



Review

Using early care and education quality measures with dual language learners: A review of the research[☆]

Ellen Peisner-Feinberg^{a,*}, Virginia Buysse^a, Allison Fuligni^b, Margaret Burchinal^a,
Linda Espinosa^c, Tamara Halle^d, Dina C. Castro^e

^a Frank Porter Graham Child Development Institute, The University of North Carolina at Chapel Hill, United States

^b California State University-Los Angeles, United States

^c University of Missouri, United States

^d Child Trends, United States

^e Arizona State University, United States

ARTICLE INFO

Article history:

Available online 10 May 2014

Keywords:

Dual language learners (DLLs)
Early childhood
Early care and education (ECE)
Quality measures
Child care quality

ABSTRACT

As the number of young dual language learners (DLLs) in early care and education (ECE) programs is increasing, it is critical to examine how well measures of the quality of practices in these settings reflect the needs of the diverse groups of children being served. This review of the research literature addressed these questions for ECE settings serving children birth-five: whether quality differs for settings serving high proportions of DLLs compared to typical samples, whether existing quality measures exhibit similar psychometric characteristics and associations with child outcomes in settings serving DLL and non-DLL children, and whether DLL-specific measures perform differently than general measures of quality. The search procedure produced 10 research studies that met the criteria for inclusion in the review out of approximately 3800 that were found initially and more than 300 that were reviewed. These studies included 10 out of 46 identified measures of ECE quality, including both general and DLL-specific measures. Findings suggested that widely used general ECE quality measures function similarly for DLLs compared to typical populations with regard to overall quality, psychometric characteristics, and child outcomes. Further research is needed to broaden the knowledge base for a wider variety of measures, beyond center-based settings, and beyond Spanish-speaking DLLs, as well as to enhance methodological approaches. There appear to be potential research opportunities through numerous existing studies that included DLL populations but had not analyzed their data in regard to these groups.

© 2014 Elsevier Inc. All rights reserved.

Contents

1. Introduction	787
1.1. Changing demography of ECE in the U.S.	787
1.2. Definition of ECE quality	787
1.3. Measures of ECE quality	787
1.4. Challenges in measuring quality of ECE for DLLs	788
1.5. Current review	789
2. Method	789
2.1. Search parameters	789
2.2. Search process	790

[☆] Author note: This paper was completed as part of the research review activities of the Center for Early Care and Education Research-Dual Language Learners (CECER-DLL) funded by the Office of Planning, Research, & Evaluation (OPRE) in the Administration for Children & Families (ACF), U.S. Department of Health and Human Services (DHHS).

* Corresponding author at: FPG Child Development Institute, The University of North Carolina at Chapel Hill, 105 Smith Level Road, Chapel Hill, NC 27516, United States. Tel.: +1 919 962 7354; fax: +1 919 966 7532.

E-mail address: epeisner.feinberg@unc.edu (E. Peisner-Feinberg).

2.3. Measures included in the review.....	790
2.4. Coding procedures.....	792
3. Results.....	792
3.1. Quality of settings.....	792
3.2. Similarity of psychometric characteristics.....	792
3.3. Associations between quality scores and child outcomes.....	792
3.4. DLL-specific measures.....	792
4. Discussion.....	798
4.1. Conclusions.....	798
4.2. Methodological considerations.....	799
4.3. Future research directions.....	800
References.....	801

1. Introduction

In the last two decades, the United States has experienced a demographic shift that has dramatically increased the linguistic diversity among young children and families participating in early care and education (ECE) programs. This changing population has resulted in a growing awareness of the unique care and education needs of young dual language learners (DLLs), children who are continuing to acquire their first or home language as they also learn English as a second language. Although as a group, young DLLs share the common characteristic of developing skills in two languages, as individuals, they differ in terms of a variety of characteristics, including home language; level of language proficiency in each language; and a host of family background factors, including culture, race, ethnicity, country of origin, immigration status, poverty level, and family composition. Given the growing numbers of young DLLs, coupled with policies in many states that promote serving these children in ECE settings, it is important to understand the current state of knowledge regarding the quality of care being received by these children and how quality is being measured with regard to this population. Therefore, the purpose of this review was to summarize the research on ECE quality for DLL children for both of these aspects.

1.1. Changing demography of ECE in the U.S

Recent data have shown that more than 30% of the children enrolled in Head Start programs speak a language other than English at home (Office of Head Start, 2009). Among DLLs, Latinos are one of the fastest-growing groups of young children in the U.S., with accordingly more research available for this population as well. As a group, Latinos represent at least 1 in 12 children in about half of the states, although there is wide variation in the proportion across states (Hernandez, Denton, & Macartney, 2007). Although some studies suggest that Latino families are less likely to use center-based care than European-American families (Magnuson, Meyers, & Waldfogel, 2007), other studies have found that use varies depending on children's age and family heritage (Winsler et al., 2014). According to Census 2000 data, substantial numbers attend center-based care, estimated at more than one-quarter of all 3-year-old and over half of all 4-year-old Latinos (Hernandez et al., 2007). Increased parental workforce participation as well as the availability of publicly funded programs such as Head Start and public pre-kindergarten programs have made enrollment in center-based early education settings increasingly common for children from a variety of linguistically diverse backgrounds. Further, studies suggest that DLLs often are at greater risk in terms of school readiness and that participation in such ECE programs can help better prepare these children for school (e.g., Fuligni & Howes, 2011; Gormley, 2008; Gormley, Gayer, Phillips, & Dawson, 2005; Peisner-Feinberg & Schaaf, 2009; Peisner-Feinberg, Schaaf, LaForett, Hildebrandt, & Sideris, 2014). Given this research

along with the demographic shifts, it is important to examine the quality of ECE for DLLs and to determine the best ways to measure the quality of these experiences for this population.

1.2. Definition of ECE quality

Quality traditionally has been defined and measured according to two basic aspects—structural features and process quality (Peisner-Feinberg & Yazejian, 2010; Vandell & Wolfe, 2000). Structural features are the more basic, easily and objectively measured aspects of quality, such as teacher education and credentials, staff-child ratios, and group size, which often are represented in licensing regulations and other program guidelines, but also may be used in research. Process quality represents the direct experiences of children in early childhood settings, and therefore requires more in-depth observation and the use of standardized instruments to measure. Key aspects of process quality include the sensitivity and responsiveness of caregivers, the materials available for learning, and the instructional interactions with the teacher. The focus of the present review was on instruments related to measuring process quality in ECE settings.

1.3. Measures of ECE quality

A number of published instruments for measuring the quality of ECE practices are available, with research showing evidence of their reliability across raters and time and their validity in terms of associations with other quality measures and with child outcomes (see Halle, Whittaker, & Anderson, 2010, for a compendium of measures). These measures typically were developed for research purposes to capture a range of practices across four main categories, from measures of global quality in center-based or home-based settings to more specific aspects of practice related to teacher-child interactions, particular content areas such as language and literacy or math, and bilingual supports for DLLs. One set of widely used quality measures, the Environment Rating Scales (ERS), focus on the global ECE environment—the extent to which children have access to age-appropriate materials and the nature of caregiver interactions and supervision. ERS measures include the *Early Childhood Environment Rating Scale-Revised* (ECERS-R; Harms, Clifford, & Cryer, 2005) for center-based preschool classrooms, the *Infant/Toddler Environment Rating Scale-Revised* (ITERS-R; Harms, Cryer, & Clifford, 2003) for center-based infant and toddler classrooms, and the *Family Child Care Environment Rating Scale-Revised* (FCCRS-R; Harms, Cryer, & Clifford, 2007) for home-based care settings. A second set of measures have focused more exclusively on the frequency and quality of teacher-child interactions. One of the most widely used measures, the *Classroom Assessment Scoring System: Pre-K* (CLASS; Pianta, La Paro, & Hamre, 2008), measures the frequency and quality of interactions between teachers and children in terms of emotional and instructional support and classroom management. Other quality measures focus on

caregiver sensitivity and responsiveness in interactions with children, such as the *Caregiver Interaction Scale* (CIS; Arnett, 1989) and the *Observational Record of the Caregiving Environment* (ORCE; NICHD Early Child Care Research Network, 1996). A third set of measures examines quality relative to a specific content area, such as the *Early Language and Literacy Classroom Observation Pre-K* (ELLCO Pre-K; Smith, Brady, & Anastasopoulos, 2008) which focuses on the language and literacy environment. Finally, a few quality measures have been explicitly designed for use in settings with DLL children. In the previous three types of measures, although DLLs are not the focus, they often include some items such as those related to cultural and linguistic diversity within the classroom. In contrast, the DLL-specific measures examine practices of particular importance for the early learning needs of DLLs, such as the use of curriculum and instructional practices that are culturally and linguistically responsive; positive teacher–child interactions and opportunities for peer interactions believed to promote DLLs’ social-emotional development; and strategies related to language development for DLLs, including explicit vocabulary instruction, strategic use of a child’s first language (L1), and frequent assessment of first (L1) and second language (L2) development (Castro, Espinosa, & Páez, 2011). In some cases, these practices may be unique to DLLs, such as those involving L1 instruction and assessment or the use of culturally and linguistically responsive practices; in other cases, these practices may be similar to practices for monolinguals, such as explicit vocabulary instruction, positive teacher–child interactions, and peer-assisted instruction, but even more critical to supporting development for DLLs who are learning English (L2) in typical ECE settings in this country (Castro, Peisner-Feinberg, Buysse, & Gillanders, 2010).

The types of measurement tools that are available, their applicability to different types of ECE settings, and their relevance for various populations served affect the evidence that can be used to inform decisions within research, policy, and practice. What is known about the level of quality in ECE settings and the ways in which it can be measured has influenced the conclusions that can be drawn from research and program evaluation studies aimed toward examining the quality of care, the ability to define and provide professional development geared toward implementing high-quality practices, and the extent to which policies are effective in promoting high-quality care (Peisner-Feinberg & Yazejian, 2010). Over the past few decades, attention has been paid to measuring quality for a variety of purposes, including monitoring related to child care licensing or other regulatory purposes, professional development efforts, and research and evaluation studies. For example, the monitoring system for Head Start relies on the CLASS, and the Quality Rating and Improvement Systems for many states include the ERS or CLASS. However, most of these uses of quality measures have been based on results for all children, without considering DLLs separately to determine whether the results are equally applicable for these children. Therefore, it is important to systematically examine the state of knowledge regarding the use of quality measures with DLLs specifically, in order to inform research, practice, and policy decisions for this population.

1.4. Challenges in measuring quality of ECE for DLLs

Similarly to most children in this country, it is likely that young DLLs do not experience high-quality ECE during the preschool years. Several large-scale studies have indicated that the general quality of ECE programs in the U.S. is in the medium range, meaning that they meet the basic health and safety needs of children but do not offer sufficient educational experiences to promote optimal development (Burchinal, Kainz, & Cai, 2011; CQO Study Team, 1995; Early et al., 2005; Mashburn et al., 2008; NICHD Early Child Care Research Network, 2000; Whitebook, Howes, & Phillips, 1989;

Zill, Sorongon, Kim, Clark, & Woolverton, 2006). Scores tend to be lower (averaging in the low-to-medium range vs the medium-to-high range) for infants and toddlers than for preschoolers and for measures of instructional quality than for caregiver sensitivity and responsiveness. This is concerning because widely used quality measures show consistent, albeit modest, associations with school readiness skills (Burchinal et al., 2009, 2011; Zaslow et al., 2010), with some evidence that those associations may be stronger for children from low-income families or families in which parents have less education (McCartney, Dearing, Taylor, & Bub, 2007; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Watamura, Phillips, Morrissey, McCartney, & Bub, 2011). Given that children from immigrant families are more likely to live in poverty and have parents with lower education (Karoly & Gonzalez, 2011; National Center for Children in Poverty, 2007), these findings are of particular relevance for young DLLs.

