Published in final edited form as:

Early Educ Dev. 2018 ; 29(3): 379–397. doi:10.1080/10409289.2017.1408371.

Teacher language in the preschool classroom: Initial validation of a classroom environment observation tool

Beth M. Phillips, Yuting Zhao, M. Jane Weekley

Florida State University

Abstract

Research Findings: This study reports initial descriptive and validity results of a new early childhood classroom observation measure, the Classroom Language Environment Observation Scales (CLEOS), designed to capture teachers' use of both implicit language supports (e.g., incidental scaffolding and shared reading) and more explicit language instruction (e.g., direct vocabulary instruction). Classrooms ($n = 122$) serving at-risk three-to-five year-old children, and representing child care, Head Start and public prekindergarten were observed; a subgroup was also observed with the Teacher Behavior Rating Scale (TBRS), a well-validated tool. Results indicated limited use of most language-support strategies, particularly those that were more explicit. Concurrent validity for the CLEOS was supported via significant correlations with TBRS subscales. Greater use of higher quality linguistic input was significantly associated with teachers' years of experience but not with their educational level. Findings supported the differential inclusion of linguistic input across settings, with large group circle time being the most frequent setting for explicit instructional input and centers being the most frequent setting for incidental supports.

Practice or Policy: Study results suggest a need to improve the professional development and preservice training for preschool teachers related to supporting rich language interactions and explicit language and vocabulary instruction within classrooms.

Ampleresearch connects early language skills to later reading achievement (e.g., Sénéchal & LeFevre, 2002; Storch & Whitehurst, 2002). Once children master decoding, reading and oral language have a reciprocal relation, such that wide reading leads to vocabulary growth (Cunningham & Stanovich, 1997), while increases in vocabulary support reading comprehension (e.g., Quinn, Wagner, Petscher, & Lopez, 2015). Before children can read, however, they are reliant on opportunities for language development provided within the linguistic input of home and early schooling settings. This paper explores a key component of young children's environments; namely the language used by teachers in preschool classrooms, a context now experienced by most children for at least one year prior to kindergarten.

Correspondence concerning this article should be addressed to: Beth M. Phillips 2010 Levy Avenue Suite 100 Tallahassee FL 32310 bphillips@fcrr.org.

Beth M. Phillips, Department of Educational Psychology and Learning Systems and Florida Center for Reading Research, Florida State University; Yuting Zhao, Department of Educational Psychology and Learning Systems, Florida State University; M. Jane Weekley, Florida Center for Reading Research, Florida State University.

Although most young children develop adequate language skills, a substantial minority have difficulty in this area, either because of intrinsic learning limitations (e.g., Leonard et al., 2007), poor home language support (e.g., Hart & Risley, 2003), or both. Children from lower-SES backgrounds are particularly likely to experience high rates of delayed or below average language skills (Hoff, 2013; Nelson, Welsh, Trup, & Greenberg, 2011). Many children's language difficulties do not resolve, but instead exert long-term influence on reading and academic performance (e.g., Dale, McMillan, Hayiou-Thomas, & Plomin, 2014).

Despite these findings, little attention is given in preschool settings to the sufficiency or pace of growth of many children's language skills (e.g., Hindman & Wasik, 2013; Neuman & Dwyer, 2009) and by kindergarten entry many children from lower-SES backgrounds have large gaps in vocabulary that can substantially impede their knowledge acquisition from reading and other educational experiences (Hammer, Farkas, & Maczuga, 2010). To close such gaps, children from lower-SES homes must develop language skills at an accelerated rate (Biemiller, 2006; Marulis & Neuman, 2010). Fostering acceleration requires high-quality, intentional language modeling and instruction within preschool classrooms (Cabell et al., 2011). Currently, however, the quality of the children's experiences in preschool and child care settings, particularly for children from lower-SES backgrounds, remains quite low (e.g., LoCasale-Crouch et al., 2007), perhaps below a threshold of meaningful impact (Burchinal, Vandergrift, Pianta, & Mashburn, 2010). Moreover, tools currently available are not optimal for careful assessment of the quality of teachers' language use and of their instructional techniques.

Measurement and quality of classroom language environments

Teachers exhibiting higher-quality language input do so through both incidental, conversational behaviors and through use of more intentional, explicit instructional techniques. Incidental teacher language strategies that influence children's language growth include using decontextualized language, asking open-ended questions (Girolametto, Weitzman, & Greenberg, 2003; Lonigan, Purpura, Wilson, Walker, & Clancy-Menchetti, 2013), using varied and abstract vocabulary (Bowers & Vasilyeva, 2011; van Kleeck, Vander Woude, & Hammett, 2006), extended discourse on topics (Dickinson & Porche, 2011) and use of syntactically complex utterances (Huttenlocher et al., 2002). With regard to intentional instruction, older (e.g., Marulis & Neuman, 2010) and recent studies support the efficacy of explicit instruction of vocabulary words (e.g., AUTHORS, 2013; Neuman, Newman, and Dwyer, 2011). In addition, robust correlational (Zucker, Cabell, Justice, Pentimonti, & Kaderavek, 2013) and experimental (Lonigan, et al., 2013) evidence supports the benefits of interactive shared reading.

Research consistently indicates that teachers vary substantially in the frequency, complexity, and form of their speech to children and in their use of intentional instructional techniques (Bowers & Vasilyeva, 2011; Turnbull, Anthony, Justice, & Bowles, 2009). These studies approach measurement of classroom language environments from different 'grain sizes'. First, studies using broad measures such as the *Classroom Assessment Scoring System* (CLASS; LaParo, Pianta, Hamre, & Stuhlman, 2002), reveal the quite low performance of preschool classrooms on the *Instructional Support* scale, which includes some focus on

language instruction (Denny, Hallam, & Homer, 2012; Justice, Mashburn, Hamre, & Pianta, 2008). Second, studies using the more instructionally-focused but still general *Teacher Behavior Rating Scale* (TBRS; Landry, Crawford, Gunewig, & Swank, 2000) suggest a wide range of teacher language quality (Lonigan et al., 2015; Landry et al. 2014). Finally, studies reporting amounts of child-directed language from preschool teachers or how much time they spend providing language-focused instruction revealed that teachers give this content area limited attention (e.g., Chien et al., 2010). On the whole, findings from many studies converge to suggest that language environment may be an especially weak aspect of classroom quality, even when teachers receive professional development (PD) supports (Cabell et al., 2011; Girolametto et al., 2003).

One common limitation among most of these investigations is that they typically grouped many different linguistic features utilized by the teachers' into broad categories or collapsed language use across multiple settings rather than detailing distinct types and patterns of language use. Measurement tools capturing more fine-grained patterns of teachers' enacted language use, such as the tool described here, may better enable understanding how and when teachers maximize the quality of their language interactions and may better support remediation of teachers' missed opportunities to foster language development. To date, however, most measures allowing fine-grained study of teacher-child language interactions have only addressed shared book reading (e.g., Blewitt, Rump, Shealy, & Cook, 2009; Zucker, Justice, Piasta, & Kaderavek, 2010). Fewer studies have captured more detailed linguistic interactions outside this one context (e.g., Dickinson, Hofer, Barnes, & Grifenhagen, 2014; Gest et al., 2006) and these have typically involved very time-consuming language sample or video analyses.

Role of Teacher Characteristics

Evidence is mixed regarding the relation between teacher background characteristics such as education and specific credentials and preschool classroom quality. Some research indicates that teachers with higher credentials engage in better quality practices (e.g., Burchinal, Cryer, Clifford, & Howes, 2002; Denny et al., 2012). For instance, Miller and Bogatova (2009) revealed significant increases in observed classroom quality as teachers accrued credits and certification, and several studies (e.g., Gerde & Powell, 2009) indicated that teachers with more education used higher quality extra-textual comments during shared reading. However, LaParo, Sexton, and Snyder (1998) found no relations between teacher characteristics and observed quality, and a relation between teachers' education and classroom quality reported by Phillipsen, Burchinal, Howes, and Cryer (1997) disappeared when structural setting features were added to their model. Most notably, Early et al. (2007) reviewed multiple studies and found little association between teacher education and classroom quality or child outcomes. Such mixed evidence requires further exploration of whether teacher education predicts language quality.

