Research Report Series | Report #1

Examining the Use of Language and Literacy Assessments with Young Dual Language Learners

Final Report
June 28, 2012

Eileen Bandel
Sally Atkins-Burnett
Dina C. Castro
Claire Smither Wulsin
Marisa Putnam

Center for
Early Care and Education Research
Dual Language Learners

Suggested citation:
Bandel, E., Atkins-Burnett, S., Castro, D. C., Wulsin, C. S., & Putman, M. (July 2012). *Examining the use of language and literacy assessments with young dual language learners. Research report* #1. Center for Early Care and Education Research-Dual Language Learners (CECER-DLL). Chapel Hill: The University of North Carolina, Frank Porter Graham Child Development Institute.

Examining the Use of Language and Literacy Assessments with Young Dual Language Learners

Final Report

June 28, 2012

Eileen Bandel
Sally Atkins-Burnett
Dina C. Castro
Claire Smither Wulsin
Marisa Putnam

Contract Number: 90YR0041/03

Mathematica Reference Number: 40001

Submitted to:
Center for Early Care and Education Research—Dual Language Learners
Frank Porter Graham
Child Development Institute
The University of North Carolina
Chapel Hill, NC
Project Officer: Dina Castro

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Sally Atkins-Burnett

www.fpg.unc.edu

MATHEMATICA
Policy Research

www.mathematica-mpr.com

ACKNOWLEDGMENTS

The authors of this report would like to thank the members of CECER-DLL who were our research partners throughout this project. Margaret Burchinal, Lisa Lopéz, Vera Gutiérrez-Clellen, and Ellen Peisner-Feinberg were involved, insightful, and supportive. We would also like to acknowledge Carol Hammer for her efforts in locating the research articles reviewed in this report.

CONTENTS

I. PROCEDURES USED TO ASSESS DUAL LANGUAGE LEARNERS (DLLS)

Introduction

The percentage of young children who are from homes where a language other than English is spoken has increased dramatically in the last 10 years. Recent reports from the Family and Child Experiences Survey (FACES) 2009 cohort, the ongoing national study of children in Head Start, indicate that more than 31% of preschoolers in Head Start live in homes where a language other than English is spoken (Aikens et al., 2011). In Early Head Start nationally, almost one-third of children live in households that report speaking a language other than or in addition to English (Vogel et al., 2011).

When study samples include children who primarily speak a language other than English or who are regularly exposed to a language other than English,[1] researchers make decisions about whether to include them in the assessment and, if so, the language(s) in which they should be assessed. Approaches to determining how to assess DLLs include (1) assessing skills in more than one language, (2) using parent or teacher report of primary language to determine whether to conduct an assessment, (3) using a screening assessment such as the English and Spanish Pre-Language Assessment Scales 2000 (PreLAS; Duncan & DeAvila, 2002) to determine the most appropriate language of assessment, and (4) using conceptually scored bilingual assessments to assess children's skills independent of language. The decisions made about the approach to assessment could potentially lead to different inferences about children's development. For example, if a DLL is assessed only in English, the results may underestimate the child's knowledge and skills. Inclusion or exclusion of children from assessment could lead to different conclusions about the characteristics and skills of a population under study. Information about the prevalence of these different approaches in relation to different types of questions (and inferences) is not available.

In this study, we examine the procedures used to assess children in large-scale government-funded studies that included DLLs and in smaller studies that examined the development of language and literacy among DLLs in the United States and Canada. Three research questions guided our analysis of the procedures used to assign children to assessments. The discussion of findings related to procedures is organized according to these questions:

- What is the prevalence of different methods for assigning assessments in different languages?
- Does the procedure for determining language of assessment vary by study purpose or sample characteristics such as children's age?
- Are results similar when different methods are used to assign children's language of assessment?

Chapter II will address questions related to the evidence for reliability and validity of the measures used.

[1] Hereafter, these children will be referred to as dual language learners, or DLLs.

Method

This study drew on the critical review of the research on the language and literacy development of DLLs conducted by the Center for Early Care and Education Research: Dual Language Learners (CECER-DLL, 2011a, 2011b). This review included peer-reviewed journal articles published between 2000 and 2010 with United States, Canadian, and international samples that included at least one direct child assessment or standardized rating of the development of DLL children prior to kindergarten entry. For the purposes of the current study, we limited our review to only include studies with a sample from the United States and its territories and Canada (n = 69), thus including only studies that had English as one of the languages that DLLs are learning. Six of the studies included in the CECER-DLLs' Critical Review of the Research did not include child assessments (Hammer, Rodriguez, Lawrence, & Miccio, 2007; Kitabayashi et al., 2008; Lao, 2004; Levey & Cruz, 2003; Perry, Kay, & Brown, 2008; Shin, 2002). These studies examined parent attitudes, practices, and beliefs regarding the development of language and literacy of young DLLs using parent surveys or questionnaires (Hammer et al., 2007; Kitabayashi et al., 2008; Lao, 2004; Perry et al., 2008) or asked only for parent reports of expressive language (Levey & Cruz, 2003; Shin, 2002) and, thus, did not need to make decisions about language of assessment. These studies were excluded from further review about procedures but were retained for consideration of the psychometric properties of parent-reported measures of child development. When we began our review, we included three more recent articles identified with the same search terms as those used for the Critical Review of the Research (Bialystok, Luk, Peets, & Yang, 2009; Gildersleeve-Neumann & Wright, 2010; Parra, Hoff, & Core, 2011). We also reviewed government reports of large-scale studies of early childhood that included at least one direct assessment of children prior to kindergarten entry. With a greater number of children, large-scale studies offer more opportunity for researchers to examine reliability and validity of assessments. Among government reports published in the last 10 years, we located only seven large-scale national studies that examined children's development prior to kindergarten entry and included DLLs. Thus, our review included a total of 79 studies.

The reports for these studies are organized in the reference list. The research studies are presented first, and the government reports from the seven national studies follow. A description of the sample characteristics and study purpose for each of the studies, as well as the approach used to determine the language of assessment, can be found in the tables in Appendix A.

Findings

What is the prevalence of different methods for assigning assessments in different languages?

As noted above, researchers use a number of approaches when making determinations about the language(s) in which to assess the skills of DLL children. We elaborate on each of these in turn.

Dual-language approach. Among the studies of DLLs' language and literacy development, the majority (76% of the smaller studies and 57% of the large-scale studies

during at least one timepoint) included a dual-language approach in which DLLs were assessed in both languages for at least one area of language or literacy development, irrespective of language proficiency or dominance.[2] The most frequent area of development assessed in both the home language and in English was receptive vocabulary.

Among the 50 studies that used a dual-language approach to assessment, 13 also included subsamples of monolingual (primarily or only English-speaking, or primarily or only Spanish-speaking) children and assessed those children only in one language.[3] Approaches for classifying children into bilingual or monolingual groups varied. Some studies used parent and/or teacher reports to ascertain children's language proficiency and/or exposure to each language.[4] In other studies, bilingual and monolingual children were recruited from different locales (e.g., U.S. versus Puerto Rico, see Páez, Tabors, & Lopez, 2007; Tabors, Páez, & López, 2003; U.S. versus Mexico, see Bunta, Fabiano-Smith, Goldstein, & Ingram, 2009; Fabiano-Smith & Barlow, 2010; Fabiano-Smith & Goldstein, 2010a; Fabiano-Smith & Goldstein, 2010b).

Parent and teacher/caregiver report. Studies used parent and teacher/caregiver report of children's language in various ways, including as background information on language exposure, as an initial step, or as sole criterion in determining language of assessment. In some cases, the assessments were parent-report measures (such as the Ages and Stages Questionnaires–Third Edition [ASQ3] and MacArthur-Bates Communicative Development Inventories [CDI]), with the parent selecting the form to use and thus determining the language of assessment. Caregiver or teacher report alone was used only in the FACES 1997, FACES 2000, and the Head Start Impact Study (HSIS).

Across studies focused on DLLs as well as large-scale studies describing representative national samples, information was usually collected from parents about language used in the home in order to determine the language of assessment. This information was used to provide a means of describing the language exposure of the sample and/or to determine the language for assessment or the initial language to use in assessment for dual-language administrations (Páez et al., 2007). Some studies reported language exposure or proficiency based solely on parent report,[5] combinations of parent and teacher report,[6] combinations of parent, teacher, and assessor reports (Dickinson, McCabe, Clark-Chiarelli, & Wolf, 2004; Farver, Xu, Eppe, & Lonigan, 2006), or some other combination (e.g., parent report and PreLAS scores [Anderson, 2004]; parent-teacher reports and grammaticality on a narrative storytelling task [Bedore, Peña, Garcia, & Cortez, 2005]).

Language proficiency screening procedures. Several of the large-scale national studies included in the review assessed the skills of DLLs in a single language as

[2] See studies in reference list: 1–4, 7–22, 29–31, 33–36, 40–43, 45, 46, 49–54, 56–60, 63–66, 68, 69; 103, 106, 108, 109.

[3] See studies in reference list: 3, 4, 11, 17, 18, 20–22, 29, 52, 56, 66, 69.

[4] See studies in reference list: 3, 4, 11, 18, 20–22, 29, 66.

[5] See studies in reference list: 7, 10, 15, 29, 34–37, 56, 58, 66.

[6] See studies in reference list: 3, 11, 20–22, 30.

determined by their performance on an English language proficiency screener. For the preschool-year assessment of the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), children who did not correctly respond to at least one of the language items were routed to the Spanish assessment (Jacobson Chernoff, Flanagan, McPhee, & Park, 2007; Najarian, Snow, Lennon, & Kinsey, 2010). In FACES 2006 and FACES 2009, the PreLAS was used to determine the language of assessment for children who were from households in which English was not the primary language (Aikens et al., 2011; Hulsey, Aikens, Xue, Tarullo, & West, 2010; Malone, Hulsey, Aikens, West, & Tarullo, 2010; Moiduddin, Aikens, Tarullo, & West, 2010; West et al. 2008). Based on their performance, children were assessed in English, in Spanish, or using an abbreviated battery consisting solely of vocabulary measures, though all children received the PPVT-4 in order to assess English vocabulary. Across all of these studies, performance on the PreLAS was coupled with parent report of home language to determine the most appropriate language of administration. None of the small-scale studies included in the review used this screening approach to assessment.

Conceptually scored assessments. Five studies (Anthony et al., 2009; Bedore et al., 2005; Castilla, Restrepo, & Perez-Leroux, 2009; Dickinson et al., 2004; FACES 2009) used at least one conceptually scored bilingual assessment. In conceptually scored assessments, children are given credit for correct responses independent of the language used. However, Anthony and colleagues (2009) used the Spanish-bilingual edition to provide a measure of the Spanish receptive vocabulary (Receptive One-Word Picture Vocabulary Test: Spanish Bilingual Edition; ROWPVT-SBE, Brownell, 2001b) and expressive vocabulary (Expressive One-Word Picture Vocabulary Test: Spanish Bilingual Edition; EOWPVT-SBE, Brownell, 2001a) and also administered these assessments in English (ROWPVT and EOWPVT). Researchers used the Bilingual English Spanish Assessment (BESA; Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, in preparation) (Bedore et al., 2005; Castilla et al., 2009), the Emergent Literacy Profile (ELP; Dickinson & Cheney, 1997b) (Dickinson et al., 2004), and the EOWPVT-SBE (Aikens et al., 2011).

Does the procedure for determining language of assessment vary by study purpose or sample characteristics such as children's age?

The procedure for determining the language of assessment varied according to study question. Some studies focused on describing children's language-specific skills and development and the relationship between children's first and second language, thus necessitating a dual-language approach to assessment. Others sought to understand children's development more generally, with less regard for language. The majority of the studies focused on preschool-age children, with fewer studies examining the development of children younger than 30 months. Tables A.1 (for studies from the research brief on language and literacy development in DLLs) and A.2 (for government reports of large-scale studies) include information about sample characteristics and study purpose for each of the studies reviewed (see Appendix A).

Differences by study purpose. As noted previously, 50 of the studies used a dual-language approach to assessment in which DLLs were assessed in both languages in at least one developmental area. This approach reflects the focus of these studies on the language skills of DLLs (e.g., the contribution of abilities in the first language [L1] to the development of skills in the second language [L2; usually English], differences in skills in L1 and L2, and examination of between-group differences between DLLs' skills in L1 and L2 and that of their monolingual counterparts). Of the 50 studies that used a dual-language approach to assessment, 26% (n = 13) included subsamples of monolingual,

primarily English-speaking or primarily Spanish-speaking children whose skills were assessed solely in one language.[7]

In large-scale studies in which DLLs are not the primary focus, children were generally assessed in a single language often determined by an English language proficiency screener. Across studies, researchers most often used the Simon Says and Art Show subtests from the PreLAS. If children did not pass the English language screener, they either were not assessed or were administered an assessment in Spanish. Although the same assessment tools were often used across these studies, the threshold for determining language of assessment varied. For example, the preschool and kindergarten data collection rounds of the ECLS-B set a low (lenient) criterion for English assessment and required children to respond correctly to only one item in English (beyond the practice item) in order to receive the assessment in English (Najarian, Snow, Lennon, & Kinsey, 2010). The report indicated that a lenient criterion was selected so that the study could be as inclusive as possible in assessing children. Thus, few children were routed into a Spanish assessment and scores were not created for the Spanish assessment. In the FACES 2006 and FACES 2009 studies, respectively, the cutpoint for Spanish versus English assessments was five consecutive errors on Simon Says and Art Show, with 54% to 57% of children from Spanish-speaking homes being assessed in Spanish at the start of preschool in different years of the study (Aikens et al., 2011; Hulsey, Aikens, Xue, Tarullo, & West, 2010; Malone, Hulsey, Aikens, West, & Tarullo, 2010; Moiduddin, Aikens, Tarullo, & West, 2010; West et al. 2008). A child who made five consecutive errors on both Simon Says and Art Show and primarily spoke a language other than English or Spanish was routed out of the cognitive assessment after the administration of the vocabulary measures (37% of non-Spanish DLLs at the start of preschool).