Both the changing demography of the U.S. and the growing international interest in quality measurement have led to increasing attention to whether existing measures of quality adequately capture key features of early experiences for children from diverse racial, ethnic, and linguistic backgrounds. In particular, the increasing number of DLLs presents potential challenges for the ECE field, in terms of the best ways to provide high-quality experiences as well as how to measure them. The research using these types of tools has focused both on measuring the quality of ECE settings as well as on the associations between quality and children’s learning and developmental outcomes. However, there is little known about the extent to which young DLLs are experiencing high-quality ECE or whether the linguistic diversity of the ECE setting is associated with quality. The results of a recent review of the literature on the effects of ECE programs for DLLs suggest that although participation in widely available and well-regulated programs such as public pre-k and Head Start may be beneficial for DLLs, little is known about the specific factors that contribute to their positive effects (Buysse, Peisner-Feinberg, Paez, Hammer, & Knowles, 2014). Other research has suggested that substantial increases in both program participation and quality could have strong effects on reducing school readiness gaps for these children (Magnuson & Waldfogel, 2005). Further support for this perspective is found from studies of several statewide pre-kindergarten programs which have shown that program participation has even stronger effects for Latino DLLs than for other children (Gormley et al., 2005; Peisner-Feinberg & Schaaf, 2009).

Although a number of studies have been conducted examining the effects of different types of curricular and instructional practices, it has not been possible to disentangle these effects from those of the language of instruction (English-only vs bilingual) within the research (Buysse et al., 2014). Therefore, it is not known whether there are specific aspects of quality that may be differentially beneficial for DLLs, although positive effects have been found both for interventions that used English as the language of instruction and for those that incorporated the home language in language and literacy curricula and instructional strategies (e.g., Bernhard et al., 2006; Buysse, Castro, & Peisner-Feinberg, 2010; Collins, 2010; Farver, Lonigan, & Eppe, 2009; Lugo-Neris, Jackson, & Goldstein, 2010; Roberts & Neal, 2004; Silverman & Hines, 2009). Findings from research with school-age children provide corresponding support for curricular approaches focused on language and literacy instruction along with high-quality instruction in general (August & Shanahan, 2006; Cheung & Slavin, 2012). The dimensions addressed in existing measures of ECE quality, such as aspects of the physical environment, instructional activities in various domains, and interactions among children and adults generally are perceived to be important for all children. However, the associations between quality ratings based on these widely used measures and developmental outcomes for DLLs have not been systematically studied.

One key question is whether ECE quality generally looks the same for settings serving DLLs as for the general population. Although there have been a number of studies examining ECE quality in recent years, the level of quality experienced by DLLs is not clearly known. Another outstanding issue with regard to measuring the quality of ECE practices is whether the existing tools function similarly for programs serving DLLs. Key questions to address this issue include whether the associations of quality with teacher and classroom characteristics or with child outcomes are the same for DLLs as for monolinguals. Further, in order to be able to take full advantage of high-quality programs, there are likely to be additional features of early learning environments that are needed by DLLs, which are addressed by DLL-specific measures of quality, but not typically included in general measures of quality. Therefore, another key question is whether general measures of ECE quality and DLL-specific measures capture different aspects of children's experiences, perhaps necessitating the use of both types of measures in order to adequately determine the quality of programs serving linguistically diverse populations. Thus, from a research perspective, it is important to examine the quality and influence of ECE experiences for DLLs in different types of settings, the performance of measures used to evaluate the quality of ECE for DLL populations, and the extent to which there are measurement issues that are unique to DLLs.

1.5. Current review

Given the above issues, the research team undertook a critical review of the research literature to examine what is known about the use of ECE process quality measures in studies of DLLs aged birth-five in center-based and/or home-based settings. First, a comprehensive set of relevant measures was defined to set the parameters for the search, identified through a variety of expert sources (as described below). The search focused on studies that used publicly available, standardized measures of quality which could be gathered across different studies (as opposed to project-specific measures designed for unique use within a single study). The studies had to report results specifically for DLLs in order to address the topic of the review. Studies could be included if they reported results in at least one of the following ways: the entire sample consisted primarily of DLLs, they included comparisons among settings serving differing proportions of DLLs, or they included comparisons between settings with and without DLL populations.

The purpose of the review was to examine what is known about ECE quality for DLL children, including the levels of quality experienced by DLL children and how existing measures of quality function with regard to DLL populations. Based on a systematic analysis of research studies published from 2000 to 2011 in peer-reviewed journals, this review specifically addressed the following questions:

- Do classrooms serving high proportions of DLL children differ from typical samples based on ECE measures of quality?
- Do ECE quality measures exhibit similar psychometric characteristics when used in classrooms serving DLL vs non-DLL samples?
- Do ECE quality measures exhibit similar associations with children's developmental outcomes for DLL and monolingual children?
- Do DLL-specific measures perform differently than general measures of ECE quality?

The results of this review would provide a summary of the state of knowledge regarding ECE quality and the use of quality measures for DLL populations, as well as help inform future research efforts

based on the patterns of findings and methodological strengths and weaknesses within the current literature.

2. Method

This review was conducted as one of a set of research reviews undertaken by the Center for Early Care and Education Research-Dual Language Learners (CECER-DLL) to summarize the current state of knowledge regarding issues of measurement and development of DLLs using common search procedures. The research team consisted of seven individuals with expertise in ECE commissioned by the CECER-DLL, in conjunction with a technical work group (TWG) of approximately 30 members affiliated with the CECER-DLL, to ensure the thoroughness and accuracy of the review, including defining specific criteria for this review (e.g., determining the measures that would be reviewed, determining inclusion and exclusion criteria), verifying search terms and procedures, and reviewing the final list of articles and results of the review. For the purposes of reviewing these studies, it is important to note that we were limited to reporting only the information that was available in the published articles, which may not always have provided complete information about the study methodologies.

2.1. Search parameters

A set of parameters was established to guide the search of existing empirical literature regarding the use of ECE process quality measures in the U.S. in relation to young DLLs in ECE programs. The search parameters for this review were defined by the CECER-DLL and included the following:

- A set of publicly available, standardized measures of process quality in ECE settings (i.e., settings serving children prior to entry into kindergarten) was defined.
- A broad definition of DLL was used, which reflects the various ways that samples have been defined by individual studies in the literature, including home language use (according to parent or teacher report), direct measures of language proficiency (based on individual assessments), or solely based on ethnicity (e.g., Latino). Although we acknowledge that these definitions are not equivalent, they are reflective of the current state of the literature. The term DLL is used in this review to refer to the various samples as defined by the individual studies.
- The search included published peer-reviewed journal articles from 2000 to 2011 (note that one article available online pre-publication and slated for a 2011 publication date was actually published in the first issue in 2012).

Individual studies had to meet the two following criteria:

- Studies that included results on one or more of the quality measures within ECE settings.
- Studies that examined the quality data for DLL populations and/or that examined the associations between quality measures and outcomes for DLL children birth to 5-years-old, either exclusively or in sub-group analyses.

Studies were included only if they examined DLL populations, either as the sole focus, based on the proportion of DLLs, or as a separate comparison group. For example, the study sample could primarily represent DLLs, the study could include comparisons among settings serving differing proportions of DLLs, or the study could include comparisons between settings with and without DLL populations. Eligible studies could provide information solely about setting quality or also could analyze whether setting quality was

associated with other variables or with DLL children's outcomes, as long as the results could be examined specifically for DLL populations.

Search terms to guide the search of the literature databases for relevant studies were defined by the research team in accordance with CECER-DLL guidelines. Two sets of search terms were defined: (1) terms for measures (names and abbreviations), and (2) terms for DLLs, based on a standard list defined by the CECER-DLL representing language status (e.g., bilingual, English language learner, dual language learner, English as a second language, Spanish, Mandarin, Chinese, Spanish-speaking, Chinese-speaking) and ethnicity or immigration status (e.g., Latino, Asian, immigrant, migrant).

2.2. Search process

First, the research team compiled a comprehensive list of measures for assessing quality in preschool settings. Interim and final lists were reviewed by the TWG, who were asked to nominate additional measures and sources. A variety of key sources were used, including recently completed compendia and books summarizing quality assessment measures (Castro et al., 2011; Halle & Vick, 2007; Halle et al., 2010; National Research Council, 2008). In addition, active researchers conducting work in this area were contacted for recommendations for additions to the list in order to ensure that any recently developed measures were included. The measures included in the review had to meet the following criteria: (1) the measure focused on measuring some aspect of the quality of care and education settings for children birth-five in center-based and/or home-based settings, (2) the measure was publicly available for use in research (commercially or otherwise), and (3) the tool was developed for general use (i.e., not a project-specific measure designed for unique use within a single study) so that it potentially could have applicability across multiple research studies. Although we acknowledge that project-specific measures sometimes may be the precursors to later standardized versions, the review hinged on identifying a comprehensive set of measures for the search process. It would not have been possible to obtain a comprehensive list of project-specific measures, as there are no authoritative sources for such measures (in contrast to the general use measures), nor would it have been consistent with the defined focus of the review on publicly available measures. This process yielded a final list of 46 measures, including both general (39) and DLL-specific (7) quality measures. (See Table 1 for a list of instruments; note that for measures with more than one published version, the search was conducted for all versions, but only the reference for the most recent version is included in the table.)

Next, we conducted systematic searches for relevant research studies in the published literature. An exhaustive search was conducted for each measure using key databases: Academic Search Premier, PsycArticles, PsycINFO, ERIC, and Google Scholar. Searches included full-text search, titles, keywords, and advanced search options; utilized the "find similar articles" feature; and cross-checked the reference lists from eligible studies for additional articles. The search process incorporated two levels. At the first level, an initial search was conducted using only the search terms for measure names. In cases where the search for a particular measure generated 15 or fewer articles, each article was reviewed to determine whether the study met the two criteria described above for inclusion in the critical review. In cases where more than 15 articles were found, a second level search was conducted, using the complete set of DLL search terms in conjunction with the measure terms, and then all articles found through the refined search were reviewed to determine whether the studies could be included in the critical review based on the criteria described above. These procedures yielded approximately 3800 articles through the first

level search (by measure names) across all search engines including duplications, with this number reduced to approximately 1600 articles with the second level search (adding DLL search terms) including duplications. In the end, this set of procedures resulted in over 300 unduplicated articles that were examined to determine whether the studies met the criteria for inclusion in the review. Three members of the research team screened all references for relevance, reviewed abstracts and full-text articles, and identified a final set of 10 articles as relevant for the review. Studies were excluded from the review because they did not meet the basic criteria of presenting results relevant to DLLs—either they did not analyze their results separately for DLL populations or they did not include defined DLL samples in the research studies themselves, despite initially having been identified through the search process. In addition, the remaining members of the research team and the TWG were asked to review the final set of 10 articles and to nominate additional articles, but no further articles meeting the review criteria were identified.