Studies exploring whether teacher experience influences classroom quality and child outcomes also reported mixed findings. Whereas Kontos and Wilcox-Herzog (2001) and Hindman and Wasik (2013) did not report a significant relation, Phillips, Gormley, and Lowenstein (2009) indicated a positive association between years of teaching and several

classroom language and shared reading quality indicators in Oklahoma prekindergarten sites. Similarly, Denny et al. (2012) reported that experience predicted CLASS-rated instructional support quality and LoCasale-Crouch et al. (2007) indicated that this variable significantly predicted classroom quality profile membership. One explanation for divergent findings is that teacher characteristics are more consequential in less administratively-supported contexts of child care centers, as opposed to in Head Start or public prekindergarten sites. Or, as alluded to by Bogard, Traylor, and Takanishi (2008), perhaps teacher expertise in delivering high quality instruction is the active agent, regardless of experience or education. Notably, few studies have investigated links between teachers' education and experience and their observed language use beyond the book reading context; this was therefore a primary aim of the current study.

Goals of the Current Study

Currently employed classroom observational tools have both strengths and weaknesses (Burchinal, Kainz, & Cai, 2011; Keys et al., 2013). As a strength, measures such as the CLASS and the TBRIS highlight intentional language scaffolding and instruction by teachers as areas of recurring weakness in early childhood settings, and they fit well within PD contexts where a mentor or supervisor might need to provide frequent, specific feedback to a teacher regarding observed behaviors. However, the weakness of these measures is that they are by design relatively broad in their treatment of teachers' language use, which is one of numerous instructional aspects being simultaneously observed; as such the grain size of the observation is larger and more generalized, rather than specific and detailed regarding numerous types of teaching behaviors. Moreover, no extant rating scale discretely scores language support by classroom setting beyond book readings. In contrast, the strength of the alternative tool, classroom language samples, is that they can capture extensive details regarding teacher-child linguistic interactions (e.g., Dickinson et al., 2014). However, these methods require highly labor-intensive transcribing and coding processes. As a result, language sampling is not feasible within large-scale studies requiring large numbers of classrooms or many observation waves or within PD contexts requiring real-time feedback. Whereas the strengths and limitations of current instruments may be complementary, neither type of observational tool is both sufficiently detailed and sufficiently practical to meet the acute need within the early childhood field to improve classroom language environments, with respect to teacher's incidental language use and their explicit, intentional instruction; this realization motivated development of our new measure.

We responded to recent calls for better measures of classroom quality (e.g., Burchinal et al., 2011) and recognized a gap in available instruments in providing a tool that is specific to language, to better enable in-depth capture of multiple facets of linguistic behavior, while simultaneously being more readily scorable than the time-consuming coding of audio-recorded language samples or videos (e.g., Massey et al., 2008; Pentimonti et al., 2012). Thus, we created the *Classroom Language Environment Observation Scales* (CLEOS; Phillips et al., 2011) to combine the feasibility of measures like the CLASS for collecting data in a large number of classrooms (e.g., Howes et al., 2008) with the narrow and rich focus of the language-sampling methods that could identify specific aspects of language quality as robust or absent in a given teacher's behavior (e.g., Dickinson et al., 2014;

Turnbull et al., 2009). As summarized earlier, both intentional, explicit vocabulary and grammar instruction and more implicit, incidental language use by teachers positively relate to children's vocabulary and language growth within preschool classrooms. Thus, in alignment with this correlational and experimental evidence, the CLEOS focuses on vocabulary, language, and book reading, rather than just one or two of these key language interaction and instructional elements. Another relatively unique feature of the CLEOS is its organization of teachers' language input into subscales based on whether teacher behaviors are focused on vocabulary acquisition or are aimed at supporting children's broader expressive language skills (e.g., production of longer utterances, participation in multi-turn conversations). The measure also separately captures implicit, often incidental vocabulary instruction (e.g., modeling of specific object labels during play interactions, reading of narrative texts) versus more explicit teaching behaviors (e.g., using props and gestures to support understanding of new vocabulary, providing child-friendly definitions for new words). The implicit/explicit distinction rests on whether the children have to infer a word's meaning themselves or are directly provided with this semantic information. The CLEOS also is coded by classroom setting, to allow researchers and educators insight into the daily activity contexts where teachers are most and least likely to need assistance. This organizational structure is designed to allow researchers to parse the specific association of these distinct aspects of teacher language with child language and is intended to readily translate into precise, actionable guidance to teachers, such as within a PD model where guidance toward higher-quality language modeling and instruction is coupled with observational feedback from mentors, and teachers themselves, as part of reflective self-study and peer support systems (e.g., Landry et al., 2014).

We conducted this initial series of studies to investigate the CLEOS' validity and its associations with teacher characteristics, and to explore the nature and quality of preschool classroom language environments. Another goal was to demonstrate that items could be reliably completed by observers with moderate levels of training and varied backgrounds. The primary aims were to: (1) Describe the classroom oral language environments for preschool classrooms representing a range of site types; (2) Demonstrate convergent validity with a widely used classroom observation measure that includes book reading and oral language subscales; (3) Explore relations with teacher background characteristics; and (4) Describe and compare teachers' use of language across varied classroom settings.

Method

Participants

Classrooms were recruited to participate in this study via one of two routes. First, early childhood sites serving 3-to-4-year-old children from backgrounds of low SES were recruited as part of a larger vocabulary intervention development project (Phillips, Oliver & Willis, 2017). Classrooms were recruited from child care, public, and Head Start site types to ensure generalizability across administrative structures. Second, we recruited other classrooms serving the same age group within the same area, which included both rural and urban locations in the southeast, but which enrolled children from varied SES backgrounds. Our goal in expanding the SES-related eligibility criteria was ensuring that classrooms observed

demonstrated diversity with respect to teacher credentials and child and teacher demographics. Whereas we requested curriculum information only from teachers in the development project (61%), those data indicated that classrooms were using an array of published and home-grown curricula (e.g., over 16 programs).

Classroom teachers participated in this study in one of two partially overlapping subgroups of classrooms serving 3- and 4- year old children, both of which included observation using the CLEOS measure. The first, primary subgroup ($n = 122$) included the observation and lead teacher completion of a background survey. Public pre-kindergarten (13%), Head Start (10%) and subsidized child care (77%) were all represented. Between 1 and 6% of teachers did not complete specific personal background items. All lead teachers were female and their ethnic/racial identities were African American (40%), White (58%) and other (2%); none self-identified as Hispanic/Latino. Lead teachers ranged in age from 16 to 62 years ($M = 39.46$, $SD = 12.14$). There also was a wide range (less than 1–37 years) of preschool teaching experience ($M = 7.33$, $SD = 7.73$). The average number of years of education was 13.95 ($SD = 2.13$); 15.0% had a high school diploma, 35.8% had a Child Development Associate's Degree (CDA), 21.7% held an associate's degree, and 27.4% had a bachelor's or higher degree as their highest credential.

The second subgroup of classrooms ($n = 43$), which included one male teacher, represented the validity sample and were observed both with the CLEOS and with the Teacher Behavior Rating Scale-Plus (TBRS-P). A minority (12%) were unique to this sample, while 21% were those for whom the same observation was included within both primary and validity analyses. The other 67% were teachers who had data from different years included in the primary and validity datasets (e.g., fall of one year for the primary sample and summer two years later for the validity sample). The auspice representation in the second sample included Public Pre-Kindergarten (5%), Head Start (9%) and subsidized child care (86%); this increase in the proportion of child care sites occurred because some of these data were collected in summer when other sites were closed. Demographic data were missing from a few teachers (3). Teachers in this group ranged in age from 20 to 60 years ($M = 40.00$, $SD = 12.93$), reported experience from 1 to 41 years ($M = 9.90$, $SD = 9.91$), and included 14.3% with the highest credential of a high school diploma, 26.2% with a CDA, 26.2% with an associate's degree, and 33.4% with a bachelor's degree or higher. Ethnic representation included 41.0% African American, 56.5% White, and 2.5% Asian or other.

Measures

Three measures were used to evaluate teachers' characteristics and classroom language environment. The background survey and CLEOS were completed for all teachers. The TBRS-P observation was completed with a subset of the sample as described above.

Teacher background survey.—Teachers were asked to complete a brief questionnaire (18 items) regarding their demographic background (i.e., gender, age, and ethnicity), educational attainment, and teaching experience (i.e., years taught at various grade levels).

Classroom Language Environment Observational Scales (CLEOS).—This measure was developed to allow detailed observational assessment of the language

environment that the lead teacher and (where present) the aide created during a typical day in preschool classrooms. Coding captured both incidental/conversational language and intentional/instructional language choices of the teachers and also identified the number of classroom settings (i.e., circle time, small group, centers, gross motor, snack/meals, and transitions) in which each of these language behaviors occurred.

