woodcock et al., 2001). DIBELS FSF and PSF = Dynamic Indicators of Basic Early Literacy Skills First Sound Fluency and Phoneme Segmentation Fluency (Good, Gruba, & Kaminski, 2002). NIH Toolbox Flanker and DCCS = Flanker Inhibitory Control and Attention and Dimensional Change Card Sort (NIH Toolbox, 2015). WJAP = Woodcock-Johnson III Applied Problems (Woodcock et al., 2001). Teacher-rated self-regulation and social skills from factor analysis of scale scores from the Teacher-Child Relationship Scale (Hightower et al., 1986), Learning Behavior Scales (McDermott, 1999), and Student-Teacher Relationship Scale (Pianta, 2001).

[a] Models 2 and 3 include the same covariates, but their coefficients are not list listed because they are so similar across mode.

References

Abt Associates, Inc. (2007). *Evaluation of child care subsidy strategies: Findings from Project Upgrade in Miami-Dade county*. U.S. Department of Health and Human Services, Administration for Children & Families. https://www.acf.hhs.gov/sites/default/files/opre/upgrade_miami_dade.pdf

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. http://dx.doi.org/10.1080/01621459.1996.10476902

Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2010). Language Interaction Snapshot with End of Visit Ratings (LISn+ EVR). *Mathematica Policy Research.*

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research Educational Effectiveness*, *10*(1), 7–39. http://dx.doi.org/10.1080/19345747.2016.1232459

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., et al. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, *79*(6), 1802–1817. http://dx.doi.org/10.1111/j.1467-8624.2008.01227.x

Bilbrey, C., Vorhaus, E., Farran, D. C., & Shufelt, S. (2010). *Teacher observation in preschool: Tools of the Mind adaptation*. Peabody Research Institute.

Bratsch-Hines, M. E., Burchinal, M., Peisner-Feinberg, E., & Franco, X. (2019). Frequency of instructional practices in rural prekindergarten classrooms and associations with child language and literacy skills. *Early Childhood Research Quarterly*, *47*, 74–88. http://dx.doi.org/10.1016/j.ecresq.2018.10.001

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

Brownell, R. (Ed.). (2000). *Expressive One-Word Picture Vocabulary Test: Manual.* Academic Therapy Publications.

Build Initiative & Child Trends. (2014). *A catalog and comparison of Quality Rating and Improvement Systems (QRIS) [Data System]*. http://qriscompendium.org

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, *12*(1), 3–9. http://dx.doi.org/10.1111/cdep.12260

Burchinal, M., Foster, T. J., Bezdek, K. G., Bratsch-Hines, M., Blair, C., Vernon-Feagans, L., et al. (2020). School-entry skills predicting school-age academic and social–emotional trajectories. *Early Childhood Research Quarterly*, *51*, 67–80. http://dx.doi.org/10.1016/j.ecresq.2019.08.004

Burchinal, M., Magnuson, K., Powell, D., & Hong, S. S. (2015). Early child care and education and child development. In M. Bornstein, R. M. Lerner, & T. Leventhal (Eds.), *Handbook of child psychology and developmental science: Vol. 4. Ecological settings and process* (7th ed., pp. 223–267). Wiley.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176. http://dx.doi.org/10.1016/j.ecresq.2009.10.004

Burchinal, M., Zaslow, M., & Tarullo, L. (2016). Quality thresholds, features and dosage in early care and education: Secondary data analyses of child outcomes. *Monographs of the Society for Research in Child Development*, *81*, 1–128. http://dx.doi.org/10.1111/mono.12240

Buysse, V., Peisner-Feinberg, E., Soukakou, E., Fettig, A., Schaaf, J., & Burchinal, M. (2016). Using recognition & response (R&R) to improve children's language and literacy skills: Findings from two studies. *Early Childhood Research Quarterly*, *36*, 11–20. http://dx.doi.org/10.1016/j.ecresq.2015.11.005