The HSIS, a randomized controlled study of the effect of Head Start, provided more information about the development of DLLs in different languages. The initial screening involved asking the primary caregiver three questions regarding a child's language ability. If two or more of the responses were English or Spanish, the child was tested in that language. At baseline, bilingual Spanish-speaking DLLs were administered the full Spanish assessment battery and two English tests, the Peabody Picture Vocabulary Test–III (PPVT-III, Dunn & Dunn, 1997) and the WJ-III Letter-Word Identification subtest. In spring 2003, the bilingual Spanish assessment included the complete English battery, the TVIP, and the WJ-III Letter-Word Identification subtest. In Puerto Rico, all children were given Spanish assessments at all data collection times. If a child's primary language was anything other than Spanish or English, teacher report was used to decide if the child could understand the assessment in English. If he or she could not, four tests (McCarthy Draw-a-Design, Color Names and Counting, Leiter-R–adapted, and Story and Print Concepts) were translated and administered to the children.

In FACES 2006 and FACES 2009, multiple approaches to language of assessment were used in order to be able to describe more fully the language development of the growing number of DLLs in Head Start. FACES is designed to describe children in Head Start nationally (along with characteristics of their families and programs). Head Start encourages support of home language, and 80% of DLLs who entered the program in the fall of 2009 were from homes where Spanish was spoken (Aikens et al., 2011). FACES

[7] See studies in reference list: 3, 4, 11, 17, 18, 20–22, 29, 52, 56, 66, 69.

includes direct child assessments in both English and Spanish. For literacy and math assessments, FACES routed children using the PreLAS. In the area of receptive language, dual assessment was used for DLLs from Spanish-speaking households. All children, including DLLs, were assessed with the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4) in order to provide an estimate of English receptive vocabulary for all children. Spanish-speaking children, regardless of performance on the screener, were also assessed with the Test de Vocabulario en Imagenes Peabody (TVIP), to provide a measure of their receptive vocabulary in Spanish. Finally, in 2009, FACES added a conceptually scored measure of expressive vocabulary, the Expressive One-Word Picture Vocabulary Test–Spanish Bilingual Edition (EOWPVT-SBE) that was administered to all Spanish-speaking children regardless of performance on the screener. The English edition of this measure (EOWPVT) was administered to all other children. DLLs in FACES with home languages other than English or Spanish and who did not pass the English screener did not receive any additional direct assessments of language and literacy, except for the PPVT-4 and EOWPVT.

Conceptually scored bilingual assessments such as the BESA (Bedore et al., 2005; Castilla et al., 2009), the ELP (Dickinson et al., 2004), and the EOWPVT-SBE (Brownell, 2001a) were included in studies with various research purposes, including to describe the language development of DLLs, examine the influence of the first language on the development of English, examine differences in estimation of vocabulary when scoring conceptually rather than in one language, and explore the relationship between vocabulary and literacy skills for DLLs (Anthony et al., 2009; Aikens et al., 2011; Bedore et al., 2006; Castilla et al., 2009; Dickinson et al., 2004).

In some studies, DLLs were assessed either in English (Bialystok, Luk, Peets, & Yang, 2010; Bland-Stewart & Fitzgerald, 2001; Gildersleeve-Neumann, Kester, Davis, & Pena, 2008; Gildersleeve-Neumann & Wright, 2010; Goldberg, Paradis, & Crago, 2008; Hammer, Miccio, & Wagstaff, 2003; Jia, 2003; Mushi, 2002; Vagh, Pan, & Mancilla-Martinez, 2009; Yavas & Core, 2001) or in Spanish (Gildersleeve-Neumann, Peña, Davis, & Kester, 2009; Guiberson, Barrett, Jancosek, & Itano, 2006; Yaden & Tardibuono, 2004). In these studies, the study purpose and research questions (that is, examining DLLs' language abilities in English or in Spanish) determined the language of assessment.

Beyond the goal of providing a snapshot of children's knowledge and abilities upon Head Start entry, FACES was designed to describe gains in children's skills over time. Similar to other preschool studies that include assessments of DLLs, FACES allows children to change language of assessment across data collection rounds (from a non-English assessment to an English assessment) based on performance on a language screener. Reports from the study typically discuss results on literacy (and mathematics) assessments based on children who remain in the same language of assessment in fall and spring. It is more difficult to interpret results of assessments when children change language of assessment from fall to spring. This means that children who change from a Spanish to an English assessment are not represented in the mean change over time. With the majority of the instruction in the U.S. occurring in English, the results on a Spanish measure of literacy may underestimate the advances that Spanish-speaking children are making in this area. FACES assessed the English vocabulary of children at all timepoints, so researchers can examine change over time in that area and relate it to progress on literacy measures.

Differences by age. Fourteen studies included samples of children younger than 2.5 years of age. Only ECLS-B and the descriptive study of Early Head Start (Baby FACES)

assessed areas beyond language with this age group and thus had a need to determine a specific language for assessments (Administration for Children and Families, 2011; Andreassen & Fletcher, 2005; Chernoff, Flanagan, McPhee, & Park, 2007; Mulligan & Flanagan, 2006; Najarian et al., 2010). Other studies of very young children collected language samples or used parent reports of vocabulary as the focus of the study. Fewer measures of early development are available in more than one language when compared to assessments available for use with preschool-aged children. With the exception of the CDI/Inventario and the ASQ-3, the early assessments used in other languages were translations of the English assessments completed by the study researchers.

The 9-month and 2-year assessments of the ECLS-B included English and Spanish versions of the direct assessments (videotapes of the Two-Bag Task were also coded in Mandarin). The procedure for determining whether to assess children in English or in Spanish is not described in the psychometric report, but apparently depended on parent report. This is in contrast to the preschool and kindergarten ECLS-B data collection timepoints that used one correct question on the English PreLAS as the criterion for conducting the assessment in English.

In Baby FACES, assessments for the 1-year-old cohort included parent- and teacher-report measures and videotaped parent-child interactions. The parent- and teacher-report measures, including the CDI/Inventario and the ASQ-3 Communication, were available in both Spanish and English. Teachers completed the Spanish forms only if both they and the child understood or spoke Spanish.

The remaining studies of infants and toddlers (n = 12) collected language samples using various approaches including audiotape and videotape of natural conversations (studies included DLLs from multiple linguistic backgrounds), or parent reports—usually the CDI/Inventario. In most cases, the parents completed word lists for the child's predominant language (as selected by the parent) or completed both the English and home language form. The parent reports of child language were a primary source of data or additional information to examine outcomes. For example, Guiberson and colleagues (2006) used a modified version of the parent-report Bilingual Language Proficiency Questionnaire (BFQ), but as background information rather than selection criteria, as they examined the loss of Spanish using language samples and the Spanish Language Assessment Procedure (SLAP). Conboy and Mills (2006) used imaging of brain activity in response to a word list derived from the CDI.

Are results similar when different methods are used to assign children's language of assessment?

Available evidence suggests that different methods for assigning children to a single language of assessment will affect the findings. In the recent FACES 2009 report, information about the performance of 693 DLLs is presented separately according to whether they did or did not pass the language screener (see Table B.10 in Aikens et al., 2011). As noted previously, the PPVT-4 was administered to all children. The EOWPVT-SBE and TVIP were administered to all children having Spanish as a home language, regardless of whether or not they passed the English language screener. The mean scores on the TVIP were similar for those who did and did not pass the English language screener (standard scores = 83.1 and 84.8, respectively). As expected, the children from Spanish-speaking homes who passed the screener scored better on the PPVT-4 than those who did not (standard scores = 76.4 and 63.0, respectively), and they also scored better on the EOWPVT-SBE (standard scores = 98.3 and 86.3, respectively). The picture

presented about the language development of these children differs somewhat by measure; different routing procedures would make different assumptions about children's language skills and would assign children to different measures.

Using methods other than direct measurement of English language skills can place children into a language of assessment with different levels of accuracy. When researchers examined relations between parent and teacher report and direct assessments of child outcomes, stronger relations (usually bivariate correlations) were found with the parent report than with the teacher report (Vagh et al., 2009). This study included both reports of language use at home and use of a standardized measure (CDI). For young DLL children, parents appeared to be better reporters of children's language exposure and vocabulary than teachers or other caregivers outside of the home.

FACES 1997[8] and the HSIS noted differences between the mainland U.S. and the Puerto Rican samples in language use and language of instruction. In the HSIS, despite an initial design that would include the sample from Puerto Rico as part of the national sample, researchers analyzed the Spanish monolingual sample from Puerto Rico separately. The children from Puerto Rico continued to be monolingual Spanish, received instruction entirely in Spanish, and were assessed exclusively in Spanish throughout the study period. Conducting a separate analysis on this smaller sample size limited the power to detect differences between the Head Start and control groups. This has implications for study designs that combine monolingual samples from countries and territories outside the U.S. mainland with samples from within the U.S.

Because the majority of the studies used dual-language administration, the summary of results of language development included in the brief Language and Literacy Development in Dual Language Learners: Annotated Bibliographies from a Critical Review of the Research (CECER-DLL, 2011a) describe the findings regarding associations between development in the home language, and language and literacy outcomes in English. We do not duplicate that discussion here.

Summary of Approaches to Assessing DLLs

Studies used a variety of approaches to assessing DLLs, including dual-language assessment, parent and teacher report, language screening assessments, conceptually scored assessments, and language samples. The selected approach is often determined by the study question but is also influenced by other study constraints. Large-scale studies attempt to collect as much information as possible as economically as possible. Fewer measures are available for DLLs under the age of 3 years. As the percentage of DLLs has increased, more careful attention has been given to how to collect valid information that portrays DLLs' development. More recent studies utilize conceptually scored approaches and/or dual-language assessment in at least one domain. The language used for assessment can affect conclusions drawn about the skills, knowledge, and abilities of young DLLs.

[8] After FACES 1997, Puerto Rico was no longer included in the sampling frame for FACES and these findings cannot be evaluated with more recent FACES data.

II. EVIDENCE OF RELIABILITY AND VALIDITY

In addition to examining the procedures used in assessing DLLs, we also examined the psychometric evidence provided for assessments conducted with these samples of DLLs. We limited our analysis of these results to studies that had more than 25 children in the sample and excluded researcher-developed study-specific measures and studies that used only language samples. More than half of the research articles (58%) collected language samples or used a researcher-developed measure with a small sample. In this section, we summarize the evidence of reliability and validity of the assessments reported in the remaining 30 research articles and the reports of the 7 government studies. Appendix B discusses the available evidence by measure for vocabulary assessments, language assessments, and literacy assessments included in the studies. The tables in Appendix C summarize the evidence of each measure by the specific study so that the reader can consider the evidence relative to the sample descriptions found in Appendix A.

Reliability informs consumers of research about the trustworthiness of the findings. There are several different indicators of reliability. The most commonly reported is internal consistency, that is, how consistently the items within an assessment measure the construct of interest. Particularly for measures that involve inferences in scoring, the inter-rater reliability of the assessors (that is, whether assessments administered by different assessors would result in the same score) is an important consideration. The stability of a score, or test-retest reliability, indicates whether assessment of a particular construct would result in the same score if repeated a week later or, for more stable constructs, months or years later. Higher coefficients on each of these indicate more positive evidence, with expected ranges between 0.65 and 1.0.

Evidence of validity provides information about the appropriateness and utility of the specific inferences that can be derived from assessment results. Is the assessment measuring what it purports to measure? Validation of an assessment is supported by the accumulation of evidence about the assessment in relation to different types of inferences. An assessment may be valid for a particular purpose or group of children, but not valid or representative of the skills of another group of children. The key issue when using an assessment is whether it really measures what it purports to measure for the particular use of that assessment, including examination of use with samples that are representative of the respondents. For the purposes of this review, we looked for evidence about whether the assessments used with DLL children measure the skills of linguistically diverse children in the intended way, and the types of inferences about children's development that could be drawn from those assessments. For many assessments, use is supported only for children with similar demographic and linguistic backgrounds. Other assessments (such as the Expressive One-Word Picture Vocabulary–Spanish Bilingual Editions) demonstrated similar item functioning across Spanish- and English-dominant samples during the development phase, allowing for use across linguistic groups.

Information about the validity of assessments allows for a better understanding of the inferences made from them and provides evidence for the appropriateness of their use in comparing different demographic groups within a single language or across linguistic groups.

There are multiple sources of bias that could render an assessment unfair or invalid for a particular group of children.

Different types of evidence are collected to support interpretation of the validity of an assessment and results obtained from that assessment. The normative group used for an assessment is a key component. Information about study samples contained in Table A.1 and the government studies described in Table A.2 indicates the representation of children from different age, language, and socioeconomic groups. An assessment may be valid for comparing the skills of children from the same linguistic group but may be biased when drawing comparisons across linguistic groups. For example, if the purpose is to assess whether a child is learning English, then an English vocabulary or language assessment would be a valid indicator of how much the child understands and communicates in English. In contrast, using norms based on children who are monolingual English speakers to determine if a DLL child's development in English suggests the presence of a language disability would not be a valid use of the assessment.

If you want to assess a child's conceptual vocabulary, that is, whether a child has words for different objects, actions, and concepts, then assessing a DLL in a single language would underrepresent the words that the child knows. Children typically acquire words for objects and activities experienced at home in the language used most often at home, and acquire words for academic concepts in the language used in school. Using one of the most commonly used measures of English vocabulary (PPVT-III), Bialystok, Luk, Peets, and Yang (2010) noted that items that referred to home objects and activities were more difficult for Spanish-dominant children than for English-dominant children, while school-related words showed similar difficulty across groups. This can have implications for the assessment of young children beyond the correctness of individual items when the assessments are adaptive using start and stop rules. Usually, names for items found in the home are easier for young children and so these are presented early in the assessments of vocabulary. When assessments are adaptive, the items are generally ordered in terms of difficulty based on the responses of the normative sample and the stop rules are designed so that children would have a very low probability of getting any items beyond that point correct; the scoring assumes that everything beyond that point is incorrect. However, if children are DLLs, they may know the English names for objects and activities related to school and academics, but not know the English words for home activities and objects. Using the published ceiling rules could result in underestimates of children's English vocabulary. None of the reviewed studies examined the appropriateness of basal and ceiling rules (start and stop rules) for young DLLs.