In all areas except book reading, the primary scoring system awarded higher points for classrooms in which the desirable teacher language behaviors occurred in a greater percentage of observed contexts out of the six possible classroom contexts. This “coverage” coding was used as a key index of quality instead of attempting to conduct frequency counts on each of many items during live coding. However, for a few behaviors, anticipated to be much more common, a minimum number of observed instances were required to score the item as observed within the setting. A secondary scoring system included a simple count of the items on each scale if they were ever observed, regardless of instructor or the number of contexts in which the item was observed. This more generous coding system is more akin to the present/absent scoring of other rating scales (e.g., *ECERS-R*; Harms, Clifford, & Cryer, 1998; *Classroom Observation Tool*; Crawford, Zucker, Williams, Bhavsar & Landry, 2013), but does not account for diversity or multiplicity of contexts and thus is a less rigorous index of the consistency of teachers’ implementation of the behaviors and the extent to which children were exposed to them.

The measure targets teachers’ language use and their vocabulary and language instruction in five subscales of General Language Environment (GLE, 8 items), Book Reading Quality (BRQ, 13 items), Incidental Language Instruction (ILI, 10 items), Incidental Vocabulary Instruction (IVI, 5 items), and Explicit Vocabulary Instruction (EVI, 8 items). Example items for each subscale and internal consistency ratings are included in Table 1. Book reading items were coded each time a teacher or aide read to the children, whether in large or small groups or with a single child. Comparable to how other items were scored across settings, book reading behaviors were scored cumulatively across readings, such that if a teacher read two times and engaged in a particular quality behavior in the first reading and in another specific behavior in the second reading, both of these items would be marked as observed. The book reading subscale items therefore were all scored as “ever observed” and summed. Consequently, to retain all classrooms in analyses, book reading total scores were set to zero for classrooms in which no book reading occurred. Circle time was indicated when children all gathered with a teacher but were not being read to; activities may involve singing, calendar activities, finger plays, etc. Small group was indicated when children had been divided by teachers into small groups to play and work together, or when a small group of children incidentally gathered with the teacher for a sustained activity while all other children were engaged in free play during Centers. Thus, both Centers and Small Group could be marked simultaneously, for example if a teacher was leading a small group while the aide was supervising Centers for all other children.

Centers were indicated when children had opportunities for free choice play at different centers such as dramatic play, block center, science center, library/listening center, and art center. Gross motor was indicated when children were doing activities involving large motor skills, either inside or outdoors. This activity was coded regardless of whether the teachers

participated. Meals were coded when children were eating snacks, breakfast, or lunch. Transitions were marked when children were in the process of ending one activity or group setting and initiating the next (e.g., moving from free play centers to line up to go outside, cleaning up the room). For each item, descriptive coding indicated whether the lead teacher and/or the aide (where present) demonstrated the behavior. However, notating of contexts in which items were observed was marked just once per item regardless of which teacher engaged in the target behavior. In this way, the measure yields comparable item-level and scale level scores all on the same metric irrespective of whether the classrooms had aides or assistant teachers, so as not to penalize classrooms without aides. The CLEOS had no missing data.

To determine concurrent validity with an established classroom observation measure, we utilized an augmented version of the TBRs, labeled here as the TBRs -Plus (TBRs-P). The original instrument (Landry et al., 2000) is a reliable and well-validated comprehensive classroom observation measure that appears responsive to instructional changes and PD (Landry, Anthony, Swank, & Monseque-Bailey, 2009; Landry et al., 2014, Lonigan et al., 2015). For several large scale projects, the first author and colleagues augmented the TBRs to include more phonological awareness and language instruction items than are in the original version. Essentially, the TBRs-P oral language subscale includes within a single composite items representing multiple language-related behaviors that the CLEOS represents more extensively in four distinct subscales. As shown in Table 1, the TBRs-P oral language scale obtained high internal consistency reliability. There are 12 total scales in the TBRs-P including three under general teaching behaviors (i.e., classroom community, sensitivity, and discipline), lesson plans, centers, book reading behaviors, print and letter knowledge, math concepts, phonological awareness, written expression, oral language use with students, and team teaching. As included sites varied both in whether they were required to use written lesson plans and in whether they had a second teacher present, we excluded consideration of those two scales.

Observers rated items in each scale both in quantity (e.g., *rarely*, *sometimes* and *often*), and quality (e.g., *low*, *medium low*, *medium high*, and *high*). Each area thus produces quantity and quality subscales. There was a very small percentage (i.e., < .005%) of missing data on TBRs-P items that were determined to be missing at random and were imputed with EM before composition of scales and total scores so as to retain all teachers in the calculation of all variables. For these analyses quantity and quality scores within each content area were averaged, as they are typically, and in this sample, very highly correlated. Total scores were used for discipline, which includes only one quality rating, and for centers, which includes only one quantity rating. Per TBRs scoring rules, instructional types not observed are assigned zeroes for all items, yielding a zero for the total score in that content area.

Procedures

Observers included research assistants with a range of backgrounds, including advanced undergraduate students, part-time research assistants who were former teachers, and three senior project coordinators with advanced degrees in speech-language or special education who also served as the “gold-standard” observers who trained the others. All observers

completed self-study review of each measure and received approximately 12 hours of workshop training in completing each of the CLEOS and TBRSP. After the training, each observer completed joint observations with a trainer where the trainer and new observer discussed the coding during and after the observation. Each then conducted 3–4 practice observations simultaneous with but independently of a trainer-observer in non-project classrooms. Initial clearance for the CLEOS required achieving the high reliability standard of over 90% inter-rater exact agreement with an experienced trainer on individual item scores (present/absent) and item-specific context counts for the CLEOS. In addition, approximately 15% of all observations were completed by two assigned observers to enable calculation of inter-observer consistency (i.e., intraclass correlations, Shrout & Fleiss, 1979). These intraclass correlations ranged from .79 to .99 ($M = .90$) across the five CLEOS subscales, an indication of excellent consistency (Cicchetti, 1994; Hallgren, 2012). To receive field clearance for the TBRSP, observers had to achieve a minimum agreement standard of at least 85% agreement (i.e., within one point of a gold-standard trainer) overall and at least 80% on each subscale. All observers completed recalibration observations each season in which they again conducted simultaneous observations, to minimize drift and to protect against rater severity biases (Kelcey, McGinn & Hill, 2014).

Approximately one-third (38) of the classrooms were observed during one school year, whereas the remainder were observed over the subsequent two school years. Including the initial 38, most (79) were observed in the late fall, when classroom routines and relationships would likely be well established. Another 43 classrooms with year-round programs were observed during late spring or summer. All observations took place during a two to three-hour morning period, on a day that was randomly selected from within a target window to the extent possible. Observers were trained to be unobtrusive and non-participatory, but to stay close enough to teachers to accurately code their language; item scoring was revised as needed as different settings were observed and descriptive field notes also were used to finalize all scoring. All classrooms recruited to participate in the vocabulary intervention project were observed for this study before introduction of any PD or novel instructional materials at those sites. For classrooms in which both the CLEOS and TBRSP were completed, both observations took place within the same two-week window and were randomly allocated to take place first or second. All validity observations were completed during spring or early summer.

Results

Most analyses utilized the primary sample of 122 preschool classrooms, for which descriptive statistics are shown in the left panel of Table 2. Across the 122 classrooms, 30% of observations indicated one teacher present, 62% indicated two (lead and assistant/aide), and approximately 9% indicated three or (in one classroom) four teachers present. The classrooms served a varied number of children, ranging from 7 to 22. As described above, the scores for each of the CLEOS language environment subscales except Book Reading Quality are weighted by the number of different classroom contexts in which the target behaviors are observed. Across classrooms, of the six possible activity contexts, 34% of the classrooms were seen engaged in four distinct contexts, and 32% and 18% were observed engaged in five or six contexts, respectively. Few classrooms were observed to engage in

fewer than four of the contexts (i.e., just 2% with two contexts and 14% with three). Approximately 25% of the classroom observations in the primary sample included no book reading, whereas about 6% included four or more sessions. Overall, the average was 1.35 ($SD = 1.19$) observed book reading sessions.