Cabell, S. Q., DeCoster, J., LoCasale-Crouch, J., Hamre, B. K., & Pianta, R. C. (2013). Variation in the effectiveness of instructional interactions across preschool classroom settings and learning activities. *Early Childhood Research Quarterly*, *28*(4), 820–830. http://dx.doi.org/10.1016/j.ecresq.2013.07.007

Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J., & Forston, L. D. (2015). Teacher-child conversations in preschool classrooms: Contributions to children's vocabulary development. *Early Childhood Research Quarterly*, *30*, 80–92. http://dx.doi.org/10.1016/j.ecresq.2014.09.004

Cabell, S. Q., Justice, L. M., Piasta, S. B., Curenton, S. M., Wiggins, A., Turnbull, K. P., et al. (2011). The impact of teacher responsivity education on preschoolers' language and literacy skills. *American Journal of Speech-Language Pathology*, *20*(4), 315–330. http://dx.doi.org/10.1044/1058-0360(2011/10-0104)

Callaghan, G., & Madelaine, A. (2012). Leveling the playing field for kindergarten entry: Research implications for preschool early literacy instruction. *Australasian Journal of Early Childhood*, *37*(1), 13–23. http://dx.doi.org/10.1177/183693911203700103

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., et al. (2010). Children's classroom engagement and gains in academic and social-emotional outcomes across pre-kindergarten. *Child Development*, *81*(5), 1534–1544. http://dx.doi.org/10.1111/j.1467-8624.2010.01490.x

Clements, D. H., & Sarama, J. (2007). Early childhood mathematics learning. In F. K. Lester Jr (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 461–555). Information Age Publishing.

Connor, C. M., Morrison, F. J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology*, *98*(4), 665–689. http://dx.doi.org/10.1037/0022-0663.98.4.665

Cummings, K. D., Kaminski, R. A., Good III, R. H., & O'Neil, M. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of first sound fluency. *Assessment for Effective Intervention*, *36*(2), 94–106. http://dx.doi.org/10.1177/1534508410392209

Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009). Teacher–child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology*, *101*(4), 912–925. http://dx.doi.org/10.1037/a0016647

Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher–child interactions in the classroom: Toward a theory of within- and cross-domain links to children's developmental outcomes. *Early Education and Development*, *21*(5), 699–723. http://dx.doi.org/10.1080/10409289.2010.497453

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. M. C., Crawford, G. M., et al. (2010). How do pre-kindergartners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *25*(2), 177–193. http://dx.doi.org/10.1016/j.ecresq.2009.10.003

Farran, D. C., Meador, D., Christopher, C., Nesbitt, K., & Bilbrey, L. (2017). Data-driven improvement in prekindergarten classrooms: Report from a partnership in an urban district. *Child Development*, *88*(5), 1466–1479. http://dx.doi.org/10.1111/cdev.12906

Farran, D. C., & Son-Yarbrough, W. (2001). Title I funded preschools as a developmental context for children's play and verbal behaviors. *Early Childhood Research Quarterly*, *16*(2), 245–262. http://dx.doi.org/10.1016/S0885-2006(01)00100-00104

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2013). Preschool classroom processes as predictors of children's cognitive self-regulation skills development. *School Psychology Quarterly*, *28*(4), 347–359. http://dx.doi.org/10.1037/spq0000031

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2011). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly*, *27*(2), 198–209. http://dx.doi.org/10.1016/j.ecresq.2011.10.001

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH Toolbox for assessment of neurological and behavioral function. *Neurology*, *80*(11 Supplement 3), S2–S6. http://dx.doi.org/10.1177/1073191118766396

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Institute for the Development of Educational Achievement.

Hamre, B. K. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives*, *8*(4), 223–230. http://dx.doi.org/10.1111/cdep.12090

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale, revised (ECERS-R).* Teachers College Press.

Harms, T., Clifford, R. M., & Cryer, D. (2014). *Early Childhood Environment Rating Scale, third edition (ECERS-3).* Teachers College Press.