Some of the studies used measures that are conceptually scored. In measuring children's knowledge of concepts rather than vocabulary in a particular language, conceptual scoring would usually present a more valid assessment of children's knowledge. However, for receptive vocabulary, comparisons across linguistic groups can be more challenging. If you present four pictures and name an item in one language, there is a 1 in 4 chance of success. If the child is unsuccessful and a prompt is given in the other language, there is now a 1 in 3 chance of selecting the correct picture. Thus, the probability of success favors children who know two languages, compared with children who know only one. This is not a problem for expressive vocabulary, as there are a large number of words that a child could draw upon to name a picture.

Analysis and Reporting of Psychometric Evidence

Next, we examined the available evidence supporting the use of the assessments with children from diverse linguistic backgrounds. The review focused on the following research questions:

- When compared with estimates from nationally representative samples, how reliable are assessments of language and literacy when used with samples of DLLs? Does this differ across studies that examine different constructs, or across different demographic groups (based on age, socioeconomic status, or cultural backgrounds)?
- Do studies of DLLs include evidence of construct validity?
- What evidence is documented for the concurrent and predictive validity of measures used in these studies? What information do authors provide about the psychometric properties of assessments used to examine the language and literacy development of children who are DLLs?

When compared with estimates from nationally representative samples, how reliable are assessments of language and literacy when used with samples of DLLs? Does this differ across studies that examine different constructs, or across different demographic groups (based on age, socioeconomic status, or cultural backgrounds)?

Many studies did not provide any study-specific evidence of reliability, but only reported the published evidence from the assessment manual. When reported, the estimates of internal consistency (that is, whether the items within a measure were measuring the same concept) were generally favorable (.80). The most notable exception to this is the Spanish version of the Story and Print Concepts used in the FACES 2000 and FACES 2006 studies, where Cronbach's alphas were much lower (< .60). Alphas for some of the Woodcock-Muñoz measures were also below 0.80 in FACES 2006 and FACES 2009, as well as for some of the ASQ measures in Baby FACES. Baby FACES, FACES 2000, FACES 2006, and FACES 2009 reported reliability estimates for measures separately for different language groups. In FACES 2006 and FACES 2009, the estimates were typically stronger for children assessed in English than for those assessed in Spanish. The estimates were similar across groups in Baby FACES.

Test-retest reliability within the same time period was reported only for subtests of the Bilingual English Spanish Oral Language Screener (BESOS; Peña, Bedore, Gutierrez-Clellen, Iglesias, & Goldstein, in preparation). When examined with a small sample of DLLs (n = 20), the BESOS Semantics and Morphosyntax[9] subtests had stronger test-retest reliability for the Spanish version than for the English one for both subtests (Semantics r = .70 for Spanish and .64 for English; Morphosyntax r = .86 for Spanish and .75 for English).

Stability across longer periods of time (more than three months) was examined for several measures of vocabulary and literacy. Literacy measures (including the EPAP, with

[9] See glossary in Appendix E for definitions.

estimates ranging from .32 to .41, and the TERA-2, with estimates ranging from .34 to .53) generally had weaker stability than vocabulary measures (with estimates for the PPVT-III in the preschool years ranging from .48 to .75), but both the length of time between assessments and the sample varied. As expected, when samples include children who were monolingual Spanish along with simultaneous DLLs at the start of preschool, the correlation between administrations is lower than for samples with all DLLs.

Do studies of DLLs include evidence of construct validity?

Evidence of construct validity was present in the inclusion of related measures (such as vocabulary and broader language measures) in latent trait models in two studies of DLLs. Hammer, Lawrence, and Miccio (2007) used principal component analysis (PCA) to combine the PPVT-III/TELD-3 and TVIP/Spanish PLS-3 for measure of overall receptive language ability in English and Spanish, respectively. Findings of the PCA revealed that the first component for the English language measure captured an average 97% of the variance (ranging from 96% to 99% over four measurement occasions); the first component for the Spanish language measure captured an average 95% of variance (ranging from 93% to 97%). However, the study authors provide limited information about the derived latent trait model, including the unique contribution of each component measure on the latent trait (that is, the amount of variance each measure alone explained, above and beyond the other). Rinaldi and Páez (2008) reported low predictive relations between individual English subtests of vocabulary and first-grade reading in English; however, inclusion of Spanish subtests of vocabulary in the model bolstered the amount of explained variance in children's English-reading skills.

What evidence is documented for the concurrent and predictive validity of measures used in these studies?

For assessments of children's language and literacy development,[10] we describe the available evidence of reliability and validity found with the study samples in Appendix C (e.g., correlations with other assessments of language and literacy, and associations with sample characteristics).

The most commonly provided type of evidence for measures used in examining the language and literacy development of DLLs was evidence of a relationship with child's age or exposure to English. Mean scores are presented at different ages as evidence of the increase in scores across time. Correlations with age ranged from .25 to .54. Correlations with parent-reported exposure to English at the time of assessment ranged from .36 to .72 for English assessments, with absolute value of reported correlations ranging from .33 to .57 for Spanish assessments.

For bilingual samples, researchers reported correlations between Spanish and English versions of assessments for the BESA (Bunta et al., 2009; Castilla et al., 2009; r ranges from .45 to .79) and the EPAP (Mishina-Mori, 2005; r = .29 to .48).

Researchers also examined the relations among measures of vocabulary, language, and literacy, looking at both bivariate correlations and hierarchical regressions.

[10] We excluded researcher-developed assessments used in single studies.

Vocabulary and language measures generally had moderate to strong correlations (r = .46 to .79) with measures in the same language. Parent reports of vocabulary and parent-teacher composites were related significantly with direct assessments of vocabulary, but teacher reports of the vocabulary of DLLs were not significantly related to direct assessments of either English or Spanish vocabulary. The absolute value of correlations between vocabulary and direct assessments of literacy ranged from .32 to .72. Farver et al. (2006) found correlations between vocabulary (PPVT-R/TVIP) and parent education (r = .28 to .31), children's interest in literacy (r = .38) and the parent's literacy involvement (r = .25).

Correlations of measures of literacy to other literacy measures in the same language ranged from .55 to .67. Correlations of literacy measures across languages ranged from .40 to .67 for measures of similar literacy constructs, and a Korean version of the CTOPP was related strongly to the original CTOPP (r = .84).

Predictive validity evidence in these studies usually examined whether children's scores increased across timepoints and whether vocabulary and language assessments predicted later literacy. Rinaldi and Páez (2008) found that scores on the preschool Woodcock Language Proficiency Battery (WLPB) English and Spanish picture vocabulary, English WLPB Memory for Sentences, and Spanish Word reading explained 31% of the variance in first-grade English word reading. Interestingly, the English Test of Early Reading Ability-2 (TERA-2) was significantly related to a Spanish vocabulary and language composite (TVIP/Spanish PLS-3) (r = -.33 to -.48), but not to the English vocabulary and language composite (PPVT/TELD) in a sample of 72 Spanish-English bilingual children (Hammer et al., 2009).

None of the reviewed studies provided estimates of validity separately for monolingual children and DLLs. The sample sizes were often too small for separate subgroup analyses of validity. Other than the government studies, only three of the research studies had sample sizes greater than 205 (Bialystok, Luk, Peets, & Yang, 2010; Tabors, Páez, & López, 2003; Rinaldi & Páez, 2008). Bialystok and colleagues (2010) had a sample size of 1,788 children ages 3 to 10 years and compared the performance of monolingual English and bilingual children at each of the ages. They found significant effects of age and language but no interaction of age and language; monolingual children outperformed bilingual ones at every age. Rinaldi and Páez (2008) studied 234 Spanish-English bilingual children. Tabors et al. (2003) studied both a bilingual sample and a monolingual Spanish sample from Puerto Rico, but examined correlations between early language and literacy measures with the full sample of 4-year-olds (and compared mean performance by subsample). With the exception of Memory for Sentences, Spanish-English bilinguals scored consistently higher than monolingual Spanish-speaking 4-year-olds from Puerto Rico.

Limitations

Most of the studies reviewed in this report included only low-income Spanish-English DLLs, limiting the generalizability of the findings to other groups of DLLs. However, nationally, the majority of young DLLs reside in homes with limited income and have Spanish as a home language (Shin & Kominski, 2010).

Discussion

The reviewed studies add to the evidence of validity for different assessments in answering questions about DLLs. Reported internal consistency estimates were often somewhat weaker for Spanish assessments, but were usually within acceptable ranges when more than 15 items were administered. For example, in fall FACES 2009, the Cronbach's alpha estimates for the WM-III Letter-Word Identification and Spelling subtests were low (.67 and .66, respectively) but these estimates were based on only 14 administered items compared to the 17–26 items administered to the children taking the assessments in English (including some DLLs), with internal consistency ranging from .79 to .81. Similarly, the range of reported concurrent validity coefficients was weaker in strength when compared to coefficients found across measures in studies of young monolingual English samples,[11] although most concurrent coefficients were in the moderate to moderately high range.

Available evidence of predictive validity is scarce. The report that included information about the predictive validity of the measures used in FACES 1997 (Zill et al., 2003) only included children who took the tests in English at each timepoint and did not clearly indicate how many of them had Spanish as a home language. The one study that reported analysis of predictive relations to first-grade reading in English (Rinaldi & Páez, 2008) indicated low relations for individual subtests, but a combination of several subtests across both Spanish and English increased the amount of explained variance in English reading.

Many factors can affect the strength of concurrent and predictive validity coefficients of early childhood measures. Typically, the younger the child is at the initial assessment, the less reliable the measure; accordingly, associations to other measures, both within and across time, tend to be weak in magnitude. Changes in the mode of assessment over time (e.g., moving from parent-reported to direct assessments of language comprehension), as well as a greater time period between assessments (e.g., 4 months versus one year) also contribute to weaker observed associations between assessments. When the constructs measured by two different assessments tap different areas (for example, phonemic awareness and knowledge of grammar), the strength of the coefficient will be weaker. Early childhood assessments among English monolingual samples typically demonstrate low to moderate predictive correlation coefficients, with less than 25% of the overall variance in early academic performance predicted from any single preschool measure (LaParo & Pianta, 2000). Kim and Suen (2003) used Hierarchical Linear Modeling (HLM) to perform a "validity generalization study" of 716 predictive correlation coefficients from 44 studies of early assessments to later achievement or success in school. The predictive coefficients of the tests in their study ranged from .12 to .81 and differed by study for similar assessments. They concluded that "the predictive power of any early assessment from any single study is not generalizable, regardless of design and quality of research. The predictive power of early assessments is different from situation to situation" (p. 561). When samples include DLLs, the number of

[11] An example of the high end of associations found between two different English language assessments includes the PLS-5 and the CELF P-2, with r = .79 for total scores and r = .82 for expressive language scores for each. Similarly, correlations of two English vocabulary assessments, the PPVT-4 (receptive vocabulary) with the Expressive Vocabulary Test-Second Edition (EVT-2), ranged from .80 to .84.

additional variables that can affect the strength of the coefficient increases, for example, the age of introduction to the language used, the amount of exposure to the language of assessment, and intervention or preschool experiences.

Assessing DLLs requires careful consideration and care in selecting the instruments and methods that match the question to be answered. When children come from multiple linguistic backgrounds, researchers must consider whether the methods and items will fairly represent the child's knowledge, skills, and behaviors. Even within a linguistic group, difference in dialect may bias results unless accounted for by the assessment. The most reliable and valid source of information about children's skills and development may differ for children who are DLLs. For DLLs, parents may be better sources of information about children's vocabulary than teachers. Teachers may not have knowledge of children's vocabulary in languages used at home. Among DLLs, stronger relations are found between parent reports and direct assessments of children's vocabulary than between teacher reports and direct assessments of children's vocabulary skills.

Researchers should use care when discussing study results. When comparing groups of children, information about differences in opportunity to learn (due to different socioeconomic, cultural, or linguistic backgrounds) should be included in considering differences in performance and progress. When discussing standard scores, researchers should help readers understand the similarities and differences between the normative group and the study sample. Even for the same sample of children, mean standard scores vary across measures of the same construct. The information in FACES 2009 (Aikens et al. 2011) provides the clearest picture of potentially different interpretations of DLLs' knowledge of vocabulary and concepts depending on the assessment and norm group used for standard scores. FACES 2009 assessed DLLs' vocabulary with both receptive English vocabulary (PPVT-4) and expressive conceptual vocabulary (using the EOWPVT-SBE) and provided standard scores for 4-year-olds based on the PPVT-4, the TVIP, the English version of the EOWPVT, and the Spanish bilingual (EOWPVT-SBE) version. The EOWPVT-SBE is based on a nationally representative sample of Spanish dominant and bilingual children who come from homes with limited maternal education. The PPVT-4 and the EOWPVT standard scores are based on a nationally representative sample of English-speaking children. Four-year-old DLLs in FACES 2009 who were routed into fall assessments in Spanish had mean standard scores of 56.3, 66.3, and 86.3 on the PPVT-4, the EOWPVT, and the EOWPVT-SBE, respectively. English-speaking 4-year-old children in FACES 2009 had standard scores of 86.2 and 81.3 on the PPVT-4 and the EOWPVT, respectively. The standard scores range from a mean approximately one standard deviation below the national mean for the scores based on the Spanish bilingual sample to almost three standard deviations below the national mean for the PPVT-4. This highlights the difficulty in interpreting children's skills without information about the assessments and the normative samples used to generate the standard scores.

Recommendations

- **Researchers need to consider if assessments are valid for the DLL children in their sample**. Assessing DLLs requires careful consideration and care in selecting the instruments and methods that match the question to be answered. When samples include children from multiple linguistic backgrounds, researchers must consider whether the methods and items will fairly represent all children's knowledge, skills, and behaviors. Even within a

single linguistic group, differences in dialect may bias results unless accounted for by the assessment.