The first research aim explored average classroom language quality for each CLEOS subscale. As seen in Table 2, using the primary scoring method weighted by the range of contexts, the general language environment (GLE) had a mean of 5.69 ($SD=1.47$); this scale mean ranged from 2.00 to the maximum 8.00, whereas the other four subscales of ILI, IVI, EVI and BRQ all were considerably lower, relative to their specific maximum possible scores. These four scales all included zero in their score range, but also included classrooms scoring at or near the maximum, suggestive of considerable variability in the language environments between classrooms. Overall, descriptive results indicated that many of these preschool classrooms had quite low quality language support and minimal explicit instruction, although the teachers' incidental language use as indexed by GLE was relatively strong. Use of the more generous simple additive scoring (i.e., "ever observed") yielded somewhat higher mean scores (again relative to the maximum possible), particularly for ILI, which benefited the most from the more liberal scoring method, suggesting that many teachers engaged in a small amount of language support but that it was not robustly observed across multiple settings. The more conservative weighted scores were utilized in all subsequent analyses, including those in the validity sample. As approximately one-third of the classrooms were observed during the spring or summer, sensitivity analyses explored whether means were different for classrooms seen in different seasons. For only one scale, ILI ($F[1, 120] = 11.72, p < .01$), were there any significant differences. ILI data collected in the spring/summer indicated a slightly higher mean than for data collected in the fall (i.e., 3.30 [$SD=1.37$] vs. 2.50 [$SD=1.16$]). However, the spring/summer classrooms also included most serving middle-income populations, so it is not possible to distinguish the influence of time of year from that of other classroom characteristics. Comparably, other sensitivity analyses compared classrooms observed for 2.5 hours or longer and those observed for a briefer time; minimal differences were found; length of observation was related to the actual length of the day--and thus language exposure-- provided by each program.

Correlations among CLEOS subscales, presented in Table 3, indicate positive, generally significant, and moderate but not overly high relations among the subscales. Whereas these positive relations indicate that teachers strong in one aspect of language support are likely to also be at least moderately good in another area, the correlations also indicate that teachers' quality of language-input, when conceptualized by focus (language versus vocabulary) and technique (incidental versus intentional/explicit) vary somewhat independently and can be measured distinctly. In these analyses with all 122 classrooms, neither GLE nor ILI were correlated with book reading quality. Given the substantial number of classrooms in which book reading did not occur, we replicated analyses regarding book reading quality for only those classrooms (92) in which the activity was observed. As expected, the mean was somewhat higher ($M=6.48, SD=3.03$). Correlations for most other CLEOS scales with book reading quality for this restricted sample were higher (i.e., $r_s = .16, .42, .32$, and $.40$, for GLE, ILI, IVI and EVI, respectively) and the correlation with ILI now obtained significance.

Concurrent Validity Analyses

The second set of analyses concerned aim two, the concurrent validity of the CLEOS measure with the TBRIS-P. These analyses utilized the smaller subsample of 43 classrooms, for which the descriptive statistics are shown in the right panel of Table 2. The descriptive statistics for the TBRIS-P subscales are presented in Table 4. As shown in the Tables, this subgroup of classrooms had generally comparable CLEOS scores to the primary sample, with the exception of somewhat lower book reading quality behaviors and less frequent explicit vocabulary instruction. On the TBRIS-P, scores for general instructional behaviors (e.g., sensitivity and discipline) were moderate but means for many of the specific instructional areas indicated lower frequency and quality of instruction. For the oral language scale, the classroom instructional environment was moderate, but with a wide range from poor to strong. In contrast, phonological awareness was rated as weak, on average, with no classroom achieving a high score. On most scales there were wide ranges, indicating substantial variability among classrooms.

One finding of particular note was that the percentage of classrooms not observed as engaging in book reading was even greater in the validity sample, where 49% of the CLEOS observations and 28% of the TBRIS-P observations did not include any book reading interactions among teachers and children. To account for the scoring rules allotting zeros to classrooms where no book readings occurred, we also calculated both book reading quality means and standard deviations for only those sites at which a book reading was actually observed. As expected, these scores were higher than those reported in Tables 2 and 4 (i.e., CLEOS BRQ: $M = 7.18$, $SD = 2.95$; TBRIS-P Book Reading: $M = 16.28$, $SD = 5.66$).

The primary question of interest for these analyses was whether scales from the CLEOS would demonstrate significant concurrent correlations with scales assessing corresponding constructs on the TBRIS-P. As depicted in Table 5, all of the CLEOS scales except GLE, which measures ambient language use not instruction, had significant and sizable correlations with the oral language scale on the TBRIS-P. Likewise, the two book reading scales were moderately and significantly correlated. Most CLEOS scales also were significantly and substantially correlated with the TBRIS-P Writing scale, suggesting that high quality language interactions may be particularly likely to occur in the context of shared or scaffolding writing activities, or at least in classrooms where higher quality writing instruction was observed. In contrast, most correlations between the CLEOS subscales and the math and code-focused scales on the TBRIS-P (e.g., print knowledge and phonological awareness) were nonsignificant. Correlations with the general classroom environment aspects of the TBRIS-P were mixed but generally positive and statistically significant, with each CLEOS scale yielding significant correlations with at least two of TBRIS-P sensitivity, community, discipline, and centers, respectively. For example, the ILI scale was significantly correlated with both sensitivity and discipline, both of which emphasize teachers' use of responsive and supportive behaviors and encouragement of self-regulation by children through linguistic expression (e.g., describing their own emotions).

Relations between CLEOS and Teacher Characteristics

The third set of analyses utilized bivariate correlations to address the research aim of exploring relations between CLEOS scores and teacher characteristics. Teachers' years of experience was positively related to ILI, IVI, and EVI (see Table 3). Teachers' age was significantly associated with EVI. In contrast, years of formal education was uncorrelated with all CLEOS subscales and also was unrelated to age and years of experience.

Comparing Classroom Contexts

Analyses for aim four, regarding the patterns of teacher language use and instruction in different classroom contexts were conducted using the primary sample in two ways. First, we determined the averages and variance only including those classrooms in which each particular context was observed. We then repeated these analyses with the entire sample. That is, item scores for those classrooms in which the setting did not occur were recoded from missing to zero. This recoding allowed the scores to indicate whether or not teachers' utilized the opportunity to intentionally or incidentally provide that particular context of teacher-child language interaction. Given that all six contexts are typical in early childhood educational settings, and that some may provide more optimal language interaction opportunities than others (Cabell, deCoster, LoCasale-Crouch, Hamre, & Pianta, 2013; Gest et al., 2006), this was considered an appropriate, albeit demanding, analytic approach. Descriptive statistics for each type of linguistic interaction as observed in each of the six contexts are presented in Table 6. The means and standard deviations provided in the upper panel include just those classrooms in which we observed the specific context. As depicted in the Table, the means were low, particularly for incidental and explicit vocabulary instruction. The more exacting analyses, including the entire sample, are presented in the lower panel; results generally paralleled the more generous findings in the upper panel.

Within-sample analyses were conducted using the primary sample to compare the means of each subscale across the different classroom settings. Subscripts within each row in the bottom panel of Table 6 indicate which means were and were not significantly different from one another after Bonferroni adjustment for multiple comparisons. As can be seen, many of the means were significantly distinct, with clear patterns across most scales for the quality of language use and instruction in circle time and centers being the highest, and those for meals and gross motor typically being the lowest. Whereas general language during transitions was moderately good, all other quality features were particularly weak during these frequent routines.

Discussion

The broad aims for this study were to introduce a new observational measure of the classroom language environment in early educational settings and to describe the quality of and specific nature of the language support experienced by children in these classrooms. Overall, this close-in view revealed that language environments are generally poor to moderate, at best. There is a particular weakness in the provision of explicit, intentional vocabulary instruction. The related study aim was to provide initial concurrent validation against a widely used existing measure of the classroom instructional environment.

Concurrent correlational evidence suggests that the individual CLEOS scales have significant, moderate to high convergent correlations with the scales on the TBRs-P that measure comparable features of the classroom context.

Findings are consistent with prior studies indicating that the language environment of early education classrooms do not comport with indicators of high quality (Denny et al., 2012; Neuman & Dwyer, 2009). Similar to Chien et al. (2010) and Cabell et al. (2013) who indicated that intentional language instruction was observed during very small percentages of classroom time, results from the CLEOS primary scoring system, in which higher scores represent the specific quality marker being observed in a wider array of activity settings, indicated limited inclusion of high quality incidental or intentional language support strategies throughout the day.