2.3. Measures included in the review

The 10 studies meeting the review criteria included 10 different quality measures. It is important to note that some of these studies included more than one measure. A brief description of each measure is listed below: (a) *Caregiver Interaction Scale* (CIS; Arnett, 1989), measures the sensitivity of caregiver-child interactions (two studies); (b) *Classroom Assessment Scoring System* (CLASS; Pianta et al., 2008), measures teacher-child interactions around emotional support, classroom organization, and instructional support (one study); (c) *Emerging Academics Snapshot* (Snapshot; Ritchie, Howes, Kraft-Sayre, & Weiser, 2001) measures exposure to instruction, engagement in academic activities, and adult involvement (two studies); note that additional studies used only the *Adult Involvement Scale* (AIS; Howes & Stewart, 1987) portion of this measure (four studies); (d) *Early Childhood Environment Rating Scale* and *Revised Edition* (ECERS; Harms & Clifford, 1980; ECERS-R; Harms et al., 2005), measures process quality in center-based settings, including health and safety, relationships, and opportunities for learning (five studies); (e) *Early Language and Literacy Classroom Observation* (ELLCO; Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002) measures language and literacy practices around language environment, books and reading, and print and early writing, along with general classroom environment (two studies); (f) *Early Language and Literacy Classroom Observation Addendum for English Language Learners* (ELLCO-A; Castro, 2005) measures language and literacy practices for DLLs based on the ELLCO (one study); (g) *Family Day Care Rating Scale* (FDCRS; Harms & Clifford, 1989) measures process quality in family child care homes, including health and safety, relationships, and opportunities for learning (one study); (h) *Observational Record of the Caregiving Environment* (ORCE; NICHD Early Child Care Research Network, 1996) measures the quality of caregiver-child interactions from 6 to 54 months, along with structural measures (two studies); (i) *Supports for Early Literacy Assessment* (SELA; Smith et al., 2001) measures the quality of the literacy environment and instruction (one study); (j) *Supports for English Language Learners Classroom Assessment* (SELLCA; National Institute for Early Education Research, 2005) measures the quality of supports for first language development for DLLs (one study). For two of the measures (ELLCO and FDCRS), there have been newer versions developed since these studies were conducted, as indicated in Table 1. Note that most of these measures were designed to measure ECE practices for the general population, with a focus on English-language instruction and interactions (CIS, CLASS, ECERS/ECERS-R, ELLCO, FDCRS, ORCE, SELA, Snapshot/AIS); only two of the measures were designed to measure ECE practices

Table 1
Early care and education quality measures.

Name of measure	Abbreviation	Reference
DLL-specific measures		
Bilingual Teacher Behavior Rating Scale	B-TBRS	Solari, Landry, Crawford, Gunnewig and Swank (2009)
Classroom Assessment of Supports for Emergent Bilingual Acquisition	CASEBA	Freedson, Figueras-Daniel, and Frede (2009)
Early Language and Literacy Classroom Observation Addendum for English Language Learners	ELLCO-A	Castro (2005)
Measures of Early Language and Literacy Environment Language Interaction Snapshot	ELLE	Mathematica Policy Research (2010)
Observation Measures of Language and Literacy Instruction:	LISn	Atkins-Burnett, Sprachman, and Caspe (2010)
Quality Rating of Language and Literacy Instruction/ Classroom Literacy Opportunities Checklist/ Snapshot	OMLIT: OMLIT-QUILL/ OMLIT-CLOC/ OMLIT-Snapshot	Goodson, Layzer, Smith, and Rimdzius (2006)
Supports for English Language Learners Classroom Assessment	SELLCA	National Institute for Early Education Research (2005)
General measures		
Assessment Profile for Early Childhood Programs	APECP	Abbott-Shim and Sibley (1992a)
Assessment Profile for Family Child Care Homes	APFCCH	Abbott-Shim and Sibley (1992b)
Caregiver Interaction Scale	CIS, Arnett Scale	Arnett (1989)
Caregiver Observation Form and Scale	COFAS	Fiene (1984)
Child Care Assessment Tool for Relatives	CCAT-R	Porter, Rice, and Rivera (2006)
The Child Care HOME Inventories	CC-HOME	Bradley, Caldwell, and Corwyn (2003)
Child Caregiver Interaction Scale	CCIS	Carl (2007)
The Child–Caregiver Observation System	C-COS	Boller, Sprachman, & the Early Head Start Research Consortium (1998)
Child Development Program Evaluation Scale	CDPES	Fiene (1984)
Child/Home Early Language and Literacy Observation	CHELLO	Neuman, Dwyer, and Koh (2007)
Classroom Assessment Scoring System: Pre-K	CLASS	Pianta, La Paro, and Hamre (2008)
Classroom Assessment Scoring System: Toddler Version	CLASS Toddler	Pianta, La Paro, and Hamre (2008)
Classroom CIRCLE: Classroom Code for Interactive Recording of Children's Learning Environments	CIRCLE	Atwater, Lee, Montagna, Reynolds, and Tapia (2009)
Classroom Language and Literacy Environment Observation	CLEO	Holland-Coviello (2005)
Classroom Observation of Early Mathematics Environment and Teaching	COEMET	Sarama and Clements (2007)
Classroom Practices Inventory	CPI	Hyson, Hirsh-Pasek, and Rescorla (1990)
Early Childhood Classroom Observation Measure	ECCOM	Stipek and Byler (2004)
Early Childhood Environment Rating Scale/Early Childhood Environment Rating Scale-Revised	ECERS/ECERS-R	Harms, Clifford, and Cryer (2005)
Early Childhood Environment Rating Scale-Extension	ECERS-E	Sylva, Siraj-Blatchford, and Taggart (2003)
Early Language and Literacy Classroom Observation/ Early Language and Literacy Classroom Observation Pre-Kindergarten	ELLCO/ELLCO Pre-K	Smith, Brady, and Anastopoulos (2008)
Early Literacy Observation Tool	E-LOT	Grehan and Smith (2004)
Emerging Academic Snapshot/Adult Involvement Scale	Snapshot/AIS	Ritchie, Howes, Kraft-Sayre, and Weiser (2001)
Emlen Scales: A Packet of Scales for Measuring the Quality of Child Care From a Parent's Point of View	Emlen Scales	Emlen, Koren, and Schultze (2000)
Family Day Care Rating Scale/Family Child Care Environment Rating Scale-Revised.	FDCRS/FCCERS-R	Harms, Cryer, and Clifford (2007)
Individualized Classroom Assessment Scoring System	inCLASS	Downer, Booren, Lima, Luckner, & Pianta, 2010
Infant/Toddler Environment Rating Scale/ Infant/Toddler Environment Rating Scale-Revised.	ITERS/ITERS-R	Harms, Cryer, and Clifford (2003)
Observational Record of the Caregiving Environment	ORCE	NICHD Early Child Care Research Network (1996)
The Preschool Classroom Implementation Rating Scale	PCI	Frede and Miller (1990)
Preschool Mental Health Climate Scale	PMHCS	Gilliam (2008)
Preschool Program Quality Assessment Instrument/Preschool Program Quality Assessment Instrument, 2nd Edition	PQA	High/Scope Educational Research Foundation (2003)
Preschool Classroom Mathematics Inventory/Preschool Rating Instrument for Science and Mathematics	PCMI/PRISM	Stevenson-Boyd, Brenneman, Frede, and Weber (2008)
Program for Infant/Toddler Care Program Assessment Rating Scale	PITC-PARS	Kriener-Althen and Mangione (2010)
Quality of Early Childhood Care Settings: Caregiver Rating Scale	QUEST	Goodson, Layzer, and Layzer (2005)
The Ramey Observation of Learning Essentials	ROLE	Ramey and Ramey (2002)
Supports for Early Literacy Assessment	SELA	Smith, Davidson, Weisenfeld, and Katsaros (2001)
Supports for Social-Emotional Growth Assessment	SSEGA	Smith (2004)
Teacher Behavior Rating Scale	TBRS	Landry, Crawford, Gunnewig, and Swank (2001)
Teacher Instructional Engagement Scale	TIES	Dickinson (2008)
Teaching Pyramid Observation Tool for Preschool Classrooms	TPOT	Fox, Hemmeter, and Snyder (2014)

specifically related to DLLs in the area of L1 supports (ELLCO-A, SELLCA).

2.4. Coding procedures

Each of the 10 articles was reviewed by three members of the research team, with a standard set of information about each study extracted and entered into an excel database to serve as the basis for analysis. Key study elements defined in the review guidelines were coded and verified, including the purpose and design, participant and setting characteristics, definitions and methods of determination of DLL status, constructs and measures, and statistical approaches and results. Consensus was reached for all study elements. This common set of information provided the key elements for the summary table as well. In addition, the remaining members of the research team were asked to confirm the correspondence between the summary table and the results.

3. Results

Table 2 summarizes key information for each of the 10 studies, including the original study purpose, relevance for the present review with regard to ECE quality measures and DLLs, study design, sample characteristics (including how DLLs were determined), measures, and results. Overall results from this review regarding findings about the use of quality measures for DLLs are presented below, including commonalities and differences in the quality of ECE settings, the psychometric characteristics of the measures used, the associations with child outcomes, and the performance of general vs DLL-specific measures.

3.1. Quality of settings

First, based on all 10 studies, the available descriptive data suggest that the overall quality of ECE experienced by DLL children as indicated by scores on widely used quality measures was similar to the overall quality of preschool ECE reported in most of the large child care studies in the U.S. All 10 of these studies focused only on Latino or Spanish-speaking DLLs, so the information provided represents the quality experienced by this population. For eight studies, the majority of the sample or the entire sample consisted of this population. For the remaining two studies, one study included one DLL and two non-DLL samples (Burchinal & Cryer, 2003) and one study measured the proportion of DLLs in the classrooms (Downer et al., 2012); although both reported the quality scores for these different sample groups, neither conducted analyses of the differences in quality by these groups.

Five of the studies reported ECERS or ECERS-R scores, all of which had averages that suggested moderate to moderately high quality (Barnett et al., 2007; Burchinal & Cryer, 2003; Chang et al., 2007; Howes et al., 2004; Winter et al., 2007)—about what is generally found in various types of ECE settings in the general population. Five studies also used the AIS/Snapshot; however, it is difficult to compare across them, as each used different aspects or versions of the measure (Burchinal & Cryer, 2003; Chang et al., 2007; Howes et al., 2004; Owen et al., 2008; Zuniga & Howes, 2009). For the remaining measures (CLASS, FDCRS, ELLCO, ORCE, SELA), there were only one or two studies for each. Although the scores for these other measures seem within typical ranges, the small number of studies limits the generalizability of these findings about quality for DLL populations for any particular instrument.

3.2. Similarity of psychometric characteristics

Second, two secondary data analyses that investigated the psychometric characteristics of various widely used quality measures

found no differences in their performance for DLL vs non-DLL samples. One study found that a number of widely used measures (ECERS, CIS, ORCE) demonstrated similar psychometric properties (levels of internal consistency and construct validity as indicated by correlations with teacher education, teacher training, or teacher–child ratios) across samples of Latino, African-American, and White children (Burchinal & Cryer, 2003). Similarly, a study investigating the psychometric properties of the CLASS found no differences in the factor structure of the CLASS, with the same three domains (emotional support, classroom organization, instructional support) replicated across samples on the basis of the classroom proportion of DLL children or the classroom proportion of Latino children (Downer et al., 2012). In addition, one study found positive associations between the FDCRS and provider characteristics (level of provider education and amount of scaffolding) for a Latina sample of providers (Zuniga & Howes, 2009), which is consistent with findings from other studies of general population samples.