- **In making determinations about the most appropriate language in which to assess DLLs, researchers need to place the goal of the assessment at the forefront**. It is important to distinguish between measurement of static knowledge in a particular area versus change in knowledge over time. For example, an evaluation of the effectiveness of a curriculum aimed at enhancing children's letter knowledge in English necessitates the assessment of these skills in English at both points in time, even for DLLs who may have limited English proficiency. Assessments of DLLs' letter knowledge in Spanish would not be a valid indicator of gains in letter knowledge over time, given that the language of instruction is English.

- **Researchers should report sample-specific estimates of reliability.** More information is needed about the performance of measures with DLLs. Poor reliability can limit the ability to detect associations between constructs, particularly when the sample size is small, and can lead to flawed conclusions about the relatedness of measures among DLLs.

- **More information is needed about expected performance of DLLs on assessments**. In order to provide information about the expected performance of DLLs on language measures, measures developers should provide supplemental norms for DLLs or estimates of the mean and standard deviation for the subsample of DLLs.

- **Information on the background characteristics of study participants is necessary when making interpretations about the performance of DLLs**. Beyond describing the knowledge and skills of DLL children, researchers need to report more information about the study characteristics of their DLL sample, particularly in large-scale studies in which DLLs are not the primary focus. When conducting research with samples of monolingual and DLL children, factors including socioeconomic background, age ranges represented, and, when available, differences in ethnic and cultural background are critical to contextualizing differences in performance. In the absence of this information, caution should be exercised when making interpretations, given that such characteristics may account for observed differences.

- **More information is needed about the equivalence of measures for different language groups**. Researchers should consider whether the selected assessment(s) are biased in any way. Do the tasks or items require similar levels of skill across languages and cultures? Is the task equally representative of skills across different groups? For example, rhyming is usually easier for young children when the words are only one syllable, but English has many more one-syllable rhymes than Spanish. Limited evidence was provided for the congruence of estimates of item difficulty across languages.

- **More evidence is needed for the predictive validity of early measures for later outcomes when used with DLLs.** Most of the evidence of validity of the measures was found with samples of children who were able to take assessments in English, and the DLLs were combined with English-only speakers. Separate analyses with DLLs are needed, as is more information about the Spanish versions of assessments.

REFERENCES

Research Studies[12]

[1]Anderson, R. T. (2004). Phonological acquisition in preschoolers learning a second language via immersion: A longitudinal study. *Clinical Linguistics & Phonetics, 18*(3), 183–210.

[2]Anthony, J. L., Solari, E. J., Williams, J. M., Schoger, K. D., Zhang, Z., Branum-Martin, L., & Francis, D. J. (2009). Development of bilingual phonological awareness in Spanish-speaking English language learners: The roles of vocabulary, letter knowledge, and prior phonological awareness. *Scientific Studies of Reading, 13*(6), 535–564.

Anthony, J. L., Williams, J. M., McDonald, R., Corbitt-Shindler, D., Carlson, C. D., & Francis, D. J. (2006). Phonological processing and emergent literacy in Spanish speaking preschool children. *Annals of Dyslexia, 56*, 239–270.

Anthony, J. L., Williams, J. M., McDonald, R., & Francis, D. J. (2007). Phonological processing and emergent literacy in younger and older preschool children. *Annals of Dyslexia, 57*, 113-137.

[3]Bedore, L. M., Fiestas, C. E., Peña, E. D., & Nagy, V. J. (2006). Cross-language comparisons of maze use in Spanish and English in functionally monolingual and bilingual children. *Bilingualism: Language and Cognition, 9*(03), 233–247.

[4]Bedore, L. M., Peña, E. D., Garcia, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language, Speech, and Hearing Services in Schools, 36*(3), 188-200.

[5]Bialystok, E., Luk, G., Peets, K.F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525–531.

[6]Bland-Stewart, L. M., & Fitzgerald, S. M. (2001). Use of Brown's 14 grammatical morphemes by bilingual Hispanic preschoolers: A pilot study. *Communication Disorders Quarterly, 22*(4), 171–186.

[7]Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism, 13*(3), 325–344.

[8]Brice, A. E., Carson, C. K., & O'Brien, J. D. (2009). Spanish–English articulation and phonology of 4- and 5-year-old preschool children: An initial investigation. *Communication Disorders Quarterly, 31*(1), 3–14.

[12] The studies included in this review are accompanied by a superscripted numerical value.

Brownell, R. (2001a). Expressive One-Word Picture Vocabulary Test (Spanish-Bilingual edition, manual). Novato, CA: Academic Therapy Publications.

Brownell, R. (2001b). Receptive One-Word Picture Vocabulary Test (Spanish-Bilingual edition, manual). Novato, CA: Academic Therapy Publications.

Brownell, R. (2001). Expressive One-Word Picture Vocabulary Test. San Antonio, TX: Harcourt Assessment, Inc.

[9]Buckwalter, J. K., & Gloria Lo, Y. H. (2002). Emergent biliteracy in Chinese and English. *Journal of Second Language Writing, 11*(4), 269–293.

[10]Bunta, F., Fabiano-Smith, L., Goldstein, B., & Ingram, D. (2009). Phonological whole-word measures in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics & Phonetics, 23*(2), 156–175.

[11]Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research, 50*(4), 999.

[12]Castilla, A. P., Restrepo, M. A., & Perez-Leroux, A. T. (2009). Individual differences and language interdependence: A study of sequential bilingual development in Spanish–English preschool children. *International Journal of Bilingual Education and Bilingualism, 12*(5), 565–580.

[13]Castro, D., & Gavruseva, E. (2003). Finiteness and aspect in Spanish/English bilingual acquisition. *First Language, 23*(2), 171–192.

Center for Early Care and Education Research–Dual Language Learners (CECER-DLL; 2011). Research brief #6: Language and literacy development in dual language learners: A critical review of the research. Chapel Hill: The University of North Carolina, FPG Child Development Institute, Author.

Center for Early Care and Education Research–Dual Language Learners (CECER-DLL; 2011). Language and literacy development in dual language learners: Annotated bibliographies from a critical review of the research. Chapel Hill: The University of North Carolina, FPG Child Development Institute, Author.

[14]Chung, H. H. (2006). Code switching as a communicative strategy: A case study of Korean–English bilinguals. *Bilingual Research Journal, 30*(2), 293–307.

[15]Conboy, B. T., & Mills, D. L. (2006). Two languages, one developing brain: Event-related potentials to words in bilingual toddlers. *Developmental Science, 9*(1), F1–F12.

[16]Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development, 77*(3), 712–735.

[17]Davidson, D., Raschke, V. R., & Pervez, J. (2010). Syntactic awareness in young monolingual and bilingual (Urdu–English) children. *Cognitive Development, 25*(2), 166–182.

DeAvila, E. A., & Duncan, S. E. (2000). PreLAS 2000 technical manual. Monterey, CA: CTB/McGraw-Hill.

Dickinson, D. K., & Chaney, C. (1997a). Early Phonological Awareness Profile. Newton, MA: Education Development Center.

Dickinson, D. K., & Chaney, C. (1997b). Emergent Literacy Profile. Newton, MA: Education Development Center.

[18]Dickinson, D. K., McCabe, A., Clark–Chiarelli, N., & Wolf, A. (2004). Cross-language transfer of phonological awareness in low-income Spanish and English bilingual preschool children. *Applied Psycholinguistics, 25*(03), 323–347.

Duncan, S., & DeAvila, E. (2002). Pre-Language Assessment Scales [PreLAS 2000]. Monterey, CA: CTB-McGraw Hill.

[19]Fabiano, L., & Goldstein, B. (2005). Phonological cross-linguistic effects in bilingual Spanish–English speaking children. *Journal of Multilingual Communication Disorders, 3*(1), 56–63.

[20]Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism, 13*(1), 81–97.

[21]Fabiano-Smith, L., & Goldstein, B. A. (2010a). Early-, middle-, and late-developing sounds in monolingual and bilingual children: An exploratory investigation. *American Journal of Speech-Language Pathology, 19*(1), 66.

[22]Fabiano-Smith, L., & Goldstein, B. A. (2010b). Phonological acquisition in bilingual Spanish–English speaking children. *Journal of Speech, Language, and Hearing Research, 53*(1), 160.

[23]Farver, J. A. M., Xu, Y., Eppe, S., & Lonigan, C. J. (2006). Home environments and young Latino children's school readiness. *Early Childhood Research Quarterly, 21*(2), 196–212.

[24]Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development, 78*(5), 1510–1525.

[25]Gildersleeve-Neumann, C. E., Kester, E. S., Davis, B. L., & Peña, E. D. (2008). English speech sound development in preschool-aged children from bilingual English–Spanish environments. *Language, Speech, and Hearing Services in Schools, 39*(3), 314.

[26]Gildersleeve-Neumann, C. E., Peña, E. D., Davis, B. L., & Kester, E. S. (2009). Effects on L1 during early acquisition of L2: Speech changes in Spanish at first English contact. *Bilingualism: Language and Cognition, 12*(02), 259–272.

[27]Gildersleeve-Neumann, C. E., & Wright, K. L. (2010). English speech acquisition in 3- to 5-year-old children learning Russian and English. *Language, Speech, and Hearing Services in Schools, 41*(4), 429.

[28]Goldberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second. *Applied Psycholinguistics, 29*(1), 41–65.

[29]Goldstein, B. A., Fabiano, L., & Washington, P. S. (2005). Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish–English bilingual children. *Language, Speech, and Hearing Services in Schools, 36*(3), 201.

[30]Goldstein, B., & Washington, P. S. (2001). An initial investigation of phonological patterns in typically developing 4-year-old Spanish–English bilingual children. *Language, Speech, and Hearing Services in Schools, 32*(3), 153.

[31]Gonzalez, J. E., & Uhing, B. M. (2008). Home literacy environments and young Hispanic children's English and Spanish oral language. *Journal of Early Intervention, 30*(2), 116–139.

[32]Guiberson, M. M., Barrett, K. C., Jancosek, E. G., & Itano, C. Y. (2006). Language maintenance and loss in preschool-age children of Mexican immigrants longitudinal study. *Communication Disorders Quarterly, 28*(1), 4–17.

[33]Hammer, C. S., Davison, M. D., Lawrence, F. R., & Miccio, A. W. (2009). The effect of maternal language on bilingual children's vocabulary and emergent literacy development during Head Start and kindergarten. *Scientific Studies of Reading, 13*(2), 99–121.

[34]Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2007). Bilingual children's language abilities and early reading outcomes in Head Start and kindergarten. *Language, Speech, and Hearing Services in Schools, 38*(3), 237.

[35]Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008a). The effect of summer vacation on bilingual preschoolers' language development. *Clinical Linguistics & Phonetics, 22*(9), 686–702.

[36]Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2008b). Exposure to English before and after entry into Head Start: Bilingual children's receptive language growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism, 11*(1), 30–56.

[37]Hammer, C. S., Miccio, A. W., & Wagstaff, D. A. (2003). Home literacy experiences and their relationship to bilingual preschoolers' developing English literacy abilities: An initial investigation. *Language, Speech, and Hearing Services in Schools, 34*(1), 20.

[38]Hammer, C. S., Rodriguez, B. L., Lawrence, F. R., & Miccio, A. W. (2007). Puerto Rican mothers' beliefs and home literacy practices. *Language, Speech, and Hearing Services in Schools, 38*(3), 216.

[39]Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research, 46*(6), 1297.

[40]Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism, 6*(3), 271–289.

[41]Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: Lexical-semantic skills in L1 and L2. *Journal of Speech, Language, and Hearing Research, 48*(2), 372.

[42]Kan, P. F., & Kohnert, K. (2008). Fast mapping by bilingual preschool children. *Journal of Child Language, 35*(3), 495.

Kim, J., & Suen, H. K. (2003). Predicting children's academic achievement from early assessment scores: A validity generalization study. *Early Childhood Research Quarterly, 18*, 547–566.

[43]Kim, Y. S. (2009). Crosslinguistic influence on phonological awareness for Korean-English bilingual children. *Reading and Writing, 22*(7), 843–861.

[44]Kitabayashi, K. M., Huang, G. Y., Linskey, K. R., Pirga, J., Bane-Terakubo, T., & Lee, M. T. (2008). Parent-child reading interactions among English and English as a second language speakers in an underserved pediatric clinic in Hawai'i. *Hawaii Medical Journal, 67*(10), 260–263.

[45]Kohnert, K., Kan, P. F., & Conboy, B. T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *Journal of Speech, Language, and Hearing Research, 53*(3), 684.

[46]Kyratzis, A., Tang, Y., & Bahar Koymen, S. (2009). Codes, code-switching, and context: Style and footing in peer group bilingual play. *Multilingua, 28*(2–3), 265–290.

LaParo, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years. A meta-analytic review. *Review of Educational Research*, *70* (4), 443–484.

[47]Lao, C. (2004). Parents' attitudes toward Chinese–English bilingual education and Chinese-language use. *Bilingual Research Journal, 28*(1), 99–121.

[48]Levey, S., & Cruz, D. (2003). The first words produced by children in bilingual English/Mandarin Chinese environments. *Communication Disorders Quarterly, 24*(3), 129–136.

[49]López, L. M., & Greenfield, D. B. (2004). The cross-language transfer of phonological skills of Hispanic Head Start children. *Bilingual Research Journal, 28*(1), 1–18.

[50]Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish-English bilinguals. *Journal of Child Language, 37*(04), 817–840.

[51]Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science, 7*(2), 212–224.

[52]Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin-English bilingual children. *Clinical Linguistics & Phonetics, 24*(4–5), 387–400.

[53]Mishina-Mori, S. (2005). Autonomous and interdependent development of two language systems in Japanese/English simultaneous bilinguals: Evidence from question formation. *First Language, 25*(3), 291–315.

[54]Morita, E. (2003). Children's use of address and reference terms: Language socialization in a Japanese–English bilingual environment. *Multilingua, 22*(4), 367–395.