Moreover, results from the more generous scoring method where items were credited as “ever observed” also indicated moderate quality, on average, and a wide range in practices among participating teachers. These results are especially notable given that both scoring methods gave credit for language strategies as used by either the lead or assistant teacher (where present), and that each scale includes items that are achievable without requiring elaborate, formal lesson plans. Such findings suggest that many early childhood settings are missing important opportunities to support young children’s language development through the provision of evidence-based incidental and explicit instructional strategies. Given that having weak language skills is associated with a range of negative long-term consequences for children, these results signal a clear need for improved pre-service and in-service training for teachers in all early educational settings regarding how best to create a more robust language environment.

One finding was that better quality language support was not necessarily a marker of better instructional quality in other content areas, as indicated by the general absence of significant positive correlations with TBRs-P math, or code-related instructional subscales. These results suggest that teachers may have modularized expertise in, or motivation for, evidence-based instruction in just some content areas, rather than demonstrating high quality more generally. Evidence from home literacy research indicates comparable modularity in what parents choose to emphasize (e.g., Sénéchal & Lefevre, 2002). Within early childhood classrooms, these results direct researchers and educators to think more precisely about defining process quality (see also Keys et al., 2013; Weiland, Ulvestad, Sachs, & Yoshikawa, 2013) and also to not assume that teachers possess or enact a generalized instructional expertise. It may be the case that, given many competing demands, teachers choose content areas to emphasize rather than embedding high quality language instruction throughout content areas. Future research should explore whether robust language instruction fits better or is more impactful within particular content foci and how such integrations can be successfully modeled within PD content.

One study aim was to add to the small literature regarding the classroom activities during which teachers are demonstrating their best language support behaviors (e.g., Cabell et al., 2013; Dickinson et al., 2014). Closer examination by specific contexts suggested that many teachers appeared to be restricting their provision of intentional and even incidental

instructional language behaviors primarily to circle times and centers. Whereas this finding demonstrates the promise that some teachers can and do support language, circle time represents a relatively small percentage of the daily schedule in this and other samples (e.g., Early et al., 2010), and neither of these settings likely allows as much opportunity for children to individually respond and receive scaffolded support as might small group interactions (e.g., Booren, Downer, & Vitiello, 2012; Fuligni et al., 2012), which were substantially less likely to occur. However, in the roughly 50% of classrooms where any small groups occurred, they were imbued with somewhat higher levels of incidental or explicit vocabulary instruction, relative to transitions, gross motor, and meal times. Thus, small group interactions may be a promising context in which to encourage higher quality teacher-child language interactions (Ruston & Schwanenflugel, 2010; Turnbull et al., 2009). In contrast to findings by Gest et al. (2006) and Cote (2001), snack and meal times were generally not utilized by teachers as an opportunity for rich language support. These results indicate that teachers are missing prime opportunities to support children's language in authentic contexts, perhaps because their focus is on classroom management and procedures, or perhaps because their training did not emphasize frequent conversational interactions with children.

Findings from this study add to the literature indicating that preschool teachers' own educational experiences are not necessarily related to the quality of their instruction (e.g., Early et al., 2007). Of particular relevance to this issue is that the present sample included a large proportion of community child care classrooms. Historically (e.g., Phillipsen et al., 1997; Vu, Jeon, & Howes, 2008) and conceptually, this is the program type in which a relation between quality and education level was most likely to occur, as these sites, within this study and in general, are less likely to have standardized curricula or regular PD that may in other programs mitigate the influence of education. Instead, only teachers' years of experience and age were related to the quality of their language use, perhaps because it takes time for teachers to learn how to focus on their incidental modeling and scaffolding of language in the midst of the high intensity demands of managing general instruction and children's behavior. Given that neither education or experience were related to explicit vocabulary instruction, this study lends support to the idea that it is specific pedagogical expertise, not accrual of degrees or time in a classroom, that best predicts which teachers will provide this aspect of high quality instruction (Powell, Diamond, Burchinal, & Koehler, 2010; Hindman & Wasik, 2012). The significant, positive association of teacher age with provision of explicit vocabulary instruction requires further exploration but may signify that teachers trained at specific periods of pedagogical perspectives are more likely to consider it necessary or appropriate to provide explicit instruction.

A Promising Tool for Research and Professional Development

Evidence from this study provides initial support for the reliability and validity of the new CLEOS measure. Observers, who did not all have classroom teaching experience or specialized language-development expertise, were able to achieve a very high level of inter-observer agreement on an exacting item- and context-specific comparison and on intraclass correlations at the scale level. Furthermore, validity correlations with the well-established and valid TBRS-P measure indicated significant moderate to strong interrelations. In

particular and as anticipated, these validity coefficients were strong with the TBRS-P Oral Language Scale, and between the two book reading quality ratings. Consistent with the conceptualization of CLEOS subscales as capturing related but distinct teacher behaviors, the correlations among CLEOS subscales all were significant but moderate; whether a teacher provided explicit vocabulary instruction, for instance, varied independently from whether she provided consistent high quality language models in her conversational interactions with children. Furthermore, the multi-component and smaller ‘grain-size’ of teacher instructional behaviors captured by the CLEOS may yield stronger relations with child language outcomes than have been seen with broader measures such as the CLASS and ECERS (Keys et al., 2013; Weiland et al., 2013).

One advantage to the CLEOS relative to most observational measures is that it addresses explicit vocabulary instruction not only within book reading contexts but across all typical classroom settings. Given how limited such vocabulary instruction seems to be, the capacity to capture it whenever it may occur can support more comprehensive understanding of teacher language behaviors and perhaps encourage teachers to embed such intentional instruction throughout their day. Compared to the TBRS-P and similar global quality measures, the CLEOS also has the advantage of providing a more in-depth perspective on teacher language use and has the feasibility and ease of explanation for use as a PD tool, where teachers can be given individualized feedback on the strengths and weaknesses of their language support behaviors both across types and classroom activity contexts. Such precise feedback is just what PD experts suggest is most actionable and beneficial for teachers (Crawford et al., 2013; Hamre et al 2012).

When high quality language input is provided, it is associated with positive gains in children’s language skill (e.g., Burchinal et al., 2010; Howes et al., 2008). For example, Gerde and Powell (2009), in the short term, and Dickinson and Porche (2011) and Zucker et al. (2013), in longitudinal studies, indicate that preschool teacher’s use of sophisticated language and analytic book-related comments predicted children’s vocabulary development and even fourth grade reading comprehension. The results of the current descriptive study indicate that there is a substantial gap, however, between present realities and the idealized impact of preschool environments. A new tool such as the CLEOS maybe a valuable asset both with respect to research investigating the individual, curricular, and contextual influences on teacher’s language in the classroom and with respect to its potential as a coaching and self-study tool for teachers engaged in active improvement efforts. By virtue of being exclusively focused on language, the CLEOS, particularly in an anticipated shorter but still very detailed version, can support teachers’ understanding of the full range of linguistic behaviors worthy of attention (e.g., both vocabulary and language and including both incidental and explicit pedagogical methods).

Evidence suggests that teachers are responsive to PD that is explicit and specific in identifying areas of strength and those in need of improvement (Hindman & Wasik, 2012; Crawford et al., 2013). Evidence also suggests that intensive, high quality PD can yield improvements in teacher language quality (e.g., Landry et al., 2014; Lonigan et al., 2015). Although impactful, many PD programs require extensive involvement of external experts (e.g., highly trained mentors). By making linguistic behaviors explicit, specific, and

measurable in discrete items, the CLEOS may assist preschool providers to develop concrete and attainable language use goals without such extensive, and expensive, external support. For example, the CLEOS may be useful as part of a “walk-through” toolkit, where school leaders can conduct targeted observations, or as part of a self-study professional learning community (e.g., Vescio, Ross, & Adams, 2008) both of which can build understanding of how teachers are and are not engaged in high-quality linguistic interactions and of where specific improvements are needed to improve children’s classroom experiences (e.g., Moss & Brookhart, 2013).

Limitations and Future Directions

Whereas this study had numerous strengths, including the relatively large and diverse sample size and careful attention to observer training and reliable measurement, it has limitations as well. Given limited resources, and a purely research-focused context that was not high-stakes, classrooms were observed for this study on one occasion, raising the possibility that despite including a randomly selected day within the observation window, the activities observed in some classrooms may have been atypical (Hill, Charalambous, & Kraft, 2012). Likewise, our standardized protocol of only observing during morning hours (as some programs were half-day) may have led us to miss additional teacher behaviors later in the day and there was some moderate variation in the exact length of our observations. Within-sample comparisons, however, indicated minimal differences in findings for shorter and longer observations. Although ideally we would have observed more often and for a longer time in each classroom, this study represents an important first step in validating and exploring the utility of the new measure and does represent the diversity in experiences provided by different preschool settings; programs only provided for two hours are by design limited in the learning opportunities they provide.