3.3. Associations between quality scores and child outcomes

Third, four studies found positive associations between quality scores and child outcomes across a number of different quality measures. For one of these studies, children's outcomes were assessed in both English and Spanish (Barnett et al., 2007); in two studies, children were assessed in either English or Spanish, based on an initial language proficiency screening (Chang et al., 2007; Downer et al., 2012); and in another study, they were assessed only in English (Burchinal & Cryer, 2003). Based on general measures of the global environment and teacher–child interactions (ECERS-R, ORCE, and CLASS), higher quality predicted better language outcomes in English (Barnett et al., 2007; Burchinal & Cryer, 2003; Chang et al., 2007; Downer et al., 2012) and in Spanish (Barnett et al., 2007; Downer et al., 2012), better cognitive outcomes in English (Burchinal & Cryer, 2003), and better social outcomes (Downer et al., 2012) for Latino or DLL children, with no evidence of weaker prediction of outcomes compared to non-DLL populations. Further, these associations were similar regardless of the child's specific heritage group, either Mexican-American/Chicano or other Hispanic/Latino (Chang et al., 2007) or the caregiver–child ethnic match (Burchinal & Cryer, 2003) for the two studies with these data available.

3.4. DLL-specific measures

Of the studies included in the review, five were designed as intervention studies which included DLL populations. Three of these interventions were specifically focused on DLLs and involved at least some L1 instruction (Barnett et al., 2007; Buysse et al., 2010; Durán et al., 2010), while the other two included primarily DLL (Latino) populations (Owen et al., 2008; Winter et al., 2007). All of these had data on general quality measures, and two also had DLL-specific measures which were broadly related to, but not directly aligned with the intervention. Almost no differences were found between intervention conditions or over time (for two descriptive studies examining changes within an intervention group only) on the general measures in four of these studies, including measures of the global environment, ECERS-R (Barnett et al., 2007; Winter et al., 2007) and measures of literacy practices, ELLCO and SELA (Barnett et al., 2007; Buysse et al., 2010; Durán et al., 2010). The only exceptions were one subscale difference favoring the intervention group for the frequency of literacy activities on the ELLCO in one study (Buysse et al., 2010), and the findings from the fifth study which used a composite of two measures of teacher–child interactions, ORCE and AIS, and found higher quality for the intervention group, as well as differences in the influence of the intervention on changes in quality for African-American vs Latino children (Owen

Table 2
Summary of results.

Reference	Study purpose	Relevance to current review	Study design	Participants/setting characteristics	Measures	Results
Barnett, Yarosz, Thomas, Jung, and Blanco (2007)	To compare the effects of English/Spanish two-way immersion (TWI) versus monolingual English immersion (EI) for English- and Spanish-speaking pre-kindergarten children.	The majority of the sample was Spanish-speaking, and comparisons were made between quality scores and quality associations with child outcomes by language of instruction in Spanish condition.	Quasi-experimental design with children randomly selected to two conditions: (a) intervention group (TWI) or (b) Control group (EI). TWI instruction in English or Spanish alternated weekly with different classrooms for each.	Classrooms: <i>n</i> = 36 20 TWI, 16 EI Children: <i>n</i> = 131 Total 79 Intervention, 52 Control Age: 3- & 4-year-olds Language: Intervention—47 Spanish, 28 English, 4 other Control—28 Spanish, 21 English, 3 other Language determination based on teacher report of language proficiency scores and parent report of home language use	Quality measures: • Global quality (ECERS-R) • Literacy environment (SELA) • L1 supports (SELLCA) Child outcome measures: • Receptive language (Peabody Picture Vocabulary Test-III, Test de Vocabulario en Imagenes Peabody) • Expressive vocabulary (Woodcock Johnson-R, Woodcock Muñoz-R) • Math (Woodcock Johnson-R, Woodcock Muñoz-R) • Phonological awareness • Letter identification All children in TWI and Spanish-dominant speakers in EI assessed in both English and Spanish; others assessed in English only	<ul style="list-style-type: none"> • No significant differences between conditions, with “good” ECERS-R scores (TWI = 5.11, EI = 4.86) and “fair” SELA scores (TWI = 3.23, EI = 3.01) • Better L1 supports on SELLCA in Spanish-language TWI classes (greater use of Spanish and better incorporation of cultural background); No differences in L2 supports. • Positive associations of ECERS-R scores with WJ-R vocabulary, English rhyme, and Spanish letter identification.
Burchinal and Cryer (2003)	To determine whether child care quality had different effects on outcomes for children from diverse ethnic backgrounds (Latino, African-American, White) using data from 2 large-scale studies.	Comparisons of quality scores and associations between quality and child outcomes were examined for children from 3 ethnic backgrounds, including Latino as one group.	Separate secondary data analyses of Cost, Quality, and Child Outcomes Study (CQO) and NICHD Study of Early Child Care (SECC).	Settings: CQO <i>n</i> = 177 preschool classrooms SECC <i>n</i> = ~248 child care centers, <i>n</i> = ~137 family child care, <i>n</i> = ~199 home (father, grandparent, in-home) Children: CQO <i>n</i> = ~546 Mean age = 4.3 years Race/Ethnicity: 31 Latino, 105 African-American, 410 White All English-speaking Source of language determination and race/ethnicity not specified SECC <i>n</i> = ~584 Age: 24 & 36-month assessments Race/Ethnicity: 40 Latino, 61 African-American, 483 White All English-speaking Source of language determination and race/ethnicity not specified	CQO Study Quality measures: • Global quality (ECERS) • Provider sensitivity (CIS) • Teaching style (ECOF; Early Childhood Observation Form) • Quality composite (Above measures; also included AIS provider responsiveness) Child outcome measures: • Receptive language (Peabody Picture Vocabulary Test-R) • Pre-reading (Woodcock Johnson-R) • Pre-math skills (Woodcock Johnson-R) • Social skills (Classroom Behavior Inventory) All measures in English SECC Study Quality measure: • Caregiver-child interaction (ORCE) Child outcome measures: • Cognition (Bracken School Readiness Scale) • Language (Reynell Developmental Language Comprehension Scale) • Behavior problems (Child Behavior Checklist) • Social skills (Adaptive Social Behavior Inventory) All measures in English	<ul style="list-style-type: none"> • Average scores by subsample (Latino, African-American, White): • ECERS: 3.93, 4.15, 4.55 • CIS: 2.70, 2.97, 3.07 • ECOF: -0.18, -0.04, 0.16 • Internal consistency was similarly high for all quality measures (ECERS, CIS, ECOF, ORCE) in the 3 subsamples. • Quality measures exhibited similar validity for all 3 subsamples, regardless of caregiver-child ethnic match. The CQO quality composite positively predicted receptive language and school readiness scores for all 3 subsamples. The SECC ORCE positively predicted receptive language, school readiness, social skills, and decreased behavior problems scores for all 3 subsamples.

Table 2 (Continued)

Reference	Study purpose	Relevance to current review	Study design	Participants/setting characteristics	Measures	Results
Buyse, Castro, and Peisner-Feinberg (2010)	To assess effects of the Nuestros Niños Early Language and Literacy Program on teachers' classroom practices and Latino DLL preschoolers' language and literacy skills in both English and Spanish.	The majority of the sample was Spanish-speaking, and information was provided about quality scores and comparisons by condition for this sample.	Experimental design with random assignment to two conditions: (a) intervention group (Nuestros Niños) or (b) control group.	Classrooms: $n = 55$ 26 Intervention, 29 Control Children: $n = 193$ 92 Intervention, 101 Control Mean age = 55 mos Intervention; 54 mos Control Language: Intervention—65% Spanish-speaking Control—64% Spanish-speaking Source of language determination not specified	Quality measures: • Literacy practices (ELLCO) • Literacy practices for DLLs (ELLCO-A)	<ul style="list-style-type: none"> • Average scores for ELLCO literacy practices: Classroom Observation Scale (COS) = 3.2–3.3, • Literacy Environment Checklist (LEC) = 25.6–28.6, • Literacy Activity Rating Scale (LARS) = 6.2–7.3 • Average scores for ELLCO-A: COS = 2.2–2.5, LEC = 2.8–5.0, LARS = 0.4–0.8 • Both groups made gains in ELLCO-A COS & LEC scores over school year. • Intervention was associated with greater gains for some general and DLL-specific practices: ELLCO LARS (ES = 0.63), ELLCO-A COS (ES = 0.61), ELLCO-A LEC. (ES = 0.81)
Chang, Crawford, Early, Bryant, Howes, Burchinal, Barbarin, Clifford, and Pianta (2007)	To examine the associations between the language interactions of Spanish-speaking pre-k children and their teachers and children's social and language skills.	The sample focused only on Spanish-speaking children, and quality scores and quality associations with child outcomes were examined specifically for this group.	Secondary data analysis of the National Center for Early Development and Learning's Multi-State Study of Pre-kindergarten (Multi-state Study) and the Statewide Early Education Programs Study (SWEEP). Both were large-scale studies of state-funded pre-k programs, including a total of 11 states.	Classrooms: $n = 161$ (108 public school, 53 non-public school pre-k) Children: $n = 345$ Mean Age = 4.6 yrs (fall) Ethnic heritage: 48% Mexican American, 48% Other Latino, 3% Puerto Rican, 1% Cuban Language: All Spanish-speaking All spoke Spanish based on parent report; all did not pass PreLAS 2000 English language screener in fall pre-k	Quality measures: • Global quality (ECERS-R) • Teacher-child language interactions (Snapshot) Child measures: • English proficiency (PreLAS 2000) • Receptive language (Peabody Picture Vocabulary Test-III, Test de Vocabulario en Imagenes Peabody) • Child behaviors (Child Behavior Scale, Teacher-Child Rating Scale) • Teacher-child relationship (Student-Teacher Relationship Scale) Children in both samples assessed in either English or Spanish, based on initial language proficiency screening assessment; all children in SWEEP sample also assessed in English	<ul style="list-style-type: none"> • Average ECERS-R score = 3.74. No differences in quality scores for Mexican American children vs other groups. • Children in higher quality classrooms (ECERS-R) had greater growth in English proficiency, but there were no differences for English or Spanish receptive language. • Based on the Snapshot, 44% of teachers spoke little to no Spanish in the classroom and 56% spoke some. • Teachers who spoke some Spanish had more interactions with Spanish-speaking children, rated their peer social skills and assertiveness higher, and their relationships as closer.