[55]Mushi, S. (2002). Acquisition of multiple languages among children of immigrant families: Parents' role in the home-school language pendulum. *Early Child Development and Care, 172*(5), 517–530.

[56]Páez, M. M., Tabors, P. O., & López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *Journal of Applied Developmental Psychology, 28*(2), 85–102.

[57]Paradis, J., & Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language, 30*(02), 371–393.

[58]Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology, 108*(1), 113–125.

[59]Patterson, J. L. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics, 23*(4), 493–508.

Peña, E., Gutierrez-Clellen, V., Iglesias, A., & Goldstein, B. (in preparation). Bilingual English Spanish Assessment.

[60]Peña, E. D., Bedore, L. M., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language, and Hearing Research, 45*(5), 938.

[61]Perry, N. J., Kay, S. M., & Brown, A. (2008). Continuity and change in home literacy practices of Hispanic families with preschool children. *Early Child Development and Care, 178*(1), 99–113.

[62]Reyes, I. (2006). Exploring connections between emergent biliteracy and bilingualism. *Journal of Early Childhood Literacy, 6*(3), 267–292.

[63]Reyes, I., & Azuara, P. (2008). Emergent biliteracy in young Mexican immigrant children. *Reading Research Quarterly, 43*(4), 374–398.

[64]Rinaldi, C., & Páez, M. (2008). Preschool matters: Predicting reading difficulties for Spanish-speaking bilingual students in first grade. *Learning Disabilities (Weston, Mass.), 6*(1), 71.

[65]Schnitzer, M. L., & Krasinski, E. (2003). Idiosyncratic variation in the application of phonological processes in early bilingual acquisition. *Language Sciences, 25*(1), 1–14.

[66]Sheng, L., McGregor, K. K., & Marian, V. (2006). Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research, 49*(3), 572.

[67]Shin, S. J. (2002). Birth order and the language experience of bilingual children. *TESOL Quarterly, 36*(1), 103–113.

Shin, H. B., & Kominski, R. A. (2010). Language use in the United States: 2007. American Community Survey Reports. (ACS-12). Washington, DC: U.S. Census Bureau.

[68]Silva-Corvalán, C., & Montanari, S. (2008). The acquisition of ser, estar (and be) by a Spanish–English bilingual child: The early stages. *Bilingualism: Language and Cognition, 11*(03), 341–360.

[69]Tabors, P. O., Páez, M., & López, L. M. (2003). Dual language abilities of bilingual four-year-olds: Initial findings from the early childhood study of language and literacy development of Spanish-speaking children. *NABE Journal of Research and Practice, 1*, n1.

[70]Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Child Development, 80*(5), 1545–1563.

[71]Yaden, D. B., & Tardibuono, J. M. (2004). The emergent writing development of urban Latino preschoolers: Developmental perspectives and instructional environments for second-language learners. *Reading & Writing Quarterly, 20*(1), 29–61.

[72]Yavas, M. S., & Core, C. W. (2001). Phonemic awareness of coda consonants and sonority in bilingual children. *Clinical Linguistics & Phonetics, 15*(1–2), 35–39.

Government Reports

[100] Administration for Children and Families, Office of Planning, Research, and Evaluation (ACF-OPRE). (2006). Head Start performance measures center, family and child experiences survey (FACES 2000) technical report. Administration for Children & Families, Department of Health and Human Services (DHHS).

[101] Administration for Children and Families, Office of Planning. (2001). Head Start FACES: Longitudinal findings on program performance, third progress report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[102] Aikens, N., Hulsey, L. K., Moiduddin, E., Kopack, A., Takyi-Laryea, A., Tarullo, L., & West, J. (2011). Data Tables for FACES 2009 Head Start children, families, and programs: Present and past data from FACES report. ACF-OPRE report 2011-33b. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[103] Andreassen, C., & Fletcher, P. (2005). Early childhood longitudinal study, birth cohort (ECLS-B), Psychometric Characteristics. Volume 1 of the ECLS-B methodology report for the nine-month data collection (NCES 2005-100). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

[104] Flanagan, K., & West, J. (2004). Children born in 2001: First results from the base year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS–B) (NCES 2005–036). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

[105] Hulsey, L., Aikens, N., Xue, Y., Tarullo, L., West, J., Administration for Children and Families (DHHS), Office of Planning, et al. (2010). Data tables for FACES 2006: A year in Head Start report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[106] Jacobson Chernoff, J., Flanagan, K. D., McPhee, C., & Park, J. (2007). Preschool: First findings from the preschool follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (NCES 2008025). Washington, DC: National Center for Education Statistics, Institute of Education Sciences.

[107] Malone, L., Hulsey, L., Aikens, N., West, J., Tarullo, L., ACF, OPRE, et al. (2010). Data tables for FACES 2006: Head Start children go to kindergarten report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[108] Moiduddin, E., N. Aikens, L. Tarullo, & J. West. (2010). Data tables for FACES 2006: A second year in Head Start report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS .

[109] Mulligan, G. M., & Flanagan, K. D. (2006) Age 2: Findings from the 2-year-old follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (NCES 2006-043). U.S. Department of Education. Washington, DC: National Center for Educational Statistics.

[110] Najarian, M., Snow, K., Lennon, J., & Kinsey, S. (2010). Early childhood longitudinal study, birth cohort (ECLS-B), preschool–kindergarten 2007 psychometric report (NCES 2010-009). U.S. Department of Education. Washington, DC: National Center for Education Statistics, Institute of Education Sciences.

[111] Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., et al. (2010). Head Start impact study: Technical report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[112] Vogel, C. A., Boller, K., Xue, Y., Blair, R., Aikens, N., Burwick, A., et al. (2011). Learning as we go: A first snapshot of Early Head Start programs, staff, families, and children. OPRE report #2011-7. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[113] West, J., Tarullo, L. Aikens, N., & Hulsey, L. (2008). Study design and data tables for FACES 2006 baseline report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[114] 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: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

[115] Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., Hubbell McKey, R., et al. (2003). Head Start FACES 2000: a whole-child perspective on program performance. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, DHHS.

APPENDIX A

STUDY DESCRIPTIONS AND APPROACHES TO ASSESSMENT

Table A.1. Approach to Assessing Dual Language Learners in Small-Scale Studies

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Anderson (2004)	5 children who were native speakers of Korean (n = 3), Russian (n = 1), or French (n = 1), and who were learning English as their second language Initial age = 3 yrs. 9 mos. to 4 yrs. 9 mos.	To examine L2 phonological acquisition in DLL children with a variety of L1 language backgrounds.	Children's phonological skills in both L1 and L2 were assessed. Two approaches were used to obtain information on children's English proficiency: (1) Parents were asked to estimate their child's perceived level of effective communication in English, and (2) children were individually administered the PreLAS.	Phonological skill development was assessed in L1 and L2 using a word-list measure developed by the researchers. In addition, a spontaneous speech sample was obtained while the child interacted with the experimenter (not analyzed for current study). Data were collected every one to two months for a maximum of five sessions.
Anthony et al. (2009)	130 Spanish-English ELLs Mean initial age = 4 yrs. 5 mos. (SD = 7 mos.)	To examine the contributions of vocabulary and letter knowledge on the development of phonological awareness in Spanish-speaking children learning English.	Spanish and English versions of each measure were administered to all children. Only responses provided in English were accepted as correct during administration of English tests; only responses provided in Spanish were accepted as correct during administration of Spanish tests.	Children's expressive and receptive vocabulary was assessed using the EOWPVT/EOWPVT-SBE (English and Spanish) and the ROWPVT/ROWPVT-SBE (English and Spanish). Children's phonological awareness was assessed using the PCTOPPP/SPCTOPPP (English and Spanish; Elision and Blending subtests). Children's letter knowledge was assessed using letter name and letter sound identification tasks developed by the researchers (English and Spanish). Data were collected in December and April of the 2003-2004 school year.
Bedore, Fiestas, Peña, & Nagy (2006)	22 Spanish-English bilinguals, 22 Spanish monolinguals, and 22 English monolinguals 4 yrs. 3 mos. to 7 yrs. 3 mos. (mean = 5 yrs. 9 mos.)	To examine bilingual children's verbal fluency in Spanish and English (as measured by maze use) and to compare the fluency of bilingual children with that of their monolingual peers.	Used parent and teacher report, to assign children to English, Spanish, or bilingual groups. Bilingual children had approximately equal input and output. Children who were included in the monolingual groups had over 80% output in their dominant language and less than 20% input in their nondominant language.	Children's language fluency was assessed using a spontaneous narrative for the monolingual children (Spanish or English; SALT, C-units) and two spontaneous narratives elicited from the DLLs (Spanish and English; SALT, C-units).

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Bedore, Peña, García, & Cortez (2005)	Study 1: n = 55 (13 Spanish dominant-English DLLs, 7 Spanish-English dominant DLLs, 24 primarily Spanish-speaking, 11 primarily English-speaking) 4 yrs. to 7 yrs. 11 mos. Study 2: n = 40, matched with study 1 participants on age and language background 5 yrs. to 6 yrs. 1 mo.	To examine the cross-language overlap of semantic abilities of bilingual children and whether classification analysis based on monolingual or conceptual scoring can accurately classify bilinguals' semantic development.	Children were divided into four language groups based on parent-teacher report or grammaticality in the target language on a narrative storytelling task. Children in the primarily English-speaking group were assessed in English; primarily Spanish-speaking children were assessed in Spanish; and bilingual English and bilingual Spanish children were assessed in both languages. Responses were recorded in the language(s) in which they were produced, and scores were calculated for the correct responses to each item (monolingual score in English or Spanish, total response score, and conceptual score).	Study 1: Children's semantic development was assessed using the BESA (Spanish and/or English; Characteristic Properties subtest). Study 2: Children were administered the BESA (Spanish and/or English; Phase 2 Semantic subtest).
Bialystok, Luk, Peets, & Yang (2010)	1,738 English monolinguals or English/other bilinguals. Aggregate sample of children who participated in studies conducted by the study author over a five-year period. 3 yrs. to 10 yrs.	To examine differences in English receptive vocabulary between monolingual and bilingual children.	All children were assessed in English. Based on parent report, all bilingual children were educated in English at school, were fluent in both languages, and spoke a non-English language at home with family members. Children who were learning English as a second language were excluded from analyses. Information about the non-English language spoken was also obtained from parent report.	Children's receptive vocabulary was assessed using the PPVT-III (English)
Bland-Stewart & Fitzgerald (2001)	15 Spanish-English bilingual children 2.6 to 5.0 yrs. (mean age = 3.8 yrs.)	To examine morphological development in Spanish-English DLLs and Standard American English (SAE) monolingual children, and identify potential biases in the use of SAE normative data on bilingual children.	All children participated in a language sampling session with a bilingual graduate clinician or certified SLP. Information regarding each child's use of and proficiency in English and Spanish was gathered from parents and teachers via case histories/questionnaires in which they were asked to report how well the child understood and spoke Spanish and English in the home, community, and school. Each child was required to pass a criterion-referenced speech/language screener in English.	Morphological development was assessed using spontaneous language sampling derived from 30-minute play sessions conducted in English. Language sample of 100 utterances was obtained for each child. Derived MLU scores (total morphemes divided by total utterances) and count of use of 14 grammatical morphemes. Analysis conducted by child MLU (n = 6 MLU of 3.0-3.4; n = 3 MLU of 3.5-3.9; n = 6 MLU of 4.0-4.4).

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Bohman, Bedore, Peña, Mendez-Perez, & Gillam (2010)	757 Spanish-English bilingual and monolingual children. Central Texas (District A): n = 300 (37% DLLs, 27% Spanish monolingual, 36% English monolingual); Central Texas (District B): n = 276 (42% DLLs, 31% Spanish monolingual, 27% English monolingual); Northern Utah: n = 181 (73% DLLs, 15% Spanish monolingual, 12% English monolingual) Pre-kindergarten and kindergarten-aged children	To examine the relationship between language experience and language semantics/grammar in Spanish-English DLLs and monolinguals.	Based on parent report, percentage of language output was used to classify children into functionally monolingual and bilingual groups. Children were assessed in English and Spanish.	Children's semantics was assessed using the Semantics subtest of the BESOS (Spanish and English). Children's morphosyntax was assessed using the Morphosyntax subtest of the BESOS (Spanish and English).
Brice, Carson, & O'Brien (2009)	16 Spanish-English bilinguals who were enrolled in Head Start or a preschool program 4 yrs. to 5 yrs.	To examine differences in articulation and phonological patterns present in both languages of Spanish-English DLLs.	Children's articulation and phonology was assessed in English and Spanish.	Children's articulation and phonology in Spanish were assessed using the researcher-developed CASA-P. English articulation was assessed using the GFTA-2; phonology was assessed using the KLPA-2.
Buckwalter & Gloria Lo (2002)	Case study of one Chinese child who was learning English. Family moved from Taiwan to the U.S. 5 months prior to the start of the study. 5 yrs. of age	To examine Chinese-English emergent literacy awareness and the effects of emergent literacy development in one language on the other.	Measures were obtained in Chinese and English.	Emergent literacy was assessed using reading, writing, and matching games that corresponded with books read with caregivers. Data were also collected while reading and writing familiar books/stories, mail, and looking at words with caregivers. Measures were obtained in both languages for 1.5 to 2 hours per week for 15 weeks.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Bunta, Fabiano-Smith, Goldstein, & Ingram (2009)	24 children: 8 monolingual Spanish, 8 monolingual English, and 8 Spanish-English bilinguals Monolingual Spanish = 3 yrs. 2 mos. to 4 yrs. (mean = 3 yrs. 4 mos.); monolingual English = 3 yrs. to 3 yrs. 11 mos. (mean = 3 yrs. 3 mos.); bilinguals = 3 to 4 years (mean = 3 yrs. 6 mos.)	To compare phonological complexity between 3-year-old bilingual and monolingual children as measured by percent consonants correct (PCC), phonological mean length of utterance (pMLU), and phonological target approximations.	Bilingual children were assessed in English and Spanish across separate sessions, and monolingual children were assessed in their respective language. Parent and teacher questionnaires were used to obtain information about children's length of exposure to each language and proficiency ratings in each language for bilingual speakers.	The phonology subtest of the BESA was used to elicit sounds in single words (Spanish and/or English depending on language group). A speech sample was obtained from each bilingual child in English and Spanish, and from each monolingual child in English or Spanish. All single-word and speech samples were transcribed phonetically and analyzed for PCC, pMLU, and proximity using LIPP.
Bunta & Ingram (2007)	Included 30 child and 18 adult participants, with equal representation of Spanish-English bilingual, Spanish monolingual, and English monolingual groups in each age group. Children were categorized into younger (3 yrs. 9 mos. to 4 yrs. 5 mos.) and older age groups (4 yrs. 6 mos. to 5 yrs. 2 mos.). Age of adult participants was not specified.	To examine speech rhythm acquisition in Spanish-English bilingual and monolingual children.	Bilingual children provided Spanish and English samples. Language status was based on parent-reported percentage of language use and the ability to produce and maintain a conversation in the ambient language(s). Children who used one language more than 80% of the time were considered monolingual; children who used both languages at least 20% of the time were considered bilingual. The Expressive Communication subscale of the PLS-4 was used to establish participants' age-appropriate language skills in the target languages (both in Spanish and English for bilingual children).	Children's speech rhythm acquisition was assessed using a sentence elicitation task using the normalized vocalic and intervocalic Pairwise Variability Indices (PVIs) developed by the researchers (Spanish and/or English)
Castilla, Restrepo, & Perez-Leroux (2009)	49 Spanish-speaking children attending English-only pre-kindergarten classrooms Initial age = 4 yrs. 4 mos. to 4 yrs. 11 mos. (mean = 4 yrs. 7 mos.)	To examine language influence of Spanish on the acquisition of English.	All children were assessed in Spanish at the beginning of the school year and in English at the end of school year. During the English administration of the BESA Semantics subtest, Spanish responses were counted as correct. During the Spanish administration of the BESA Semantics subtest, English responses were counted as correct.	Children's semantics and grammar were assessed using the Semantics and Morphosyntax subtests of the BESA (Spanish and English). A language sample was also obtained using a story-retelling task (Spanish; MLUw, C-units, SALT, CLAN). Data were collected twice: during the first two months of the school year (Spanish) and eight to nine months later during the last month of the school year (English).