We were not able to collect and report the educational or experience histories of the classroom aides despite including their instructional behaviors within the CLEOS and TBRS-P observational ratings. In many cases, whereas the lead teachers in these classrooms remained consistent from day to day, assistants were rotated among classrooms on an intermittent basis, such that a particular aide may be present in one classroom within a site for several hours on a Monday, but not in that same classroom the following Wednesday. Such inconsistencies, and the inclusion of classrooms with no aides, limited our capacity to explore associations between aides’ characteristics and classroom language quality indicators. Similarly, an ideal timeline would have included all classrooms being observed in the same season; however, sensitivity analyses indicated minimal differences between classrooms based on the season of observation.

Whereas efforts were made to include a variety of classroom types, the majority included were from private child care programs. The more limited number of public and Head Start classrooms precluded analyses divided by program type. At the same time, as much recent research has excluded stand-alone community child care and has less frequently included classrooms serving 3- rather than 4-year-old children, learning more about the quality of instruction in these sites makes an important contribution. Ongoing observations in 100 novel classrooms representing all program types and both age groups will enable these

comparative investigations and allow validation of the CLEOS items and subscales against teacher language sample data, the gold standard with regard to concurrent validity. Planned item-level analyses of the CLEOS will further validate the measure's structure and allow reduction of items to a briefer set that has the highest informational and discrimination characteristics. Finally, this new project includes longitudinal child-language data from both standardized assessments and multiple structured and classroom-based language samples, to allow investigation of the CLEOS' predictive validity to this most relevant criterion. Given the substantial risks that children with weak language skills face during and beyond preschool (e.g., Dale et al., 2014), tools that can support educators in their efforts to strengthen these children's skills are much needed.

Acknowledgements

This research was supported by the Institute of Education Sciences under grant R305A080476 to the first author. Preparation of this work also was supported by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (2 P50 HD0552120-10). The views expressed herein are those of the authors and have not been reviewed or approved by the granting agencies. The authors express appreciation to Karli Willis, Sarah McElhaney, Kylie Flynn, Amy Augustyn and Jeanine Clancy for their contributions to the development and validation of this measure.

References

- Ard LM, & Beverly BL (2004). Preschool word learning during jointbook reading: Effect of adult questions and comments. *Communication Disorders Quarterly*, 26, 17–28.
- Biemiller A (2006). Vocabulary development and instruction: A prerequisite for school learning In Dickinson DK & Neuman SB (Eds.), *Handbook of early literacy research: Vol. 2* (pp. 41–51). New York, NY: Guilford.
- Blewitt P, Rump KM, Shealy S, & Cook SA (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, 101, 294–304.
- Bogard K, Traylor F, & Takanishi R (2008). Teacher education and PK outcomes: Are we asking the right questions? *Early Childhood Research Quarterly*, 23, 1–6.
- Booren LM, Downer JT, & V. E. Vitiello (2012). Observations of children's interactions with teachers, peers, and tasks across preschool classroom activity settings, *Early Education & Development*, 23, 517–538. [PubMed: 25717282]
- Bowers EP, & Vasilyeva M (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics*, 32, 221–241.
- Burchinal MR, Cryer D, Clifford RM, & Howes C (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science*, 6, 2–11.
- 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, 166–176.
- Burchinal M, Kainz K, & Cai Y (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings In Zaslow M (Ed.), *Quality measurement in early childhood settings*. Baltimore, MD: Brooks.
- Cabell SQ, DeCoster J, LoCasale-Crouch J, Hamre BK, & Pianta RC (2013). Variation in the effectiveness of instructional interactions across preschool classroom settings and learning activities. *Early Childhood Research Quarterly*, 28, 820–830.
- Cabell SQ, Justice LM, Piasta SB, Curenton SM, Wiggins A, Turnbull KP, & Petscher Y (2011). The impact of teacher responsiveness education on preschoolers' language and literacy skills. *American Journal of Speech-Language Pathology*, 20, 315–330. [PubMed: 21856968]
- Chien NC, Howes C, Burchinal M, Pianta RC, Ritchie S, Bryant DM, ... & Barbarin OA (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development*, 81, 1534–1549. [PubMed: 20840239]

- Cicchetti DV (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Cote LR (2001). Language opportunities during mealtimes in preschool classrooms In Dickinson DK&Tabors PO (Eds.), *Beginning literacy with language: Young children learning at home and at school* (pp. 205–221). Baltimore: Brookes.
- Crawford AD, Zucker TA, Williams JM, Bhavsar V, & Landry SH (2013). Initial validation of the prekindergarten Classroom Observation Tool and goal setting system for data-based coaching. *School Psychology Quarterly*, 28, 277–300. [PubMed: 24059812]
- Cunningham AE, & Stanovich KE (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental psychology*, 33, 934–945. [PubMed: 9383616]
- Dale PS, McMillan AJ, Hayiou-Thomas ME, & Plomin R (2014). Illusory recovery: are recovered children with early language delay at continuing elevated risk? *American Journal of Speech-Language Pathology*, 23, 437–447. [PubMed: 24686486]
- Denny JH, Hallam R, & Homer K (2012). A multi-instrument examination of preschool classroom quality and the relationship between program, classroom, and teacher characteristics. *Early Education & Development*, 23, 678–696.
- Dickinson DK, Hofer KG, Barnes EM, & Grifenhagen JF (2014). Examining teachers' language in Head Start classrooms from a Systemic Linguistics Approach. *Early Childhood Research Quarterly*, 29, 231–244.
- Dickinson DK, & Porche MV (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, 82, 870–886. [PubMed: 21413936]
- Early DM, Iruka IU, Ritchie S, Barbarin OA, Winn DC, Crawford GM, & ... Pianta RC (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, 177–193.
- Early DM, Maxwell KL, Burchinal M, Alva S, ... Zill (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78, 558–580. [PubMed: 17381790]
- Fuligni AS, Howes C, Huang Y, Hong SS, & Lara-Cinisomo S (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly*, 27, 198–209. [PubMed: 22665945]
- Gest SD, Holland-Coviello R, Welsh JA, Eicher-Catt D, & Gill S (2006). Language development subcontexts in Head Start classrooms: Distinctive patterns of teacher talk during free play, mealtime and book reading. *Early Education and Development*, 17, 293–315.
- Gerde HK & Powell DR (2009). Teacher education, book-reading practices, and children's language growth across one year of Head Start. *Early Education & Development*, 20, 211–237.
- Girolametto L, Weitzman E, & Greenberg J (2003). Training day care staff to facilitate children's language. *American Journal of Speech-Language Pathology*, 12, 299–311. [PubMed: 12971819]
- Hallgren KA (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8, 23–34. [PubMed: 22833776]
- Hammer CS, Farkas G, & Maczuga S (2010). The language and literacy development of Head Start children: A study using the Family and Child Experiences Survey database. *Language, Speech, and Hearing Services in Schools*, 41, 70–83.
- Hamre BK, Pianta RC, Burchinal M, Field S, LoCasale-Crouch J, Downer JT, ... & Scott-Little C (2012). A course on effective teacher-child interactions effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49, 88–123.
- Harms T, Clifford M & Cryer D (1998) *Early Childhood Environment Rating Scale, Revised edition (ECERS-R)*. New York: Teachers College Press.
- Hart B, & Risley TR (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, Spring. Retrieved September 12, 2007 from http://www.aft.org/pubs-reports/american_educator.
- Hill HC, Charalambous CY, & Kraft MA (2012). When rater reliability is not enough teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41, 56–64.