Table 2 (Continued)

Reference	Study purpose	Relevance to current review	Study design	Participants/setting characteristics	Measures	Results
Downer, López, Grimm, Hamagami, Pianta, and Howes 2012	To examine the extent to which the CLASS measure of teacher–child interactions demonstrated similar psychometric properties in classrooms serving different proportions of Latino and different proportions of DLL children.	The study analyzed the psychometric characteristics of a quality measure by proportion of DLL and by proportion of Latino children.	Secondary data analysis of the National Center for Early Development and Learning's Multi-State Study of Pre-kindergarten (Multi-state Study) and the Statewide Early Education Programs Study (SWEET). Both were large-scale studies of state-funded pre-k programs, including a total of 11 states.	Classrooms: $n = 721$ By Proportion DLL: 51% No DLL, 34% Mid DLL (<50%), 15% High DLL (>50%) By Proportion Latino: 43% No Latino, 31% Mid Latino (<50%), 24% High Latino (>50%) Children: $n = 2983$ Age = eligible for kindergarten the following year Language: 32% Spanish-speaking Ethnicity: 26% Latino Language determination based on one or more criteria: 1) parent report that Spanish or language other than English spoken at home, 2) parent report that Spanish or language other than English was child's first language learned, 3) child did not pass PreLAS language screener in the fall, 4) teacher reported 100% of children were DLLs. Ethnicity (classroom proportion) based on teacher report.	Quality measure: • Classroom teacher–child interactions (CLASS) Child measures: • English proficiency screener (PreLAS 2000) • Pre-math skills (Woodcock Johnson-III, Woodcock Muñoz-R.) • Letter naming (Letter Naming Test) • Social competence and problem behaviors (Teacher–Child Rating Scale) • Emergent language and literacy skills (Academic Rating Scale) Children assessed in either English or Spanish, based on initial language proficiency screening assessment	<ul style="list-style-type: none"> • Average CLASS scores by proportion DLL • (No, Mid, High): • Classroom Organization: 3.83, 3.85, 3.89 • Emotional Support: 4.47, 4.46, 4.46 • Instructional Support: 2.08, 2.06, 2.06 • The same three-factor structure of the CLASS domains (emotional support, classroom organization, emotional support) was found for classrooms that varied by proportion of DLLs or Latinos. • CLASS scores positively predicted children's academic and social skills equally well for DLLs and Latinos as for non-DLLs and non-Latinos.
Durán, Roseth, and Hoffman (2010)	To compare the effects of English only (EO) instruction and Transitional Bilingual Education (TBE) on language and literacy development in low-income, Spanish-speaking preschoolers.	The sample was all Spanish-speaking, and comparisons were made between quality scores by language of instruction condition.	Experimental longitudinal design with children randomly assigned to two conditions: (a) intervention group (TBE) or (b) control group (EO).	Classrooms/Teachers: $n = 2$ 1 Intervention, 1 Control (All Head Start) Children: $n = 31$ 15 Intervention 16 Control Mean age = 43 mos. All Spanish-speaking Language based on parent report	Quality measure: • Literacy practices (ELLCO)	<ul style="list-style-type: none"> • ELLCO scores were not reported. • No differences between EO and TBE classrooms in ELLCO scores.

Table 2 (Continued)

Reference	Study purpose	Relevance to current review	Study design	Participants/setting characteristics	Measures	Results
Howes, Shivers, and Ritchie (2004)	Describe the process of a relationship-based intervention on changes in teacher–child relationships and behaviors in center-based child care.	The majority of the sample was Spanish-speaking, and initial quality scores were provided by program and primary population served.	Descriptive study of classroom-based intervention in child care centers.	Programs: $n = 10$ (>2 classrooms/program) Selection criterion of 4.0 or above on ECERS Primary group served by program: 3 Latino, 3 Latino and Armenian, 2 Latino and African-American, 2 African-American Children: $n = 70$ Mean age = 50 months Range = 15–62 mos Ethnicity: 59% Latino, 27% African-American, 10% Armenian immigrants, 4% Asian or biracial Source of ethnicity not specified	Quality measures: • Global quality (ECERS) • Provider sensitivity (CIS) • Provider responsiveness (AIS)	• Ranges on initial quality scores (average program scores): • ECERS Total = 4.05–5.77 Materials = 3.95–5.65 Caregiving = 4.27–6.10 • CIS Sensitivity = 3.00–3.81 ^a Harshness = 1.14–2.33 Detachment = 1.00–2.50 • AIS Responsiveness = 0.26–0.54
Owen, Klausli, Mata-Otero, and Caughy (2008)	To examine quality of care and children's outcomes in relationship-focused child care (RFC) and comparison non-relationship-focused child care (non-RFC) centers for Latino and African-American children from low-income families. RFC practices promote continuous, primary caregiver–child relationships.	The majority of the sample was Latino; differences in influence of intervention on quality of practices by race/ethnicity were examined.	Quasi-experimental design with two conditions: (a) RFC centers and (b) comparison non-RFC centers.	Child care centers: $n = 12$ 4 RFC (All Head Start), 8 non-RFC (4 Head Start, 4 non-Head Start) Children: $n = 223$ 123 RFC, 100 Non-RFC Mean age: RFC = 54.66 mos., Non-RFC = 50.97 mos. Ethnicity: RFC—53 African-American, 70 Latino Non-RFC—48 African-American, 52 Latino Some children described as Spanish-speaking, but number not reported. Race/ethnicity based on maternal report. Source of language determination not specified.	Quality measures: • Caregiver–child interaction (adapted from ORCE and AIS)	• Average scores on ORCE dimensions by 4 groups • (RFC-Latino, RFC-African-American, non-RFC-Latino, non-RFC-African-American): • Sensitivity: 2.63, 2.54, 2.44, 2.00 • Intrusiveness: 1.07, 1.06, 0.96, 1.02 • Detachment: 1.61, 1.60, 1.63, 1.83 • Cognitive Stimulation: 1.75, 1.83, 1.55, 1.38 • RFC centers had higher quality care, although race/ethnicity had a few moderating effects. Greater caregiver sensitivity and higher-level caregiver involvement were found in RFC than non-RFC centers for African-American children, while these practices were equally high across centers for Latino children. Greater child engagement of caregivers was found in RFC centers only for African-American children. • The influence of RFC centers was similar for Latino and African-American children for all other aspects of quality.

Table 2 (Continued)

Reference	Study purpose	Relevance to current review	Study design	Participants/setting characteristics	Measures	Results
Winter, Zurcher, Hernandez, and Yin (2007)	To examine the quality of care and developmental progress of children enrolled in the Early ON School Readiness Project, which serves a predominantly Latino population.	The majority of the population served was Latino; child care quality was examined descriptively for participating centers.	Descriptive study of quality of care in centers participating in the Early ON School Readiness Project.	Centers: Year 1 $n = 17$ Year 2 $n = 34$ (Centers included Head Start, nonprofit, faith-based, and for-profit) Children: n not reported Age = 3- to 5-year-olds Primarily Mexican-American population, majority English-speaking Language and ethnicity based on parent report	Quality measure: • Global quality (ECERS-R)	• There was little change over time in quality scores, from fall to spring and year 1 to year 2. ECERS-R Total = 4.9–5.3; Subscale scores ranged from 4.4 to 5.8.
Zuniga and Howes (2009)	To examine quality of care and children's engagement in pre-academic activities in family child care settings with low-income Latino providers and children.	The child care providers and the majority of the population served were Latino; family child care quality and predictors of quality were examined descriptively for this sample.	Descriptive study of family child care homes.	Family child care providers: $n = 115$ (all females) Ethnicity: All Latina of Mexican heritage Ethnicity based on self-report Children: $n = 460$ Mean age = 36 months Ethnicity: 80% Latino Ethnicity based on parent report	Quality measures: • Global quality (FDCRS) • Provider scaffolding (Snapshot) • Provider responsiveness (AIS)	• Average scores: • FDCRS Total mean = 4.04, Snapshot scaffolding = 17.75% of observations, • AIS responsiveness = 77.64% of observations Greater education levels and scaffolding behaviors predicted higher global quality. After provider behaviors (scaffolding and responsiveness) were included, training, experience, and number of children were not significant.

^a Note: A value of 4.43 is reported in the manuscript, although the scale values range from 1 to 4, and the manuscript appropriately describes the scale as a 4-point scale.

et al., 2008). In contrast, when studies included quality measures designed to capture specific aspects of the environment related to supports for DLLs, they did find positive intervention effects for these measures, both on the SELICA (Barnett et al., 2007) and the ELLCO-A (Buysse et al., 2010). All of these studies were measuring the impact of the interventions using the quality measures, yet only found positive effects for DLL-specific measures and not for general quality measures.

Similarly, when one study used a general measure, the Snapshot, to gather data relevant to bilingual instruction (i.e., the amount of Spanish spoken in the classroom by teachers), it found positive associations with child outcomes for Spanish-speaking children (Chang et al., 2007). More positive teacher ratings of social skills were associated with higher rates of Spanish language use by teachers.

4. Discussion

4.1. Conclusions

The purpose of this comprehensive review was to examine the research literature systematically in order to determine what is known about the use of ECE quality measures for DLL children in center-based and/or home-based settings during the preschool years. Based on the 10 studies that met the criteria for inclusion in the review, namely that the results were analyzed separately for the DLL samples, there are some common findings that merit attention.

One of the key conclusions of this review is that widely used instruments for measuring the quality of ECE settings yield similar results for DLL populations as for the general population. This primary finding is derived from the results across the set of studies and across a number of different measures of ECE quality. Based on several widely used measures of general ECE quality, there were no differences for DLLs and non-DLLs in the quality of experiences, the psychometric properties of these measures, and the associations of quality with children's cognitive, language, and social outcomes. These measures were designed with general ECE populations in mind, including DLL and non-DLL children, but not necessarily with a focus on the particular needs of DLLs.

In contrast, although only two studies included measures specifically designed to examine supports for DLLs, when such measures were included, they seemed to tap into different dimensions of the environment than the general measures and likely captured experiences of specific importance for these children. Both of these studies were designed as intervention studies, with a focus on changing classroom practices for DLLs, although the DLL-specific quality measures themselves were not directly aligned with the intervention. Similarly, results from a third non-intervention study found positive associations between measurement of relevant aspects of bilingual instruction and child outcomes. In all cases, these findings suggest that these DLL-specific measures may be sensitive to assessing aspects of quality that are particularly important for DLLs. Although the research evidence has not clearly determined the most effective instructional practices for preschool-age DLLs, there is some support from both the preschool and the school-age literature for curricular approaches which focus on language and literacy instruction, in addition to high-quality instruction in general (August & Shanahan, 2006; Buysse et al., 2014; Castro et al., 2010; Cheung & Slavin, 2012). The DLL-specific measures were largely focused on the language and literacy environment, and L1 supports in particular, for DLLs in the classroom.

Given the number of studies, limited conclusions can be drawn about the typical quality of child care experiences for DLL populations in center-based or home-based settings for any particular

measure. However, across the various measures, the scores generally were similar to those found in the broader literature, which typically shows that quality tends to be in the medium range, with higher scores for aspects related to interaction and emotional support than to instructional activities or other opportunities for learning (Burchinal et al., 2011; Peisner-Feinberg & Yazejian, 2010). Along these same lines, a recent analysis based on the ECERS-R and CLASS showed few differences in the quality of center-based ECE for immigrant vs native preschoolers in California, suggesting that the less than optimal level of quality is of similar concern for all children in this country (Karoly & Gonzalez, 2011).

One question that this review was not able to answer is what are the "best" quality measures to use with DLL populations, both for research and practice. The variability in the definitions of DLLs, the different methods used for determining DLL status, and the often limited information included in published articles about these methods all make it difficult to compare across studies. This limitation is common across all types of studies that incorporate linguistically and culturally diverse samples, and further compounds the difficulty of assessing the state of knowledge about young DLLs using studies that may not always be designed specifically to address this issue. Further, there were many other studies that may have included adequate DLL samples and ECE quality measures, but this information was not gathered or analyzed in a way that provided results about these measures in relation to DLLs. For example, studies may have reported that DLLs attended the programs included in the research, but did not report any information about the percentage of DLLs included in the sample or did not examine the associations between the quality scores and the proportion of DLLs or a DLL subsample.

In interpreting the findings from this review, it is important to acknowledge that the issue of interest—the performance of measures used to evaluate the quality of ECE for DLL populations—was not the primary question for many of these studies. There was one secondary data analysis which compared the quality of experiences for DLL/non-DLL groups as the focus (Burchinal & Cryer, 2003), two studies which examined the quality of experiences specifically for DLL populations as the focus (Chang et al., 2007; Zuniga & Howes, 2009), and one secondary data analysis which examined the functioning of a quality measure in relation to the proportion of DLLs as the primary question of interest (Downer et al., 2012). In contrast, for the six remaining studies, the primary focus was on the effects of some type of intervention (Barnett et al., 2007; Buysse et al., 2010; Durán et al., 2010; Howes et al., 2004; Owen et al., 2008; Winter et al., 2007), although data on quality measures were reported in a manner that allowed inclusion in this review.