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Castro & Gavruseva (2003)	1 Spanish-English bilingual child Initial age = 1 yr. 10 mos.; age at end = 2 yrs. 6 mos.	To investigate similarities and differences in English and Spanish verb morphology (use of finite and non-finite root predicates) in a Spanish-English bilingual child.	Spontaneous play sessions were conducted in English and Spanish.	Child's verb morphology was assessed using spontaneous language sampling. Data were collected over five English-language and six Spanish-language sessions. Language samples were coded to examine use of verb predicates in each language.
Chung (2006)	2 Korean-English bilingual children (siblings) 4.5 years of age and 11 years of age.	To examine the use of code-switching in Korean-English bilingual children.	Spontaneous language sampling was conducted in Korean and English. Information on language proficiency was reported by the study author (children's mother) via in-depth observations and knowledge of the participants.	Children's code-switching was assessed using spontaneous language sampling (Korean and English). Data were collected in one, three-hour videotaped session.
Conboy & Mills (2006)	30 Spanish-English bilingual children (n = 16 English-dominant, n = 14 Spanish-dominant) 1 yr. 7 mos. to 1 yr. 10 mos. (mean = 20.3 mos.)	To determine whether separate processing systems for each language of Spanish-English DLL toddlers are evident in the brain activity elicited by words in each language (event-related potentials, or ERPs).	Parents reported on children's vocabulary knowledge in Spanish and English. Children's dominant language was determined using TCV scores derived from the CDI and parent proficiency ratings in each language.	Children's vocabulary size was assessed using the CDI and its Spanish counterpart, the Inventario II (English and Spanish; Word and Sentences). Children were assigned to higher and lower vocabulary groups using a median split of the derived TCV scores. Children's ERPs to words was assessed using electroencephalograms (EEG).
Conboy & Thal (2006)	64 Spanish-English bilingual children 1 yr. 7 mos. to 2 yrs. 7 mos.	To examine the relationship between the grammatical abilities of Spanish-English DLLs and vocabulary development in each language.	Caregivers reported on children's language skills using the MacArthur CDI and its Spanish equivalent, the Inventario II.	Children's semantics and morphosyntax were assessed using the CDI and its Spanish counterpart, the Inventario II (English and Spanish; Word and Sentences). Cross-sectional data (n = 30) were collected at 19, 20, 21, or 22 months of age. Longitudinal data (n = 34) were collected for a maximum of three timepoints initially between the ages of 19 and 22 months and again between the ages of 24 and 26 months and/or 28 and 31 months.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Davidson, Raschke, & Pervez (2010)	Study 1: 20 Urdu-English bilingual and English monolingual children. Urdu-English DLLs: 5 yrs. 6 mos. to 6 yrs. 3 mos. (mean = 6 yrs.). English monolinguals: 5 yrs. 7 mos. to 6 yrs. 2 mos. (mean = 6 yrs.) Study 2: 72 Urdu-English bilinguals and English monolinguals. In each language group, 18 children were between the ages of 3 and 4 years and 18 children were between the ages of 5 and 6 years. Younger age group: Urdu-English DLLs = 3 yrs. 8 mos. to 4 yrs. 7 mos. (mean = 4 yrs. 2 mos.); English monolinguals = 3 yrs. 7 mos. to 4 yrs. 9 mos. (mean = 4 yrs. 4 mos.). Older age group: Urdu-English DLLs = 5 yrs. 8 mos. to 6 yrs. 1 mo. (mean = 6 yrs.); English monolinguals = 5 yrs. 6 mos. to 6 yrs. 3 mos. (mean = 6 yrs.)	To compare syntactic awareness of bilingual Urdu-English speakers and monolingual English-speaking children.	Study 1: All children were assessed in English (PPVT-III and syntax task); bilingual children were also assessed using an Urdu translation of the PPVT-III. Study 2: All children were assessed in English using the PPVT-III; bilingual children were also assessed using an Urdu translation of the PPVT-III. Monolingual children were administered the syntax task in English, and bilingual children received either the English or Urdu version of the instrument (method of determination not specified).	In both studies, children's receptive vocabulary was assessed using the PPVT-III (English). Bilingual children were also administered an Urdu translation of the PPVT-III. In both studies, children's syntax awareness was assessed using a syntax task developed by the researchers (English and English and/or Urdu).

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Dickinson, McCabe, Clark-Chiarelli, & Wolf (2004)	123 Spanish-English bilingual children Initial age = 2 yrs. 10 mos. to 5 yrs. (mean = 4 yrs. 1 mo.)	To examine the development of Spanish and English phonological awareness in Spanish-English DLLs.	Language dominance was assessed using three approaches: teachers were asked to indicate the child's primary language; parents were asked to choose the language in which their child should be assessed; and assessor ratings based on informal conversation with the child, which was confirmed by administration of a test of receptive language in that language. All children were assessed using the PPVT-III/TVIP and were administered the EPAP in English and Spanish. Based on language dominance, children were either administered the English or Spanish version of the ELP (responses accepted in either language).	Children's vocabulary was assessed using the PPVT-III/TVIP (English and Spanish). Children's emergent literacy was assessed using the ELP (Spanish or English, based on language dominance). Children's phonological awareness was assessed using the EPAP (Spanish and English). Data were collected in both the fall (October/November) and spring (April/May).
Fabiano & Goldstein (2005)	3 bilingual Spanish-English speaking children 5 yrs. to 7 yrs.	To examine the frequency and types of phonological cross-linguistic effects that occur over time in bilingual Spanish-English speaking children.	Language samples were obtained from each child in Spanish and English during a single session. Language history was obtained from parent or clinician report. The parent interview included two proficiency scales: language ability and language use. Children deemed bilingual based on a measure of grammaticality using conversation and narrative samples.	Children's phonology was assessed using the phonology subtest of the BESA (single-word sampling, LIPP); spontaneous language sampling (SALT/LIPP); and narrative language sampling (SALT/LIPP). Measures were obtained in Spanish and English.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Fabiano-Smith & Barlow (2010)	24 children categorized into three groups based on language history: 8 bilingual Spanish-English speakers, 8 monolingual Spanish speakers, and 8 monolingual English speakers Bilingual = 3 to 4 yrs. (mean = 3 yrs. 6 mos.); monolingual Spanish = 3 yrs. 2 mos. to 4 yrs. (mean = 3 yrs. 4 mos.); monolingual English = 3 yrs. to 3 yrs. 11 mos. (mean = 3 yrs. 3 mos.).	To explore the effect of cross-linguistic interaction in the phonological acquisition in Spanish-English DLLs.	Each bilingual child was assessed in Spanish and English, and each monolingual child was assessed in their respective language. Parent and/or teacher reports were used to determine each child's language status (i.e., monolingual or bilingual), bilingual status (i.e., simultaneous or sequential), and phonological status (i.e., typically developing). Bilingual children had at least eight months of exposure to English, received at least 20% input in both languages, and produced at least 20% output in both languages (based on parent description of child's schedule on a typical day, and language typically used during interactions). Parents also rated their child's proficiency in English and Spanish along 0-4 scale, with all children rating as native or near-native in their proficiency in both languages. Simultaneous/sequential bilinguals aggregated in analyses given similar levels of input, output, and proficiency.	Single word samples (Spanish and/or English) were collected from each child using the phonology subtest of the BESA; phonetic inventories were derived from these data.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Fabiano-Smith & Goldstein (2010a)	24 children categorized into three groups based on language history: 8 bilingual Spanish-English speakers, 8 monolingual Spanish speakers, and 8 monolingual English speakers Bilingual = 3 to 4 yrs. (mean = 3 yrs. 6 mos.); monolingual Spanish = 3 yrs. 2 mos. to 4 yrs. (mean = 3 yrs. 4 mos.); monolingual English = 3 yrs. to 3 yrs. 11 mos. (mean = 3 yrs. 3 mos.).	To explore and compare the early-, middle-, and late-developing speech sounds (EML) of Spanish-English DLLs with that of monolinguals.	Each bilingual child was assessed in Spanish and English, and each monolingual child was assessed in their respective language. Parent and/or teacher reports were used to determine each child's language status (i.e., monolingual or bilingual), bilingual status (i.e., simultaneous or sequential), and phonological status (i.e., typically developing). Bilingual children had at least 8 months of exposure to English, received at least 20% input in both languages, and produced at least 20% output in both languages (based on parent description of child's schedule on a typical day, and language typically used during interactions). Parents also rated their child's proficiency in English and Spanish along 0-4 scale, with all children rating as native or near-native in their proficiency in both languages. Simultaneous/sequential bilinguals aggregated in analyses given similar levels of input, output and proficiency.	The phonology subtest of the BESA was used to elicit sounds in single words (Spanish and/or English), and samples were analyzed using LIPP. PCC was calculated for all language groups (monolingual English, monolingual Spanish, DLL English productions, and DLL Spanish productions).

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Fabiano-Smith & Goldstein (2010b)	24 children categorized into three groups based on language history: 8 bilingual Spanish-English speakers, 8 monolingual Spanish speakers, and 8 monolingual English speakers Bilingual = 3 to 4 yrs. (mean = 3 yrs. 6 mos.); monolingual Spanish = 3 yrs. 2 mos. to 4 yrs. (mean = 3 yrs. 4 mos.); monolingual English = 3 yrs. to 3 yrs. 11 mos. (mean = 3 yrs. 3 mos.).	To determine how between-language interaction contributes to phonological acquisition in bilingual Spanish-English speaking children.	Each bilingual child was assessed in Spanish and English, and each monolingual child was assessed in their respective language. Parent and/or teacher reports were used to determine each child's language status (i.e., monolingual or bilingual), bilingual status (i.e., simultaneous or sequential), and phonological status (i.e., typically developing). Bilingual children had at least 8 months of exposure to English, received at least 20% input in both languages, and produced at least 20% output in both languages (based on parent description of child's schedule on a typical day, and language typically used during interactions). Parents also rated their child's proficiency in English and Spanish along 0-4 scale, with all children rating as native or near-native in their proficiency in both languages. All bilingual children were categorized as early bilinguals, having been exposed primarily to Spanish input and output in the home up to age 3 years, or to both languages in the home from birth.	The phonology subtest of the BESA was used to elicit sounds in single words (Spanish and/or English), and samples were analyzed using LIPP (phonetic analysis, PCC). A spontaneous speech sample was also obtained (Spanish and/or English) and analyzed using SALT (MLUm, NDW, TNU). Data were collected in May and June.
Farver, Xu, Eppe, & Lonigan (2006)	122 preschoolers: 43 Spanish-English bilinguals, 57 Spanish monolinguals, and 22 English monolinguals 3 yrs. 3 mos. to 4 yrs. 1 mo. (mean = 3 yrs. 9 mos.)	To examine the relationships among literacy activities, the home environment, and DLLs' school readiness as measured by social functioning and oral language.	Children were tested in the language in which they were the most fluent, based on parent report, teachers' experiences with the child, and a short conversation the assessor carried out with the children prior to the assessment (PPVT-R = 22 children and TVIP = 100 children).	Children's receptive vocabulary was assessed using the PPVT-R or TVIP (English or Spanish, based on language dominance). Home literacy practices were assessed using the Home Literacy Environment Questionnaire and the TRT (English or Spanish). Parental stress was assessed using the PSI. Children's social functioning was assessed using the teacher-reported BASC.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Fennell, Byers-Heinlein, & Werker (2007)	Study 1: 48 bilingual infants in each of three age groups (14, 17 and 20 months old). 14 mos. subsample = 1 yr. 1 mo. to 1 yr. 3 mos.; 17 mos. subsample = 1 yr. 4 mos. to 1 yr. 6 mos.; 20 mos. subsample = 1 yr. 7 mos. to 1 yr. 8 mos. Study 2: 25 Chinese-English and 28 French-English bilinguals. 17 mos. subsample = 1 yr. 4 mos. to 1 yr. 6 mos.; 20 mos. subsample = 1 yr. 7 mos. to 1 yr. 9 mos.	To explore the relationship between speech perception and lexical acquisition in DLLs.	n.a.	In both studies, children's early phonological discrimination was assessed using the Switch Word-Object Associative Task.
Gildersleeve-Neumann, Kester, Davis, & Peña (2008)	33 children divided into monolingual English (n = 10), predominantly English (n = 20), and balanced bilingual Spanish-English (n = 3) groups. 3 yrs. 1 mo. to 3 yrs. 10 mos.	To compare consonant, vowel, and word and syllable shape inventories, accuracy rates, and phonological error patterns for English monolinguals and Spanish-English DLLs.	The single-word picture naming task was administered in English to all children.	Children's phonology was assessed using a single-word picture naming task developed by the researchers (English; LIPP, PCC, PVC). Data were collected both in the fall and spring.
Gildersleeve-Neumann, Peña, Davis, & Kester (2009)	6 Spanish-English bilinguals Initial age = 3 yrs. 2 mos. to 3 yrs. 10 mos. (mean = 3 yrs. 5 mos.)	To explore Spanish phonological development in Spanish-English bilingual children when first exposed to English and eight months after first exposure.	Single-word speech samples were collected in Spanish. Language development was established through researcher observation, parent report, and standardized assessments, including performance on the Spanish versions of the Expressive and Receptive One-Word Picture Vocabulary Tests, the Comprehension Subtest of the Stanford Binet Intelligence Scale (4th edition), and dynamic assessment procedures (Kester, Peña, & Gillam, 2001).	Children's phonology was assessed in Spanish using single-word speech samplings (LIPP, independent analyses, relational analyses); words were elicited with a picture identification task. Data were collected twice, at the point of contact with English and eight months later.
Gildersleeve-Neumann & Wright (2010)	42 typically developing children, including 14 Russian-English bilingual children and 28 English monolingual children. 3 yrs. 3 mos. to 5 yrs. 7 mos.	To examine the effects of Russian phonetic and phonological properties on English single-word productions in Russian-English DLL children.	Information on children's language exposure was obtained via parent and teacher report. Single-word speech samples were collected in English from all children.	Children's phonology was assessed using the Phonological and Articulatory Bilingual Assessment, a measure developed by the researchers (English; LIPP, IPC, PVC, PCC)