- Hindman AH, & Wasik BA (2012). Unpacking an effective language and literacy coaching intervention in Head Start. *The Elementary School Journal*, 113, 131–154.
- Hindman AH, & Wasik BA (2013). Vocabulary learning in Head Start: Nature and extent of classroom instruction and its contributions to children's learning. *Journal of School Psychology*, 51, 387–405. [PubMed: 23816231]
- Hoff E (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, 49, 4–14. [PubMed: 22329382]
- 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 programs. *Early Childhood Research Quarterly*, 23, 27–50.
- Huttenlocher J, Vasilyeva M, Cymerman E, & Levine S (2002). Language input and child syntax. *Cognitive Psychology*, 45, 337–374. [PubMed: 12480478]
- Justice LM, Mashburn AJ, Hamre BK, & Pianta RC (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, 23, 51–68. [PubMed: 22773887]
- Kelcey B, McGinn D, & Hill H (2014). Approximate measurement invariance in cross-classified rater-mediated assessments. *Frontiers in Psychology*, 5, 10–3389. [PubMed: 24478742]
- Keys TD, Farkas G, Burchinal MR, Duncan GJ, Vandell DL, Li W, ... & Howes C (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development*, 84, 1171–1190. [PubMed: 23331043]
- Kontos S, & Wilcox-Herzog A (2001). How do education and experience affect teachers of young children? *Research in Review. Young Children*, 56, 85–91.
- Landry SH, Anthony JL, Swank PR, & Monseque-Bailey P (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, 101, 448–465.
- Landry SH, Zucker TA, Taylor HB, Swank PR, Williams JM, Assel M, ... Klein A (2014). Enhancing early child care quality and learning for toddlers at risk: The responsive early childhood program. *Developmental Psychology*, 50, 526–541. [PubMed: 23772822]
- Landry SH, Crawford A, Gunnewig S, & Swank PR (2000). The CIRCLE-Teacher Behavior Rating Scale. Unpublished research instrument.
- La Paro K, Pianta R, Hamre B, & Stuhlman M (2002). Classroom Assessment Scoring System (CLASS). Pre-K version. Charlottesville, VA: UVA.
- LaParo K, Sexton D, & St Snyder P (1998). Program quality characteristics in segregated and inclusive early childhood settings. *Early Childhood Research Quarterly*, 13, 151–167.
- Leonard LB, Weismer SE, Miller CA, Francis DJ, Tomblin JB, & Kail RV (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*, 50, 408–428.
- LoCasale-Crouch J, Konold T, Pianta R, Howes C, Burchinal M, Bryant D, Clifford R, et al. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, 22, 3–17.
- Lonigan CJ, Purpura DJ, Wilson SB, Walker PM, & Clancy-Menchetti J (2013). Evaluating the components of an emergent literacy intervention for preschool children at risk for reading difficulties. *Journal of Experimental Child Psychology*, 114, 111–130. [PubMed: 23073367]
- Marulis LM, Neuman SB (2010). The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Review of Educational Research*, 80, 300–335.
- Massey SL, Pence KL, Justice LM, & Bowles RP (2008). Educators' use of cognitively challenging questions in economically disadvantaged preschool classroom contexts. *Early Education and Development*, 19, 340–360.
- Miller JA, & Bogatova T (2009). Quality improvements in the early care and education workforce: Outcomes and impact of the TEACH Early Childhood® Project. *Evaluation and Program Planning*, 32, 257–277. [PubMed: 19285728]

- Moss CM, & Brookhart SM (2013). A new view of walk-throughs. *Educational Leadership*, 70(7), 42–45.
- Nelson KE, Welsh JA, Trup EMV, & Greenberg MT (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Language*, 31, 164–194.
- Neuman SB, & Dwyer J (2009). Missing in action: Vocabulary instruction in pre-k. *Reading Teacher*, 62, 384–392.
- Neuman SB, Newman EH, & Dwyer J (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly*, 46, 249–272.
- Pentimonti JM, Zucker TA, Justice LM, Petscher Y, Piasta SB, & Kaderavek JN (2012). A standardized tool for assessing the quality of classroom-based shared reading: Systematic Assessment of Book Reading (SABR). *Early Childhood Research Quarterly*, 27, 512–528.
- Phillips BM, Oliver F, & Willis KB (2017). Intensive preschool vocabulary instruction: An initial efficacy trial of media-enhanced instruction. Manuscript in preparation.
- Phillips BM, Weekley MJ, Flynn KS, Augustyn A, Clancy J, Willis KB, & Shiver G (2011). The Classroom Language Environment Observation Scales. unpublished measure. Tallahassee, FL: Authors.
- Phillips D, Gormley W, & Lowenstein A (2009). Inside the pre-K door: Classroom climate and instructional time allocation in Tulsa's pre-K program. *Early Childhood Research Quarterly*, 24, 213–228.
- Phillipsen LC, Burchinal MR, Howes C, & Cryer D (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, 12, 281–303.
- Powell DR, Diamond KE, Burchinal MR, & Koehler MJ (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, 102, 299.
- Quinn JM, Wagner RK, Petscher Y, & Lopez D (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, 86, 159–175. [PubMed: 25201552]
- Ruston HP, & Schwanenflugel PJ (2010). Effects of a conversation intervention on the expressive vocabulary development of prekindergarten children. *Language, Speech, and Hearing Services in Schools*, 41, 303–313.
- Shrout PE, Fleiss JL (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. [PubMed: 18839484]
- Sénéchal M, & LeFevre J (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, 73, 445–460. [PubMed: 11949902]
- Smith MW, & Dickinson DK (1994). Describing oral language opportunities and environments in Head Start and other preschool classrooms. *Early Childhood Research Quarterly*, 9, 345–366.
- Storch SA, & Whitehurst GJ (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38, 934–947. [PubMed: 12428705]
- Turnbull KP, Anthony AB, Justice L, & Bowles R (2009). Preschoolers' exposure to language stimulation in classrooms serving at-risk children: The contribution of group size and activity context. *Early Education and Development*, 20, 53–79.
- vanKleeck A, Vander Woude J, & Hammett L (2006). Fostering literal and inferential language skills in Head Start preschoolers with language impairment using scripted book-sharing discussions. *American Journal of Speech-Language Pathology*, 15, 85–95. [PubMed: 16533095]
- Vescio V, Ross D, & Adams A (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, 24(1), 80–91.
- Vu J, Jeon H-J, & Howes C (2008). Formal education, credential, or both: Early childhood program classroom practices. *Early Education and Development*, 19, 479–504.
- Walsh BA, & Blewitt P (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal*, 33, 273–278.

- 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, 199–209.
- Zucker TA, Cabell SQ, Justice LM, Pentimonti JM, & Kaderavek JN (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, 49, 1425–1439. [PubMed: 23066674]

Table 1.

Sample Items and Internal Consistency Values for CLEOS Subscales and TBRS-P Oral Language Scale

Subscale (Internal Consistency)	Example Items
CLEOS: General Language Environment (.75)	<p>28. Adult speech is clear and articulate, rather than mumbled or dysfluent</p> <p>29. Adult speech to children includes some sophisticated sentence structures (e.g., embedded clauses, conjunctions)</p> <p>32. Adult speech to children includes equal or greater number of questions and comments as directives</p>
CLEOS: Incidental Language Instruction (.78)	<p>37. Models unusual or novel use of typical materials to encourage child comments and interest</p> <p>39. In response to child speech, encourages further utterances by asking follow-up questions or making comments</p> <p>40. Supports (via redirecting and encouraging) peer-to-peer speech such as children asking one another for help, inviting peers to play, etc.</p>
CLEOS: Incidental Vocabulary Instruction (.66)	<p>44. Models specific language by labeling objects and actions in conversation with children</p> <p>46. Uses newly taught vocabulary in conversation with children</p> <p>48. Models use of emotion words by using "I statements" to describe own feelings</p>
CLEOS: Explicit Vocabulary Instruction (.89)	<p>49. Explicitly labels and defines multiple new objects/actions/descriptors</p> <p>50. Provides child-friendly definition when introducing new vocabulary words (e.g., using synonyms/antonyms, semantic links)</p> <p>55. Elicits newly taught labels from children (e.g., "now you say...")</p>
CLEOS: Book Reading Quality Behaviors (.89)	<p>18. Asks children to label objects/actions in illustrations using wh-questions</p> <p>19. Asks feature questions about illustrations (color, shape, function, role in narrative, etc.)</p> <p>22. Encourages child speech by asking questions that relate book to children's own experiences</p>
TBRS-P: Oral Language Subscale (.92)	<p>2. Models for children how to express their ideas in complete sentences</p> <p>7. Encourages children to use new vocabulary words in their speech and writing</p> <p>13. Uses teachable moments as they occur to develop vocabulary and syntax</p>

Table 2.
Descriptive Statistics for CLEOS Language Environment Subscales in Primary and Validity Samples