Moreover, one of the key findings from this review is that a total of only 10 studies that analyzed results for DLL populations were available in the published literature. There were no studies providing such information for many of the measures, and only a limited number available for others. The findings from this review are based on these 10 studies, and although the total number is limited, the results provide a summary of the current state of knowledge, indications of both strengths and weaknesses of the research in this area, and guidance for future directions. Although there is still a great need for research related to the topic of measuring quality in ECE settings for DLLs, there is useful information that can be learned from the studies to date.

Further, the conclusions derived from this review are consistent with the findings from other recently conducted studies not available in the peer-reviewed literature. For example, a study described in a recent edited volume found few differences across different quality measures (CLASS, ECERS-R, and Snapshot) for ECE classrooms with varying proportions of DLLs (Fuligni & Howes, 2011). There were a few associations of teachers' reports of a greater proportion of "limited English proficient" (LEP) students with higher

scores on some dimensions of instructional support as measured by the CLASS and more time spent in language and literacy activities as measured by the Snapshot. However, at the individual child level, there were no differences in the quality experienced by DLLs and non-DLLs. Similarly, another study found no associations between the linguistic composition of pre-k classrooms and the quality of language and literacy instruction (Layzer & Maree, 2011).

4.2. Methodological considerations

Future research efforts may be improved by addressing several common methodological issues found in the reviewed studies which limited the conclusions that could be drawn about the performance of ECE quality measures for DLL populations. These include issues related to the samples represented in these studies, the definitions of DLLs across studies, the measurement of DLLs at the setting and/or classroom level, the assessment of child outcomes, and the training and skills of the data collectors to capture information related to DLLs. Addressing such issues in the design of future research studies could strengthen the knowledge base about ECE quality measures in relation to DLL populations.

First, there were some constraints across the samples included in these studies, many of which mirrored the broader literature. Sample sizes varied greatly across the 10 studies, with most of these in the small to moderate range. Individual sample sizes ranged from 2 to 721 classrooms, providers, or programs, depending on the unit of analysis, with very low numbers for three studies (up to 12), modest numbers for three studies (35–55), moderate numbers for three studies (115–250), and large numbers for one study (over 700). The picture is similar with regard to sample sizes for children, when applicable, with a range from 31 to 2983, with two studies having sample sizes up to 70, three of the studies having sample sizes between 130 and 225, three having sample sizes between 345 and 600, and one secondary data analysis having a sample size near 3000 children. Consistent with much of the ECE literature, the age groups studied primarily represented 3- to 5-year-olds; only one study included 2-year-olds in the sample and no studies included any younger children. Similarly to the ECE literature in general, most of the studies (8 of the 10) included only center-based settings; only one study included some family child care and home settings (Burchinal & Cryer, 2003), and one study included only family child care settings (Zuniga & Howes, 2009). However, a variety of center-based settings were represented across studies, including public preschool, community child care centers, and Head Start. In addition, all of the studies focused on Spanish-speaking and/or Latino populations, either exclusively or in comparison to English-speaking groups; no other language groups were examined in the data analyses. Thus, future research incorporating more linguistically diverse samples and covering the full range of ECE settings would be valuable. However, it should be acknowledged that these constraints are typical of many related areas of research, not just this particular body of literature.

Second, the information available in the published articles as well as the methods used for determining DLL status were often quite limited, although it should be kept in mind that DLLs were not necessarily the primary focus for all of these studies. The definition of the DLL sample population was quite variable across studies, with samples defined based on language, ethnicity, or sometimes both. Further, the method used to determine children's ethnic or linguistic groups often was constrained, or in many cases, inadequately described. Several studies relied on parent or teacher report of home language (Barnett et al., 2007; Durán et al., 2010), ethnicity (Zuniga & Howes, 2009), or both (Winter et al., 2007) for specifying the sample or subgroups within the sample. In only two studies, both of which utilized the same sample, were more rigorous methods of individual assessments of children's language

proficiency used in addition to parent or teacher report (Chang et al., 2007; Downer et al., 2012). In the remainder of the studies, however, the children were described as "Spanish-speaking" or "Latino/Hispanic" without any information provided about how those determinations were made (Burchinal & Cryer, 2003; Buysse et al., 2010; Howes et al., 2004; Owen et al., 2008). Although these studies may have utilized relevant methods for determining their DLL samples, these methods were not adequately described in the published materials. It would greatly enhance future research efforts to develop common definitions and methods for determining DLL status that could be used across different studies—a valuable research endeavor itself.

Third, the specification of the proportion of DLLs or DLL and non-DLL groups at the setting level (classroom or program) often was not reported, which limited the conclusions of this review about the implications of these measures for DLLs. There are many other studies that included a substantial number of DLL children in the overall sample, but these could not be included in this review because information about the distribution of DLLs was not provided at the level at which the quality data were gathered (i.e., classroom or program) or about the quality results with regard to this distribution. In these cases, it could not be determined whether the overall results were similar for settings that varied by the proportion of DLLs served, or for DLL compared to non-DLL groups. Of the studies included in the present review, only three specifically addressed this issue in the analyses (Burchinal & Cryer, 2003; Downer et al., 2012; Owen et al., 2008); for the remainder, it is only known that these measures were used with a sample that consisted of a majority of DLLs (an alternative criterion for inclusion in this review so that the results could be interpreted with regard to DLLs). However, it is important to note that the overall characteristics of the samples reported in these studies may not have reflected the specific characteristics of each individual setting included in the analysis, as that information was not typically provided at the unit of analysis. For future studies, it would be useful to report information about the actual distribution of DLLs in the sample at the level at which these data are gathered. For example, for studies using classroom-level measures, it would be helpful to report information about the proportion of DLLs in each classroom, rather than just the program overall, as these numbers may vary across classrooms.

Fourth, it is important to note that most of the studies examining child outcomes in relation to quality did not employ parallel assessment procedures in which children were assessed in both English and their home language. Some studies assessed children in either English or their home language, depending on initial language proficiency, while others assessed children only in English. Only one study which examined the associations between quality and outcomes used parallel assessment procedures, although some of the other studies also used such assessment procedures but did not examine the quality-outcome associations. In order to fully characterize the associations between the quality of ECE experiences and children's outcomes, it is important to ensure that the child outcome measures that are used fully represent the range of children's skills. For DLLs who may demonstrate skills in one language or the other, parallel assessment procedures are necessary to capture this breadth of knowledge.

Fifth, it would be important to consider the skills of data collectors in regard to gathering observational measures in settings that include DLLs. Little information was provided in most studies about the training and ability of data collectors to capture interactions with non-English-speaking children or providers for relevant measures. Given that most of these measures of ECE quality focus at least to some extent on aspects of the environment such as teacher-child interactions, it would be important to have data collectors who can understand children's home languages in order to ensure that such items are appropriately scored for the

particular DLL population being studied. Of the 10 studies included in this review, only two noted that bilingual or Spanish-speaking data collectors gathered the classroom observation data (Buysse et al., 2010; Zuniga & Howes, 2009). It is worth noting, however, that several of the studies indicated that bilingual data collectors gathered child assessment data. In conducting research on DLL populations, it is critical to ensure that the methods used for gathering data with observational measures, even classroom-wide measures, give adequate attention to bilingual issues.

4.3. Future research directions

It is worth noting that, in addition to new research, there appear to be opportunities for further exploration of these issues through numerous existing studies. Hundreds of studies were examined that included DLL populations, but had not analyzed their data in relation to these populations and therefore, could not be included in this review. Thus, there is the potential for many more studies to address the issues raised by this review using existing datasets. Moreover, there are numerous other measures of quality, both general and DLL-specific, for which there was no existing published literature at the time of this review that examined quality with regard to defined DLL populations. Studies including DLL populations were found for only 10 of the 46 quality measures included in the search, with only one or two studies found for most of these instruments. Thus, the existing knowledge base is limited at this point to a number of the more widely used general measures and a few DLL-specific measures, with these questions still unanswered for many measures.

Most of the studies included in this review have been published in the past few years, suggesting that this is a burgeoning area of research; there is hope that much more may be known about this issue in the next decade. This topic also represents the intersection between two growing areas of broader interest in the field—measuring ECE quality and examining the experiences of DLLs. There is a need for further research to learn about the actual quality of experiences for DLL populations, including both center-based and home-based settings. Based on the results of this review, there are few studies which have examined ECE quality for DLL children exclusively or based on the proportions served, for both center-based and home-based settings. Therefore, we have limited information about DLL children's ECE experiences.

Even though the available literature so far seems to suggest that quality may look similar based on widely used measures, it does not necessarily mean that DLLs experience those settings in the same way. Most of these studies have relied on general measures of quality gathered at a classroom-wide level. The methodological concerns that were raised around data collector training and skills for observing DLLs suggest the possibility that some aspects of their experiences could have been missed with the use of these measures. Moreover, many of these measures rely on an average or typical rating, which may obscure the experiences of DLLs who represent a smaller proportion of the classroom in many settings. Further, the few studies using DLL-specific measures offer indications that there are other aspects of these settings that may be important to capture in order to fully understand these children's experiences, such as the L1 supports available in the classroom, the opportunities for L2 development, and the extent to which the interactions are culturally and linguistically responsive. Therefore, it would be important to design studies which are able to directly assess the experiences of DLLs. Such studies would need to include features such as the collection of both general and DLL-specific measures, or the use of different quality measures designed to gather data at both the classroom and the individual child level, or the collection of quality data using the same measures separately for DLLs and non-DLLs in the classroom.

With regard to the measures themselves, research is needed to better understand whether existing measures perform similarly in settings that vary by the proportion or population (e.g., different language groups) of DLL children served and the extent to which various measures adequately capture critical factors of ECE quality for DLL children. There may be some aspects of environments which would need to look different for DLLs to serve the same function as for non-DLLs; further, it may be that the same aspect of an environment might be used differently by DLLs or might have a different impact on DLLs. It is important to understand how the same item on a general measure of quality would need to look in practice to serve the same function for a child who speaks a different language. When considering issues around interactions, for example, it would be critical to consider the role language plays as well as the extent to which there are any language barriers in teachers' and children's communications with and understanding of one another (e.g., Do behavioral interactions take place in the child's home language while instructional interactions take place in English?, Do DLLs receive the same amount and level of response from teachers and from peers as monolinguals?) When examining the classroom environment, for example, it would be important to consider whether DLLs understand and are equally as familiar with materials from both literacy and cultural perspectives. (e.g., Are there books for children in their home language as well as in English? Are there dramatic play materials that represent diverse cultural and linguistic backgrounds?) When examining assessment practices, for example, it would be important to consider the language of the assessment measures, how that information is used to inform instruction, and the language of instruction. (e.g., Are parallel assessment approaches in both English and the child's home language used for ongoing assessment? How are assessment results from each language used to inform instruction—for example, how is instruction in English managed if a child understands a concept in the home language but not in English?)