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Goldberg, Paradis, & Crago (2008)	19 bilingual children whose first language was Korean, Mandarin, Cantonese, Spanish, Romanian, Arabic, Japanese or Farsi Initial age = 4 yrs. 2 mos. to 6 yrs. 9 mos. (mean = 5 yrs. 4 mos.)	To examine English lexical acquisition of children learning English as a second language.	All assessments were administered in English.	Children's receptive vocabulary in English was assessed using the PPVT-III. Expressive vocabulary was assessed using spontaneous language sampling (English; CHAT, CLAN, NDW). Nonverbal IQ was assessed using the Columbia Mental Maturity Scale (English). Data were collected over five sessions every six months for two years. Maternal level of education, use of English in the home, and age of onset of English learning were assessed via a parental questionnaire.
Goldstein, Fabiano, & Washington (2005)	15 Spanish-English bilingual children further classified as bilingual Spanish-English speakers (n = 5), predominantly Spanish speakers (n = 5), or predominantly English speakers (n = 5) 5 yrs. to 5 yrs. 5 mos. (mean = 5 yrs. 2 mos.)	To investigate the relationship between amount of output in DLLs' two languages and examine the phonological skills of Spanish-English DLL and monolingual children.	Bilingual Spanish-English speakers were assessed in Spanish and English; predominant Spanish speakers in Spanish; and predominant English speakers in English. The children's language status and language profile in each language (e.g., years of exposure, proficiency, input/output) were determined by parent report. All children in the bilingual group were reported to use each language more than 20% of the time; children assigned to the monolingual Spanish or monolingual English group were reported to use the predominant language more than 20% of the time and the nonpredominant language less than 20% of the time.	Children's language output was assessed using a single-word phonological assessment developed by the researchers (Spanish and/or English; PCC, syllable type, phonological patterns).
Goldstein & Washington (2001)	12 Spanish-English bilingual children. Girls were overrepresented (n = 10) 4 yrs. to 4 yrs. 11 mos. (mean = 4 yrs. 7 mos.)	To examine and compare the phonological development of Spanish-English DLL and monolingual children.	All children were assessed in English and Spanish. Based on informal parent and teacher report, children were reported to be developing both languages simultaneously.	Children's phonological development was assessed using the Phonological Measure of Bilingual Latino/a Children, a single-word assessment developed by the first author (Spanish and English; LIPP, independent and relational analyses)

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Gonzalez & Uhing (2008)	48 preschool children from Hispanic Spanish-speaking migrant and immigrant families 3.4 yrs. to 4.8 yrs. (mean = 4.3 yrs.)	To examine aspects of the home literacy environment in relation to Spanish and English oral language development in high-risk Hispanic children.	All children were administered the PreLAS 2000 in Spanish and English.	Parent-reported family/home literacy was assessed using the Familia Inventory (Spanish or English). Children's language proficiency was assessed using the PreLAS 2000 (Spanish and English).
Guiberson, Barrett, Jancosek, & Itano (2006)	10 Spanish-English bilingual children of Mexican immigrant parents: n = 8 language maintenance, n = 2 language loss Initial age = 2 yrs. 9 mos. to 3 yrs. 1 mo.; age at end = 4 yrs. 11 mos. to 5 yrs. 1 mo.	To examine Spanish language development trajectories with interest in capturing language maintenance and loss in Spanish-English DLLs.	All assessments were administered in Spanish. Based on the parent-child video recorded interactions conducted at each timepoint, Spanish lexical diversity (Spanish D) was calculated and used in trajectories that categorized children into one of two profiles: language maintenance and language loss.	Children's language usage was assessed using a modified version of the parent-report Bilingual Language Proficiency Questionnaire (BFQ). Children's language behaviors were assessed using spontaneous language sampling (Spanish; video recorded parent-child interactions, VBS, Spanish D). Children's expressive vocabulary was assessed using SLAP (Spanish). Children were assessed annually at three points in time.
Hammer, Davison, Lawrence, & Miccio (2009)	72 Spanish-English bilinguals Mean age = 4 yrs. 1 mo.	To examine the effects of maternal home language usage on Spanish-English DLLs' vocabulary and emergent literacy.	Children were assessed using both the PPVT and TVIP. The TERA-2 was administered only in English.	Mother's language usage was assessed using a questionnaire developed by the researchers (Spanish or English). Children's receptive vocabulary was assessed using the PPVT-III/TVIP (English and Spanish). Children's early literacy abilities were assessed using the TERA-2 (English). Children were assessed annually over a three-year period.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Hammer, Lawrence, & Miccio (2007)	88 Spanish-English bilinguals Initial mean age = 3 yrs. 9 mos.	To examine the relationship between DLLs' receptive language growth and their later reading outcomes.	Mothers reported on the ages at which children were spoken to and expected to communicate in both Spanish and English. Children spoken to in two languages from birth were classified as HEC; children spoken to in Spanish from birth and not expected to communicate in English on a regular basis until age 3 (Head Start entry) were classified as SEC. All children were assessed in both languages.	English and Spanish receptive language was assessed using the PPVT-III and TVIP, respectively. Oral comprehension was assessed in English using the TELD-3 and in Spanish using the PLS-3. Measures were administered in the fall and spring of two consecutive Head Start years. Children's early literacy abilities were assessed using the WLPB-R (Spanish or English; Letter-Word Identification subtest) and the TERA-2 (English). Measures were administered during the spring of the kindergarten year.
Hammer, Lawrence, & Miccio (2008a)	83 Spanish-English bilingual children Initial mean age = 3 yrs. 9 mos.	To examine the effects of time of exposure to English on the development of Spanish and English receptive language in DLLs.	Mothers reported on the ages at which children were spoken to and expected to communicate in both Spanish and English. Children spoken to in two languages from birth were classified as HEC; children spoken to in Spanish from birth and not expected to communicate in English on a regular basis until age 3 (Head Start entry) were classified as SEC. All children were assessed in both languages.	Children's receptive vocabulary was assessed using the PPVT-III/TVIP (English and Spanish). Children's language comprehension was assessed using the PLS-3 (Spanish; Auditory Comprehension) and the TELD-3 (English; Receptive Language). Children were assessed during the fall and the spring of two consecutive years of Head Start.
Hammer, Lawrence, & Miccio (2008b)	83 Spanish-English bilingual children Initial mean age = 3 yrs. 9 mos.	To examine the effects of summer vacation on language development in Spanish-English DLLs.	Mothers reported on the ages at which children were spoken to and expected to communicate in both Spanish and English. Children spoken to in two languages from birth were classified as HEC; children spoken to in Spanish from birth and not expected to communicate in English on a regular basis until age 3 (Head Start entry) were classified as SEC. All children were assessed in both languages.	Children's language comprehension was assessed using the PLS-3 (Spanish) and the TELD-3 (English). Children were assessed during the fall and the spring of two consecutive years of Head Start.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Hammer, Miccio, & Wagstaff (2003)	43 Spanish-English bilinguals Mean age = 3 yrs. 8 mos.	To investigate the relationship between home literacy experiences and bilingual preschoolers' early literacy outcomes.	Based on parent report, children who were spoken to in Spanish and English from birth were classified as simultaneous language learners. Children who were spoken to in Spanish from birth and exposed to English at Head Start entry were considered sequential learners. All children were assessed in English.	Children's home literacy experiences were assessed using a questionnaire developed by the researchers (Spanish or English). Children's early literacy knowledge was assessed using the TERA-2 (English). Children were assessed at two timepoints.
Hammer, Rodriguez, Lawrence, & Miccio (2007)	Parents of 81 Spanish-English bilingual children: n = 51 simultaneous bilingual (SI), n = 30 sequential bilingual (SE). Mean age = 4 yrs. 8 mos.	To examine relationships between Puerto Rican mothers' parenting beliefs and home literacy practices.	Only parent questionnaires were used in this study (Spanish or English). Mothers reported on the ages at which children were spoken to and expected to communicate in both Spanish and English. Children spoken to in two languages from birth were classified as HEC; children spoken to in Spanish from birth and not expected to communicate in English on a regular basis until age 3 (Head Start entry) were classified as SEC.	Participant background information was obtained using a questionnaire developed by the researchers. Parent beliefs about education and literacy were assessed using the Parental Modernity Scale and the Rank Order of Parental Values. Home literacy practices were assessed using the Home Literacy Activities Questionnaire developed by the researchers.
Jia (2003)	10 native Mandarin Chinese-speaking children and adolescents who recently immigrated to the U.S. Children ranged in age from 5 yrs. to 16 yrs.	To examine the acquisition of the English plural morpheme in Mandarin L1 - English DLLs.	All assessments were administered in English.	Children's language environment was assessed using an annual parental questionnaire developed by the researchers (Mandarin Chinese), child and parent self-report, and interviewer observations. Children's productions of the English plural morpheme were assessed using a picture naming task (English) and spontaneous speech sampling (English). Children were assessed over seven monthly sessions during Year 1, four quarterly sessions during Year 2, two semi-annual sessions each during Years 3 and 4, and a single session during Year 5.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Johnson & Wilson (2002)	2 Japanese-English bilingual siblings 2 yrs. 10 mos. and 4 yrs. 8 mos.	To examine issues affecting data collection and interpretation of phonetic-phonological and to present voice onset time (VOT) data as a measure of early language differentiation from Japanese-English DLLs.	Recordings of structured parent-child conversations were conducted in Japanese and English.	Voice Onset Time was assessed using recordings of structured parent-child conversations (Japanese and English; VOT). Children were assessed three times, with one week between the first two sessions and the third session two and a half months later.
Kan & Kohnert (2005)	19 Hmong-English bilinguals. Participants were divided into two groups: younger (n=10) and older participants (n=9). Overall sample: 3 yrs. 4 mos. to 5 yrs. 2 mos. (mean = 4 yrs. 4 mos.); younger participants: mean = 3 yrs. 11 mos.; older participants: mean = 5 yrs.	To examine the lexical-semantic development of Hmong L1 - English DLLs.	All children completed four testing sessions: two in English and two in Hmong.	Children's expressive and receptive lexical skills were assessed using a picture naming task and picture identification task, respectively, developed by the researchers (Hmong and English). Data were collected over four different sessions (two in English, two in Hmong).
Kan & Kohnert (2008)	26 Hmong-English bilingual children. 3 yrs. to 5 yrs. 3 mos. (mean = 4 yrs. 3 mos.)	To investigate the relationship between fast mapping skills and vocabulary knowledge in Hmong (L1) and English DLLs.	All children were assessed in English and Hmong.	Children's expressive and receptive lexical skills were assessed using a picture naming task and picture identification task, respectively, developed by the researchers (Hmong and English). Children's fast mapping abilities were assessed using a fast mapping task developed by the researchers (Hmong and English).