Heckman, J. J. (2011). The economics of inequality: The value of early childhood education. *American Educator*, *35*, 31–47. https://eric.ed.gov/?id=EJ920516

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., et al. (1986). The Teacher-Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review*, *15*(3), 393–409. http://dx.doi.org/10.1037/t05347-000

Hindman, A. H., & Wasik, B. A. (2012). Unpacking an effective language and literacy coaching intervention in Head Start: Following teachers' learning over two years of training. *The Elementary School Journal*, *113*(1), 131–154. http://dx.doi.org/10.1086/666389

Hong, S. L., Sabol, T. J., Burchinal, M. R., Tarullo, L., Zaslow, M., & Peisner-Feinberg, E. S. (2019). ECE quality indicators and child outcomes: Analyses of six large child care studies. *Early Childhood Research Quarterly*, *49*, 202–217. http://dx.doi.org/10.1016/j.ecresq.2019.06.009

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten program. *Early Childhood Research Quarterly*, *23*(1), 429–430. http://dx.doi.org/10.1016/j.ecresq.2008.08.001

Jones, S. M., Bouffard, S. M., & Weissbourd, R. (2013). Educators' social and emotional skills vital to learning. *The Phi Delta Kappan*, *94*(8), 62–65. http://dx.doi.org/10.1177/003172171309400815

Justice, L. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Childhood Research Quarterly*, *42*, 79–92. http://dx.doi.org/10.1016/j.ecresq.2017.09.003

Kaminski, R. A., & Good, R. H. I. I. I. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, *25*(2), 215–227. http://dx.doi.org/10.1037/h0088938

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. http://dx.doi.org/10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. http://dx.doi.org/10.2307/2529310

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee prekindergarten program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, *45*, 155–176. http://dx.doi.org/10.1016/j.ecresq.2018.03.005

Lonigan, C. J., Farver, J. M., Phillips, B. M., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing*, *24*(3), 305–337. http://dx.doi.org/10.1007/s11145-009-9214-6

Lonigan, C. J., & Shanahan, T. (2009). *Developing early literacy: Report of the National Early Literacy Panel. Executive summary. A Scientific synthesis of early literacy development and implications for intervention*. National Institute for Literacy. https://files.eric.ed.gov/fulltext/ED508381.pdf

Manship, K., Farber, J., Smith, C., & Drummond, K. (2016). *Case studies of schools implementing early elementary strategies: Preschool through third grade alignment and differentiated instruction*. Office of Planning, Evaluation and Policy Development, US Department of Education. https://www2.ed.gov/rschstat/eval/implementing-early-strategies/report.pdf

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

McCoy, D. C., Morris, P. A., Connors, M. C., Gomez, C. J., & Yoshikawa, H. (2016). Differential effectiveness of Head Start in urban and rural communities. *Journal of Applied Developmental Psychology*, *43*, 29–42. http://dx.doi.org/10.1016/j.appdev.2015.12.007

McDermott, P. A., Green, L. F., Francis, J. M., & Stott, D. H. (1999). Learning Behaviors Scale. *Edumetric and Clinical Science*.

NICHD Early Child Care Research Network. (2006). Child-care effect size for the NICHD Study of Early Child Care and Youth Development. *American Psychologist*, *61*(2), 99–116. http://dx.doi.org/10.1037/0003-066X.61.2.99

NICHD Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*(5), 1456–1475. http://dx.doi.org/10.1111/1467-8624.00617

Office of Head Start. (2011). *Report to congress on the final Head Start program designation renewal system*. U.S. Department of Health and Human Services. https://eclkc.ohs.acf.hhs.gov/report/report-congress-final-head-start-program-designation-renewal-system

Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*, *46*, 112–125. http://dx.doi.org/10.1016/j.ecresq.2018.04.001

Peisner-Feinberg, E. S., LaForett, D. R., Schaaf, J. M., Hildebrandt, L. M., Sideris, J., & Pan, Y. (2014). *Children's outcomes and program quality in the North Carolina Pre-Kindergarten Program: 2012-2013 statewide evaluation*. The University of North Carolina, FPG Child Development Institute. Pre-K Eval 2012-2013 Report.pdf. https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/NC