Relatedly, there is a need for research to examine the additional information contributed by DLL-specific measures beyond that from general quality measures. There are likely some aspects of ECE environments which may be specific to DLL populations, while there are other aspects of the environment which would not necessarily differ by DLL status. When choosing or designing tools for use with DLL populations, researchers need to consider the type of information that is needed and to ensure that both specific and general measures are being used in ways that are appropriate for their intended purpose. Although the studies that have been conducted to date with DLL-specific measures have been designed as intervention studies, it would be helpful to examine how such measures operate in everyday ECE settings which are serving DLL children. The findings from the current review suggest that some of the general quality measures may operate similarly for DLL and non-DLL populations, but that other measures focusing more specifically on aspects of the environment related to supports for DLLs may capture different dimensions and have different effects on child outcomes. Thus, the combination of the two types of measures may offer the most complete representation of DLL children's experiences in ECE settings. Similarly, when measuring children's outcomes, using parallel assessment procedures would provide the most accurate picture of the breadth of children's skills related to their experiences in ECE settings.

Although not meeting the criteria for this review, preliminary information from development studies is emerging about the use of some of these measures with DLL populations. This is particularly the case for the DLL-specific measures, with information about several measures in a recently published volume (Howes, Downer & Pianta, 2011), including the Bilingual Teacher Behavior Rating Scale (B-TBRS; Solari, Landry, Zucker & Crawford, 2011), the Classroom Assessment of Supports for Emergent Bilingual

Acquisition (CASEBA; Freedson, Figueras-Daniel, Frede, Jung & Sideris, 2011), the Language Interaction Snapshot (LISn; Atkins-Burnett, Sprachman, López, Caspe & Fallin, 2011), and the Observation Measures of Language and Literacy (OMLIT; Layzer & Maree, 2011). Although generally based on preliminary results, the picture presented by these studies supports many of the conclusions from the present review. This emerging body of research includes a focus on observation from the perspective of DLLs within the classrooms, considers both classroom-wide and child-level measurement to allow for differences in the nature of DLLs' experiences within the classroom, examines both L1 and L2 use, and includes measurement of the linguistic composition of the classroom.

In sum, the findings from this review call for five considerations for future research: (1) Development of consensus within the research field of common definitions and methods for determining DLL status, as well as research to further the path toward such an endeavor; (2) Research that examines a broader range of aspects of ECE quality, including using both general and DLL-specific measures within the same studies, both to document the level of quality for DLLs as well as to capture the full range of their experiences in ECE; (3) Research that examines hypothesized associations between specific aspects of quality and specific outcomes for children across different measures, in order to begin to disentangle the aspects of quality that differ for DLLs from those that are common for all children, regardless of the particular type of quality measure; (4) Research that includes more diverse and better specified samples in order to enhance the generalizability and interpretation of the findings, including linguistically diverse samples beyond Latinos or Spanish-speaking children, a range of age groups and ECE settings, specification and examination of information about DLLs at the setting level, and adequate sample sizes; and (5) Research that utilizes bilingual observers who are able to adequately capture non-English interactions within the classroom and assess child outcomes in both their home language and English, as relevant. Such considerations will be important for measures development efforts and for providing information that will be useful and informative to researchers, practitioners, and policymakers interested in the quality and effects of ECE experiences for young DLLs.

References¹

- Abbott-Shim, M., & Sibley, A. (1992a). *Assessment profile for early childhood programs: Research Edition I*. Atlanta, GA: Quality Assist.
- Abbott-Shim, M., & Sibley, A. (1992b). *Assessment profile for family child care homes: Research Edition I*. Atlanta, GA: Quality Assist.
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541–552.
- Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2010). *Language interaction snapshot (LISn)*. Princeton, NJ: Mathematica Policy Research.
- Atkins-Burnett, S., Sprachman, S., López, M., Caspe, M., & Fallin, K. (2011). The Language Interaction Snapshot (LISn): A new observational measure for assessing language interactions in linguistically diverse early childhood programs. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom* (pp. 117–146). Baltimore, MD: Paul H. Brookes.
- Atwater, J., Lee, Y., Montagna, D., Reynolds, L., & Tapia, Y. (2009). *Classroom CIR-LE: Classroom code for interactive recording of children's learning environments*. Kansas City, KS: Juniper Gardens Children's Project.
- August, D., Shanahan, T., et al. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Mahwah, NJ: Erlbaum. <http://dx.doi.org/10.1017/S0272263108080182>
- *Barnett, W. S., Yarosz, D. J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly, 22*(3), 277–293. <http://dx.doi.org/10.1016/j.ecresq.2007.03.003>
- Bernhard, J., Cummins, J., Campoy, F., Ada, A., Winsler, A., & Bleiker, C. (2006). Identity texts and literacy development among preschool English language learners: Enhancing learning opportunities for children at risk for learning disabilities. *Teachers College Record, 108*(11), 2380–2405. <http://dx.doi.org/10.1111/j.1467-9620.2006.00786.x>
- Boller, K., Sprachman, S., & the Early Head Start Research Consortium. (1998). *The child-caregiver observation system*. Princeton, NJ: Mathematica Policy Research.
- Bradley, R. H., Caldwell, B. M., & Corwyn, R. F. (2003). The Child Care HOME Inventories: Assessing the quality of family child care homes. *Early Childhood Research Quarterly, 18*, 294–309.
- Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow (Ed.), *Reasons to take stock and strengthen our measures of quality* (pp. 11–31). Baltimore, MD: Paul H. Brookes.
- Burchinal, P., Kainz, K., Cai, K., Tout, K., Zaslow, M., Martinez-Beck, I., et al. (2009). *Early care and education quality and child outcomes*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, and Child Trends.
- *Burchinal, M. R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly, 18*, 401–426. <http://dx.doi.org/10.1016/j.ecresq.2003.09.003>
- *Buysse, V., Castro, D. C., & Peisner-Feinberg, E. (2010). Effects of a professional development program on classroom practices and outcomes for Latino dual language learners. *Early Childhood Research Quarterly, 25*(2), 194–206. <http://dx.doi.org/10.1016/j.ecresq.2009.10.001>
- Buysse, V., Peisner-Feinberg, E., Paez, M., Hammer, C., & Knowles, M. (2014). Effects of early education programs and practices on the development and learning of dual language learners: A review of the literature. *Early Childhood Research Quarterly, 29*. <http://dx.doi.org/10.1016/j.ecresq.2013.08.004> (in press)
- Carl, B. (2007). *Child Caregiver Interaction Scale* Unpublished research instrument. Indiana, PA: Dissertation for Indiana University of Pennsylvania.
- Castro, D. C. (2005). *Early language and literacy classroom observation, addendum for English language learners*. Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute.
- Castro, D. C., Espinosa, L. M., & Páez, M. M. (2011). Defining and measuring quality in early childhood practices that promote dual language learners' development and learning. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 257–280). Baltimore, MD: Paul H. Brookes.
- Castro, D., Peisner-Feinberg, E., Buysse, V., & Gillanders, C. (2010). Language and literacy development in Latino dual language learners: Promising instructional practices. In O. Saracho, & B. Spodek (Eds.), *Language and cultural diversity in early childhood education* (pp. 65–93). Charlotte, NC: Information Age.
- *Chang, F., Crawford, G., Early, D., Bryant, D., Howes, C., Burchinal, M., et al. (2007). Spanish-speaking children's social and language development in pre-kindergarten classrooms. *Early Education and Development, 18*(2), 243–269. <http://dx.doi.org/10.1080/10409280701282959>
- Cheung, A. C. K., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research, 82*(4), 351–395. <http://dx.doi.org/10.1080/15235882.2005.10162835>
- Collins, M. (2010). ELL preschoolers' English vocabulary acquisition from storybook reading. *Early Childhood Research Quarterly, 25*(1), 84–97. <http://dx.doi.org/10.1016/j.ecresq.2009.07.009>
- CQO Study Team. (1995). *Cost, quality, and child outcomes in child care centers* [Technical report]. Denver: Department of Economics, Center for Research in Economic and Social Policy, University of Colorado at Denver.
- Dickinson, D. (2008). *Teacher Instructional Engagement Scale* Unpublished research instrument. Nashville, TN: Vanderbilt University.
- Downer, J. T., Booren, L. M., Lima, O. A., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly, 25*, 1–16. <http://dx.doi.org/10.1016/j.ecresq.2009.08.004>
- *Downer, J. T., López, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the Classroom Assessment Scoring System in diverse settings. *Early Childhood Research Quarterly, 27*(1), 21–32. <http://dx.doi.org/10.1016/j.ecresq.2011.07.005>
- *Durán, L. K., Roseth, C. J., & Hoffman, P. (2010). An experimental study comparing English-only and transitional bilingual education on Spanish-speaking preschoolers' early literacy development. *Early Childhood Research Quarterly, 25*(2), 207–217. <http://dx.doi.org/10.1016/j.ecresq.2009.10.002>
- Early, D. M., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., et al. (2005). *Pre-kindergarten in eleven states: NCEdL's multi-state study of pre-kindergarten & study of State-Wide Early Education Programs (SWEEP), preliminary descriptive report*. Chapel Hill, NC: FPG Child Development Institute, University of North Carolina. Retrieved from <http://projects.fpg.unc.edu/~ncedl/pdfs/swEEP.ms.summary.final.pdf>
- Emlen, A. C., Koren, P. E., & Schultze, K. H. (2000). *A packet of scales for measuring quality of child care from a parent's point of view*. Portland, OR: Regional Research Institute for Human Services, Portland State University.
- Farver, J., Lonigan, C., & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An experimental study of two methods. *Child Development, 80*(3), 703–719. <http://dx.doi.org/10.1111/j.1467-8624.2009.01292.x>

¹ Reference marked with (*) denotes studies included in the review of the use of program quality measures with DLLs.