Observational Subscale	Primary Sample (<i>n</i> = 122)		Validity Sample (<i>n</i> = 43)		Maximum Possible
	Mean (<i>SD</i>)	Range	Mean (<i>SD</i>)	Range	
General Language Environment	5.69 (1.47)	2.00 – 8.00	5.93 (1.34)	3.00 – 8.00	8
General Vocabulary Instruction - Ever Observed	7.23 (1.24)	3.00 – 8.00	7.63 (0.62)	6.00 – 8.00	8
Incidental Language Instruction	2.78 (1.29)	0.00 – 7.00	3.40 (1.06)	1.50 – 6.00	10
Incidental Language Instruction -Ever Observed	6.28 (1.95)	0.00 – 10.00	6.40 (1.68)	3.00 – 10.00	10
Incidental Vocabulary Instruction	1.21 (0.73)	0.00 – 3.00	1.26 (0.78)	0.00 – 3.33	5
Incidental Vocabulary Instruction -Ever Observed	3.01 (1.41)	0.00 – 5.00	2.91 (1.49)	0.00 – 5.00	5
Explicit Vocabulary Instruction	0.98 (0.92)	0.00 – 3.75	0.72 (0.72)	0.00 – 2.50	8
Explicit Vocabulary Instruction - Ever Observed	3.20 (2.66)	0.00 – 8.00	2.37 (2.28)	0.00 – 7.00	8
Book Reading Quality Behaviors- Ever Observed	4.89 (3.84)	0.00 – 13.00	3.67 (4.19)	0.00 – 12.00	13

Note. Variables indicated as “ever observed” represent means for subscales based on occurrence in just one or more classroom setting; those *not* indicated as “ever observed” represent means weighted by the number of observed classroom settings in which the items on the scale were seen to occur.

Table 3.
Correlations among CLEOS Language Environment Subscales and Teacher Characteristics in Primary Sample

	1.	2.	3.	4.	5.	6.	7.	8.
1. GLE	---							
2. ILI	.31 ***	---						
3. IVI	.32 ***	.52 ***	---					
4. EVI	.27 **	.37 ***	.57 ***	---				
5. BR Quality	.12	.14	.23 *	.34 ***	---			
6. Years Teaching Preschool (Lead)	.20 *	.26 **	.23 *	.10	.03	---		
7. Teacher Education (Lead)	.12	-.03	.00	.08	.03	-.03	---	
8. Teacher Age (Lead)	.04	.12	.24 *	.24 **	.16	.50 ***	.06	---

Note. GLE = General Language Environment; ILI = Incidental Language Instruction; IVI = Incidental Vocabulary Instruction; EVI = Explicit Vocabulary Instruction; BR = Book Reading. *N* = 122; sample size for relations with age = 114; for relations with lead teacher education = 117.

* $p < .05$,

** $p < .01$,

*** $p < .001$.

Table 4.

Descriptive Statistics for TBRs - Plus in Validity Sample

Observational Subscale	Mean (SD)	Range	Maximum Possible
TBRs-P Classroom Community	10.69 (2.41)	5.00 – 16.00	16
TBRs-P Teacher Sensitivity	17.54 (5.30)	8.00 – 27.50	28
TBRs-P Teacher Discipline	8.77 (2.14)	5.00 – 11.00	13
TBRs-P Centers	21.43 (6.60)	0.00 – 34.00	34
TBRs-P Book Reading	12.29 (8.73)	0.00– 28.00	28
TBRs-P Oral Language	26.97 (8.48)	14.50 – 45.50	46
TBRs-P Print Knowledge	11.80 (6.60)	0.00 – 26.50	28
TBRs-PPhonological Awareness	9.21 (11.25)	0.00 – 32.00	51.50
TBRs-P Writing	9.85 (5.14)	0.00 – 22.00	24.50
TBRs-P Math	9.98 (6.68)	0.00 – 23.00	28

Note. TBRs-P = Teacher Behavior Rating Scale-Plus.

Table 5.

Concurrent Validity Correlations between CLEOS and TBRs-Plus Subscales

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. CLEOS:GLE	---														
2. CLEOS:ILI	.37**	---													
3. CLEOS:IVI	.51***	.61***	---												
4. CLEOS:EVI	.32*	.38*	.50**	---											
5. CLEOS:BRQ	.17	.13	.22	.51***	---										
6. TBRs-P:Com.	.58***	.28	.36*	.36*	.23	---									
7. TBRs-P:Sens.	.29	.43**	.27	.19	.44**	.49**	---								
8. TBRs-P:Disc.	.23	.35*	.33*	.21	.19	.52***	.72***	---							
9. TBRs-P:Curs.	.41*	.34*	.33*	.32*	.25	.77***	.43**	.47**	---						
10. TBRs-P:BR	.10	.08	-.03	.20	.48**	.28	.46**	.32*	.36*	---					
11. TBRs-P:OL	.15	.53***	.43**	.42**	.46**	.39*	.52***	.52***	.46**	.38*	---				
12. TBRs-P:PK	.30	.20	-.13	.28	.14	.51***	.27	.22	.41**	.25	.20	---			
13. TBRs-P:PA	.26	.23	-.11	.23	.17	.31*	.27	.15	.16	.25	.38*	.62***	---		
14. TBRs-P:Wt.	.26	.36*	.37**	.40*	.11	.52***	.28	.43**	.54***	.16	.54***	.39*	.20	---	
15. TBRs-P:Mt.	.18	.23	.03	.25	.32*	.34*	.34*	.17	.33*	.27	.37*	.51**	.41**	.16	---

Note. CLEOS = Classroom Language Environment Observational Scales; GLE = General Language Environment; ILI = Incidental Language Instruction; IVI = Incidental Vocabulary Instruction; EVI = Explicit Vocabulary Instruction; BRQ = Book Reading Quality; TBRs-P = Teacher Behavior Rating Scale-Plus; Com. = Classroom Community; Sens. = Sensitivity; Disc. = Discipline; Ctrs. = Centers; BR = Book Reading; OL = Oral Language; = PK = Print Knowledge; PA = Phonological Awareness; Wt. = Writing; Mt. = Math.

* $p < .05$,

** $p < .01$,

*** $p < .001$.

Table 6.

CLEOS Weighted Means, (Standard Deviations) and [Percentage of Classrooms Observed in Context] by Scale and by Context in Primary Sample

Selected Classrooms where Context was Observed						
Scale (Maximum Possible)	Circle Time [94%]	Centers [82%]	Small Group [53%]	Gross Motor [68%]	Meals [57%]	Transitions [97%]
GLE (8)	6.56 (1.74)	6.48 (2.00)	5.64 (3.00)	5.01 (2.36)	4.56 (2.65)	5.18 (2.10)
ILI (10)	2.87 (1.92)	4.59 (1.99)	3.50 (2.24)	2.51 (1.98)	1.66 (1.80)	1.31 (1.44)
IVI (5)	1.97 (1.46)	1.82 (1.41)	2.25 (1.44)	0.64 (0.89)	0.50 (0.81)	0.48 (0.74)
EVI (8)	2.69 (2.47)	1.25 (1.93)	1.42 (2.05)	0.17 (0.46)	0.13 (0.38)	0.13 (0.50)
Full Sample						
Scale (Maximum Possible)	Circle Time	Centers	Small Group	Gross Motor	Meals	Transitions
GLE (8)	6.24 _a (2.21)	5.43 _{ac} (3.02)	3.34 _b (3.64)	3.48 _b (3.01)	2.79 _b (3.02)	5.01 _c (2.26)
ILI (10)	2.71 _a (1.98)	3.76 _b (2.53)	1.86 _c (2.38)	1.71 _c (2.01)	0.95 _d (1.59)	1.27 _{cd} (1.44)
IVI (5)	1.85 _a (1.49)	1.49 _{ab} (1.46)	1.18 _b (1.53)	0.43 _c (0.79)	0.29 _c (0.66)	0.46 _c (0.73)
EVI (8)	2.53 _a (2.48)	1.03 _b (1.81)	0.75 _b (1.64)	0.12 _c (0.40)	0.08 _c (0.29)	0.23 _c (0.49)

Note. GLE = General Language Environment; ILI = Incidental Language Instruction; IVI = Incidental Vocabulary Instruction; EVI = Explicit Vocabulary Instruction. $N = 122$ for lower panel and varied for upper panel; [%] represents proportion of the full sample included. Means within rows with different subscript letters differed from one another at $p < .05$.