Pentimonti, J. M., Justice, L. M., Yeomans-Maldonado, G., McGinty, A. S., Slocum, L., & O'Connell, A. (2017). Teachers' use of high- and low-support scaffolding strategies to differentiate language instruction in high-risk/economically disadvantaged settings. *Journal of Early Intervention*, *39*(2), 125–146. http://dx.doi.org/10.1177/1053815117700865

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., Duncan, G. J., Dynarsky, M., Magnuson, K., & Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. A consensus statement. In K. Dodge (Ed.), *Issues in pre-kindergarten programs and policy* https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf

Pianta, R. C. (2001). *Student-Teacher Relationship Scale: Professional manual*. Psychological Assessment Resources.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System manual*. Pre-K. Brookes.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review*, *33*(3), 444–458. http://dx.doi.org/10.1080/02796015.2004.12086261

Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., et al. (2020). Children's school readiness skills across the pre-k year: Associations with teacher-student interactions, teacher practices, and exposure to academic content. *Journal of Applied Developmental Psychology*, *66*, 1–10. http://dx.doi.org/10.1016/j.appdev.2019.101084

Powell, D. R., Burchinal, M., File, N., & Kontos, N. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool. *Early Childhood Research Quarterly*, *23*(1), 108–123. http://dx.doi.org/10.1016/j.ecresq.2008.05.002

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, *102*(2), 299–312. http://dx.doi.org/10.1037/a0017763

Raver, C. C. (2003). Does work pay psychologically as well as economically? The role of employment in predicting depressive symptoms and parenting among low-income families. *Child Development*, *74*(6), 1720–1736. http://dx.doi.org/10.1046/j.1467-8624.2003.00634.x

Raver, C. C., & Blair, C. (2016). Neuroscientific insights: Attention, working memory, and inhibitory control. *The Future of Children*, *26*(2), 95–118. http://dx.doi.org/10.1353/foc.2016.0014

Reardon, S. F., Valentino, R. A., & Shores, K. A. (2012). Patterns of literacy among U.S. students. *The Future of Children*, *22*(2), 17–37. http://dx.doi.org/10.1353/foc.2012.0015

Rimm-Kaufman, S. E., La Paro, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *The Elementary School Journal*, *105*(4), 377–394. http://dx.doi.org/10.1086/429948

Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emerging Academics Snapshot*. University of California at Los Angeles (unpublished measure).

Sawyer, B., Atkins-Burnett, S., Sandilos, L., Scheffner Hammer, C., Lopez, L., & Blair, C. (2018). Variations in classroom language environments of preschool children who are low income and linguistically diverse. *Early Education and Development*, *29*(3), 398–416. http://dx.doi.org/10.1080/10409289.2017.1408373

Sprachman, S., Caspe, M., Atkins-Burnett, S., & Kennen, B. (2009). *Language Interaction Snapshot (LISn) Coding Manual*. Mathematica Policy Research, Inc.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, *98*(1), 63–74. http://dx.doi.org/10.1037/0022-0663.98.1.63

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*(2), 199–209. http://dx.doi.org/10.1016/j.ecresq.2012.12.002

Weizman, A., & Snow, C. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. http://dx.doi.org/10.1037/0012-1649.37.2.265

Whorrall, J., & Cabell, S. (2016). Supporting children's oral language development in the preschool classroom. *Early Childhood Education Journal*, *44*(4), 335–341. http://dx.doi.org/10.1007/s10643-015-0719-0

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *The Woodcock-Johnson Tests of Achievement* (3rd ed.). Riverside.

Vygotsky, D. H. (2001). *Pedagogy*. Routledge Falmer.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., et al. (2013). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development. https://www.fcd-us.org/the-evidence-base-on-preschool/

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*, 16–33. http://dx.doi.org/10.1111/mono.12032