- Fiene, R. (1984). *Child Development Program Evaluation Scale and COFAS*. Washington, DC: Children's Services Monitoring Consortium.
- Fox, L. K., Hemmeter, M. L., & Snyder, P. (2014). *Teaching Pyramid for Preschool Classrooms (TPOT) for Preschool Classrooms (Research Edition)*. Baltimore, MD: Brookes Publishing.
- Frede, E. C., & Miller, A. K. (1990). *Preschool classroom implementation rating instrument – high/scope manual*. Unpublished research instrument. Available from the first author at efrede@tcnj.edu
- Freedson, M., Figueras-Daniel, A., & Frede, E. (2009). *Classroom assessment of supports for emergent bilingual acquisition*. New Brunswick, NJ: National Institute for Early Education Research.
- Freedson, M., Figueras-Daniel, A., Frede, E., Jung, K., & Sideris, J. (2011). The Classroom Assessment of Supports for Emergent Bilingual Acquisition: Psychometric properties and key initial findings from New Jersey's Abbott Preschool Program. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom*. Baltimore, MD: Paul H. Brookes.
- Fulgini, A. S., & Howes, C. (2011). Experiences of low-income dual language learning preschoolers in diverse early learning settings. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom* (pp. 19–44). Baltimore, MD: Paul H. Brookes.
- Gilliam, W. S. (2008). *Development of the Preschool Mental Health Climate Scale. Final report*. Unpublished research instrument. New Haven, CT: Yale Child Study Center.
- Goodson, B. D., Layzer, J. I., & Layzer, C. J. (2005). *Quality of early childhood care settings: Caregiver Rating Scale (QUEST)*. Cambridge, MA: Abt Associates.
- Goodson, B. D., Layzer, C. J., Smith, W. C., & Rimdzius, T. (2006). *Observation measures of language and literacy instruction in early childhood (OMLIT)*. Cambridge, MA: Abt Associates.
- Grehan, A. W., & Smith, L. J. (2004). *The early literacy observation tool*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.
- Gormley, W. T., Jr. (2008). The effects of Oklahoma's pre-K program on Hispanic children. *Social Science Quarterly*, 89(4), 916–936. <http://dx.doi.org/10.1111/j.1540-6237.2008.00591.x>
- Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, 41(6), 872–884. <http://dx.doi.org/10.1037/0012-1649.41.6.872>
- Halle, T., & Vick, J. E. (2007). *Quality in early childhood care and education settings: A compendium of measures*. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Halle, T., Whittaker, J. E., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures* (second edition). Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Harms, T., & Clifford, R. M. (1980). *Early Childhood Environment Rating Scale*. New York, NY: Teacher College Press.
- Harms, T., & Clifford, R. M. (1989). *The Family Day Care Rating Scale*. New York, NY: Teacher College Press.
- Harms, T., Clifford, R., & Cryer, D. (2005). *Early Childhood Environment Rating Scale—Revised*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale—Revised*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family Child Care Environment Rating Scale—Revised*. New York, NY: Teachers College Press.
- Hernandez, D. J., Denton, N. A., & Macartney, S. E. (2007). Young Hispanic children in the 21st century. *Journal of Latinos and Education*, 6(3), 209–228. <http://dx.doi.org/10.1080/15348430701312859>
- High/Scope Educational Research Foundation. (2003). *Preschool Program Quality Assessment, 2nd edition (PQA) Administration Manual*. Ypsilanti, MI: High/Scope Press.
- Holland-Coviello, R. (2005). *Language and literacy environment quality in early childhood classrooms: Exploration of measurement strategies and relations with children's development*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 3187493).
- Howes, C., Downer, J. T., & Pianta, R. C. (Eds.). (2011). *Dual language learners in the early childhood classroom*. Baltimore, MD: Paul H. Brookes Publishing Co.
- *Howes, C., Shivers, E. M., & Ritchie, S. (2004). Improving social relationships in child care through a researcher-program partnership. *Early Education & Development*, 15(1), 57–78. http://dx.doi.org/10.1207/s15566935eed1501_4
- Howes, C., & Stewart, P. (1987). Adult Involvement Scale. *Developmental Psychology*, 23(3), 423–430.
- Hyson, M. C., Hirsh-Pasek, K., & Rescorla, L. (1990). The classroom practices inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4- and 5-year old children. *Early Childhood Research Quarterly*, 5, 475–494.
- Karoly, L. A., & Gonzalez, G. C. (2011). Early care and education for children in immigrant families. *The Future of Children*, 21(1), 71–101. Retrieved from <http://www.jstor.org/stable/41229012>
- Kriener-Althen, K., & Mangione, P. (2010, April). The PITC program assessment rating scale: Assessing structural, global, and process quality in infant-toddler care. In A. M. Daniels (Ed.), *Defining, providing, and assessing quality child care in group settings that support infants and toddlers. Symposium conducted at the American Educational Research Association annual meeting*. Denver, CO.
- Landry, S. H., Crawford, A., Gunnnewig, S., & Swank, P. R. (2001). *Teacher Behavior Rating Scale*. Unpublished research instrument. Center for Improving the Readiness of Children for Learning and Education, University of Texas Health Science Center at Houston.
- Layzer, C., & Maree, K. R. (2011). Using the Observation Measures of Language and Literacy Instruction (OMLIT) to characterize early literacy classrooms—focus on dual language learners. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom*. Baltimore, MD: Paul H. Brookes.
- Lugo-Neris, M., Jackson, C., & Goldstein, H. (2010). Facilitating vocabulary acquisition of young English language learners. *Language, Speech, and Hearing Services in Schools*, 41(3), 314–327. <http://dx.doi.org/10.1044/0161-1461>
- Magnuson, K. A., Meyers, M. K., & Waldfogel, J. (2007). Public funding and enrollment in formal child care in the 1990s. *Social Service Review*, 81, 47–83.
- Magnuson, K. A., & Waldfogel, J. (2005). Early childhood care and education: Effects on ethnic and racial gaps in school readiness. *The Future of Children*, 15(10), 169–196.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., 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>
- Mathematica Policy Research. (2010). *Measures of Early Language and Literacy Environment (ELLE)*. Unpublished research instrument. Princeton, NJ: Author.
- McCartney, K., Dearing, E., Taylor, B. A., & Bub, K. L. (2007). Quality child care supports the achievement of low-income children: Direct and indirect pathways through caregiving and the home environment. *Journal of Applied Developmental Psychology*, 28(5–6), 411–426. <http://dx.doi.org/10.1016/j.appdev.2007.06.010>
- National Center for Children in Poverty. (2007). *Low-income children in the United States: National and state trend data 1996–2006*. New York, NY: Author.
- National Institute for Early Education Research. (2005). *Support for English language learners classroom assessment*. Rutgers, NJ: Author.
- National Research Council. (2008). *Early childhood assessment: Why, what, and how?* Committee on Developmental Outcomes and Assessments for Young Children. In C. E. Snow, & S. B. Van Hemel (Eds.), *Board on Children, Youth and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.
- Neuman, S., Dwyer, J., & Koh, S. (2007). *Child/Home Early Language & Literacy Observation Tool (CHELLO)*. Baltimore, MD: Paul H. Brookes.
- NICHD Early Child Care Research Network. (1996). *Observational record of the caregiving environment*. New York, NY: Ablex.
- NICHD Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science*, 4(3), 116–135. http://dx.doi.org/10.1207/S1532480XADS0403_2
- Office of Head Start. (2009). *Historical overview of Head Start's support for children and families who speak languages other than English*. Washington, DC: Author.
- *Owen, M. T., Klausli, J. F., Mata-Otero, A., & Caughy, M. O. (2008). Relationship-focused child care practices: Quality of care and child outcomes for children in poverty. *Early Education and Development*, 19(2), 302–329. <http://dx.doi.org/10.1080/10409280801964010>
- Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The Cost, Quality, and Outcomes Study. *Merrill-Palmer Quarterly*, 43(3), 451–477.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534–1553. <http://dx.doi.org/10.1111/1467-8624.00364>
- Peisner-Feinberg, E. S., & Schaaf, J. M. (2009). *Evaluation of the North Carolina More at Four Pre-kindergarten Program: A look across time at children's outcomes and classroom quality from pre-k through kindergarten (2003–2009)*. The University of North Carolina, FPG Child Development Institute: Chapel Hill.
- Peisner-Feinberg, E. S., Schaaf, J. M., LaForett, D., Hildebrandt, L. M., & Sideris, J. (2014). *Effects of Georgia's Pre-K Program on children's school readiness skills: Findings from the 2012–2013 evaluation study*. Chapel Hill: The University of North Carolina, FPG Child Development Institute.
- Peisner-Feinberg, E. S., & Yazejian, N. (2010). Research on program quality: The evidence base. In P. W. Wesley, & V. Buysse (Eds.), *The quest for quality: Promising innovations for early childhood programs* (pp. 21–45). Baltimore, MD: Paul H. Brookes.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System: Pre-K (CLASS Pre-K)*. Baltimore, MD: Paul H. Brookes.
- Porter, T., Rice, R., & Rivera, E. (2006). *Assessing quality in family, friend and neighbor care: The Child Care Assessment Tool for Relatives*. New York, NY: Institute for a Child Care Continuum.
- Ramey, S. L., & Ramey, C. T. (2002). *The Ramey Observation of Learning Essentials (ROLE)*. Unpublished research instrument. Washington, DC: Georgetown University Center for Health and Education.
- Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emerging Academic Snapshot Scale*. Unpublished research instrument. Los Angeles, CA: UCLA.
- Roberts, T., & Neal, H. (2004). Relationships among preschool English language learner's oral proficiency in English, instructional experience and literacy development. *Contemporary Educational Psychology*, 29(3), 283–311. <http://dx.doi.org/10.1016/j.cedpsych.2003.08.001>
- Sarama, J., & Clements, D. H. (2007). *Classroom Observation of Early Mathematics Environment and Teaching*. Unpublished research instrument. Buffalo, NY: University at Buffalo, SUNY.
- Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology*, 101(2), 305–314. <http://dx.doi.org/10.1037/a0014217>

- Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). *Early Language & Literacy Classroom Observation Pre-k Tool*. Baltimore, MD: Paul H. Brookes.
- Smith, M. W., Dickinson, D. K., Sangeorge, A., & Anastasopoulos, L. (2002). *Early Language & Literacy Classroom Observation Toolkit: Research Edition*. Baltimore, MD: Paul H. Brookes.
- Smith, S. (2004). *Supports for Social-Emotional Growth Assessment (SSEGA)* Unpublished research instrument. New York: New York University.
- Smith, S., Davidson, S., Weisenfeld, G., & Katsaros, S. (2001). *Supports for Early Literacy Assessment (SELA)* Unpublished research instrument. New York: New York University.
- Solari, E. J., Landry, S. H., Crawford, A., Gunnewig, S., & Swank, P. R. (2009). *Bilingual-Teacher Behavior Rating Scale* Unpublished research instrument. Children's Learning Institute University of Texas Health Science Center at Houston.
- Solari, E. J., Landry, S. H., Zucker, T. A., & Crawford, A. D. (2011). The importance of sensitive measurement tools for understanding what instructional practices promote school readiness for dual language learners. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom* (pp. 45–68). Baltimore, MD: Paul H. Brookes.
- Stevenson-Boyd, J., Brenneman, K., Frede, E., & Weber, M. (2008). *Preschool Rating Instrument for Science and Mathematics* Unpublished research instrument. New Brunswick, NJ: National Institute for Early Education Research.
- Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, 19, 375–397.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years. Early Childhood Environment Rating Scale Extension (ECERS-E): Four circular sub-scales*. Stoke on Trent, UK: Trentham Books.
- Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* (Special Report No. 78.). Retrieved from Institute for Research on Poverty website <http://www.irp.wisc.edu/publications/sr/pdfs/sr78.pdf>
- Watamura, S. E., Phillips, D. A., Morrissey, T. W., McCartney, K., & Bub, K. (2011). Double jeopardy: Poorer social-emotional outcomes for children in the NICHD SECCYD experiencing home and child-care environments that confer risk. *Child Development*, 82(1), 48–65. <http://dx.doi.org/10.1111/j.1467-8624.2010.01540.x>
- Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America*. Oakland, CA: Child Care Employee Project.
- Winsler, A., Burchinal, M. R., Tien, H., Peisner-Feinberg, E. S., Espinosa, L., Castro, D. C., et al. (2014). Early development among dual language learners: The roles of language use at home, maternal immigration, country of origin, and socio-demographic variables. *Early Childhood Research Quarterly*, <http://dx.doi.org/10.1016/j.ecresq.2014.02.008> (in press)
- *Winter, S. M., Zurcher, R., Hernandez, A., & Yin, Z. (2007). The early on school readiness project: A preliminary report. *Journal of Research in Childhood Education*, 22(1), 55–68. <http://dx.doi.org/10.1080/02568540709594609>
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., & Burchinal, M. (2010). *Quality dosage, thresholds, and features in early childhood settings: A review of the literature (OPRE 2011-5)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Zill, N., Sorongon, A., Kim, K., Clark, C., & Woolverton, M. (2006). *FACES 2003 research brief: Children's outcomes and program quality in Head Start*. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.
- *Zuniga, S. A., & Howes, C. (2009). Predictions of children's experiences with Latina family child care providers. *Early Education and Development*, 20(2), 265–284. <http://dx.doi.org/10.1080/10409280802595433>