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Kim (2009)	33 Korean-English bilingual children: n = 11 enrolled in a Korean bilingual program on the East Coast; n = 22 enrolled in a program on the West Coast. Mean age = 5.2 yrs.	To examine phonological (speech sound) awareness of Korean-English bilinguals and possible cross-language influences.	All children were assessed in both Korean and English.	Children's phonemic awareness was assessed using the CTOPP (English; Blending, Matching, Segmenting) and phoneme awareness tasks for blending and segmenting developed by the researchers (Korean; a syllable awareness task, an onset-rime awareness task, a body-coda awareness, a blending and segmenting CV and CVC monosyllabic words task). Children's sight-word reading abilities were assessed using the Ready-to-Read Word Test: List C (English) and the Ready-to-Read Word Test developed by the researchers (Korean). A pseudoword task and real-word reading task were also administered (English).
Kitabayashi et al. (2008)	Parents of 103 bilingual and monolingual English-speaking children: 52 monolingual English-speaking and 51 ESL children who spoke English and a variety of second languages. 6 mos. to 5 yrs. (mean = 2 yrs.)	To compare reading attitudes of English-speaking and DLL families from a variety of language backgrounds.	Participants were categorized as ESL or English-speaking based on the primary language spoken at home. Respondents with English primary home language were assigned to the English-speaking group; all others were assigned to the ESL group. All respondent surveys were conducted in English.	Families' attitudes about reading were assessed using a validated survey instrument that was modified by the researchers (favorite parent-child activity, child activities aimed at promoting school success, the frequency of reading to their child, and the number of children's books in the home).
Kohnert, Kan, & Conboy (2010)	19 Hmong-English bilinguals 2 yrs. 11 mos. to 5 yrs. 2 mos. (mean = 4 yrs. 3 mos.)	To investigate Hmong and English lexical and grammatical development in Hmong-English DLLs.	All tasks were administered to children in English and Hmong, and were clustered by language (English-only session, Hmong-only session).	Children's grammatical development was assessed using a story re-tell task (Hmong and English; SALT, MLUm, NDW). Children's lexical development was assessed using a picture naming task developed by the researchers (Hmong and English), as well as a picture identification task developed by the researchers (Hmong and English).
Kyratzis, Tang, & Bahar Koymen (2009)	4 Mexican Spanish-English bilinguals and 1 English monolingual 3 yrs. to 5 yrs. 6 mos.	To examine the communicative resources (e.g., code-switching) used by a Spanish-English DLL peer group during pretend play.	Children's language during spontaneous play with their peer group was videotaped in Spanish and/or English.	Children's code-switching was assessed using spontaneous language sampling (Spanish and/or English). Children were assessed twice weekly over one year.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Lao (2004)	Parents of 86 Chinese-English bilinguals who were grouped based on Chinese (n = 52) or English (n = 34) language dominance. 3 yrs. to 5 yrs.	To explore Chinese-English DLL parents' views on bilingual education and bilingual practices.	Parents completed the questionnaire in English or in Chinese.	Parents' attitudes toward bilingual education were assessed using a questionnaire developed by the researchers (Chinese or English). Gathered information on participants' language proficiency and language use, parent's expectations for their children and reasons for enrolling their children in Chinese-English bilingual school, and attitudes towards bilingual education.
Levey & Cruz (2003)	17 Mandarin Chinese-English bilingual children 1 yr. 10 mos. to 4 yrs. (mean = 3 yrs. 4 mos.)	To determine whether exposure to a bilingual care environment influences the language (English versus Mandarin Chinese) or the category (noun versus verb) of children's early word productions.	Parents were interviewed either in Mandarin-Chinese or English.	Children's first words were assessed using a parent interview (Chinese/English) and personal baby book records.
Lopez & Greenfield (2004)	100 Hispanic Spanish-speaking children learning English 4 yrs. to 5 yrs. 6 mos. (mean = 4 yrs. 8 mos.)	To examine the relationship between the oral language and phonological awareness skills of DLLs in Spanish and English.	All children were assessed in English and Spanish.	Children's receptive and expressive languages were assessed using the PreLAS 2000 (Spanish and English). Children's phonological awareness was assessed using the Phonological Sensitivity Test developed by the researchers (Spanish and English).
Marchman, Fernald, & Hurtado (2010)	26 Spanish-English bilingual children 2 yrs. 5 mos. to 2 yrs. 10 mos. (mean = 2 yrs. 6 mos.)	To examine the relationship between vocabulary development and processing efficiency in Spanish-English simultaneous DLLs.	Parents completed both the CDI and Inventario II. Children were assessed in English and Spanish.	Children's language background was assessed using the Language background interview (Spanish or English). Children's vocabulary was assessed using the CDI and its Spanish counterpart, the Inventario II (Spanish and English; Words and Sentences, TVS, TVC). Children's language processing efficiency was assessed using a looking-while-listening procedure developed by the researchers (Spanish and English; reaction time)

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Marchman, Martinez-Sussmann, & Dale (2004)	113 Spanish-English bilinguals 1 yr. 5 mos. to 2 yrs. 6 mos. (mean = 2 yrs.)	To examine language-specific and language-general predictors of lexical and grammatical development in Spanish-English simultaneous DLLs.	Spontaneous language sampling was conducted in Spanish and English. Parents and/or teachers completed the CDI and/or the Inventario II based on their ability to read and write in that language, as well as their familiarity with the child's use of that language.	Children's degree of language exposure was assessed using the Bilingual Background Interview (Spanish/English). Children's vocabulary was assessed using the CDI or Inventario II (English or Spanish, M3Lw; n = 49 completed both forms). Children's behavioral language was assessed using spontaneous language sampling (Spanish and English, NDW, MLUw; n = 26)
Marinova-Todd, Zhao, & Bernhardt (2010)	144 Mandarin-English bilingual and monolingual children: 62 Mandarin-English bilinguals, 61 Mandarin monolinguals, and 21 English monolinguals. 5 yrs. to 6 yrs.	To compare performance on measures of phonological awareness between Mandarin-English DLLs and Mandarin and English monolinguals.	Bilingual children were assessed in both Mandarin and English. Monolingual children were assessed in their respective languages.	Children's expressive vocabulary comprehension was assessed using the PPVT-III and a Chinese translation of the PPVT-R (English and/or Mandarin). Children's phonological awareness was assessed using the CTOPP (English; Elision, Blending, Sound matching) and tasks developed by the researchers (Mandarin; syllable deletion, onset-rime combination, initial sound identification, rhyme detection, tone discrimination).
Mishina-Mori (2005)	2 Japanese-English bilingual children Initial age = 1 yr. 11 mos. to 2 yrs. 4 mos.; age at end = 3 yrs. 2 mos. to 3 yrs. 3 mos.	To investigate the effects of simultaneous bilingualism on the development of both Japanese and English in children.	Speech samples of spontaneous parent-child play interactions were collected in Japanese and English.	Children's productions of interrogatives was assessed using spontaneous language sampling (Japanese and English; CHAT, CLAN, MLU). Children were observed for approximately one year (number of observations ranged from 11 to 14).
Morita (2003)	2 Japanese-English bilingual children Initial age = 4 yrs. 7 mos. to 5 yrs.	To examine the use of personal reference systems in the spontaneous conversations of Japanese-English DLLs.	Both the English- and Japanese-speaking environments of each bilingual child were examined.	Children's pragmatics was assessed using participant-observation ethnographic approach (Japanese and English). Data were collected over 19 hours in a 12-week period.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Mushi (2002)	42 immigrant DLLs from various non-English language backgrounds 18 mos. to 5 yrs.	To examine attitudes and language practices of immigrant parents and teachers.	Parent-child linguistic interactions were assessed using audio recordings of English communication.	Parents' attitudes and preferences toward English and their L1 were assessed using a parent questionnaire developed by the researchers. Parent-child linguistic interactions were assessed using audio recording (English) and an observation checklist developed by the researchers (L1 and English).
Páez, Tabors, & Lopez (2007)	319 Spanish-English bilingual children in the U.S. (ECS) and 144 Spanish monolingual children in Puerto Rico (PRC). ECS group: Initial mean age = 4.4 yrs.; mean age at end = 4.9 yrs. PRC group: Initial mean age = 4.5 yrs.; mean age at end = 4.9 yrs.	To examine oral language and early literacy skills of Spanish-English bilingual children and compare their development with that of Spanish-speaking monolingual children.	Children in the ECS group were administered all measures in English and in Spanish (assessed in their stronger language first, as determined by parent report). Children in the PRC group were assessed only in Spanish.	Children's phonological awareness was assessed using the Phonological Awareness Task developed by the researchers (Spanish and English; rhyme recognition, rhyme production, initial phoneme recognition, sentence segmenting, and syllable segmenting subtests). Four subtests of the WLPB-R were administered to children in Spanish and English: Letter-Word Identification, Dictation, Picture Vocabulary, and Memory for Sentences. Children were assessed in the fall and spring.
Paradis & Navarro (2003)	1 Spanish-English bilingual child (Cuban/Panamanian Spanish), and 2 Spanish monolingual children (Spain) Initial age: 1 yr. 8 mos. to 1 yr. 9 mos.	To examine cross-language relationships in subject use by Spanish-English bilinguals.	Naturalistic language samples were collected for all children.	Children's input and production of overt subjects were assessed using naturalistic language samples (Spanish; CHAT, CLAN, MLUw). Children were assessed once a month over a period from 4 months to 10 months.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Parra, Hoff, & Core (2011)	41 Spanish-English bilingual first language learners Initial mean age = 1 yr. 11 mos.; mean age at end = 2 yrs. 2 mos.	To examine the relationships among phonological memory, language experience, and language development in Spanish-English DLLs.	Children were assessed in both languages. Primary caregivers completed the CDI/Inventario II in their native language; or completed both inventories if they were Spanish-English proficient.	Children's language exposure was assessed using the HLEQ (Spanish/English). Children's phonological memory was assessed using a nonword repetition task developed by the researchers (Spanish and English, PCC). Children's vocabulary and grammatical development was assessed using the CDI and its Spanish counterpart, the Inventario II (English and Spanish; Words Children Use, Sentences and Grammar). Children were assessed at 22 and 25 months.
Patterson (2002)	64 Spanish-English bilingual children 1 yr. 9 mos. to 2 yrs. 3 mos. (mean = 2 yrs.)	To examine the relationships between home literacy practices, television watching, and expressive vocabulary size in Spanish-English DLLs.	Approximately half of the words on the SEVC are in Spanish; the remaining half are in English.	Children's exposure to language was assessed using a parent interview (Spanish/English). Children's expressive vocabulary was assessed using the parent-report SEVC (Spanish and English)
Peña, Bedore, & Zlatic-Giunta (2002)	44 Spanish-English bilingual children 4 yrs. 5 mos. to 7 yrs. 1 mo. (mean = 5 yrs. 9 mos.). Children below the mean age constituted the younger group (mean = 5 yrs. 1 mo.); those at or above the mean age constituted the older group (mean = 6 yrs. 5 mos.).	To examine Spanish-English bilingual children's performance on a category-generation task as a means of understanding qualitative changes in vocabulary development.	The category-generation task was administered to children in English and Spanish.	Children's categorization strategies were assessed using a category-generation task adapted by the researchers (English and Spanish).
Perry, Kay, & Brown (2008)	13 Latino families whose primary language was Spanish Initial age = 2.7 yrs. to 4.9 yrs. (mean = 3.8 yrs.). Maternal mean age = 32.7 yrs.	To investigate how Latino immigrant parents incorporated school-based literacy activities into their home literacy practices.	Instructions and written literacy materials were provided in English and Spanish. All journal entries were entered in Spanish and translated into English for analysis.	How families incorporated school-based interactive literacy activities into their existing home practices was assessed using home literacy activities, parent journals (Spanish), and teacher reports.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Reyes (2006)	3 Spanish-English bilinguals Mean age = 4 yrs.	To examine emergent biliteracy development in Mexican Spanish-English DLLs and their home language and literacy practices.	n.a.	Children's emergent biliteracy was assessed using observation, field notes, writing samples, informal conversation with participants and their parents/teachers, video recordings of children's interactions with different family members, peers, and teachers.
Reyes & Azuara (2008)	12 Spanish-English bilingual Mexican immigrant children 4 yrs. to 5 yrs.	To examine the effect of a biliterate environment on the development of biliteracy in Mexican Spanish-English DLLs.	Children receptive language was assessed in English and Spanish.	Children's emergent literacy was assessed using the Environmental Print Awareness Task developed by the researchers, a book handling task developed by the researchers, the State Arizona Language Assessment, and the PPVT-R/TVIP. Children's context and specific language environments were assessed using a child interview and home observations (TRANSANA). Children were assessed over a year at an average of twice a month for at least two hours per visit.
Rinaldi & Páez (2008)	234 Spanish-English bilingual children Mean age of preschoolers = 4.6 yrs.; kindergarteners = 5.5 yrs.; and first graders = 6.6 yrs.	To examine the relationship among DLLs' Spanish and English word reading skills, phonological awareness, and oral language abilities.	All children were assessed in English and Spanish.	Children's phonological awareness was assessed using a task developed by the researchers (Spanish and English; rhyme recognition, rhyme production, initial phoneme recognition, sentence segmenting, and syllable segmenting subtests). Three subtests of the WLPB-R were administered to children in Spanish and English: Letter-Word Identification, Picture Vocabulary, and Memory for Sentences. Children were assessed longitudinally at the end of the pre-kindergarten, kindergarten, and first grade.
Schnitzer & Krasinski (2003)	2 Spanish-English bilingual siblings Initial age = 1 yr. 1 mo. to 1 yr. 6 mos.; age at end = 3 yrs. 9 mos. to 4 yrs. 6 mos.	To investigate language maintenance strategies used by DLLs who were acquiring unrelated and related languages.	Video recordings were collected at least monthly, alternating the two languages (mother speaking Spanish one month, father speaking English the following month).	Children's phonology processes were assessed using language diaries (Spanish and English) and video recordings (Spanish and English)

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Sheng, McGregor, & Marian (2006)	12 Mandarin Chinese-English bilinguals and 12 English monolingual children 5 yrs. 7 mos. to 8 yrs. 5 mos. (mean = 7 yrs. 1 mo.)	To examine lexical-semantic organization in both languages of Mandarin Chinese-English DLLs and to compare results with pre-existing English monolingual data.	The bilingual children were tested in Mandarin and English during two separate sessions, with a different version of the word association test used each time. The monolingual children were tested once in English. The PPVT-III was administered in English to children in both groups.	Children's lexical-semantic organization was assessed using a Repeated Word Association Test (Mandarin and/or English) and the PPVT-III (English). Monolinguals served as controls for the bilingual children; groups were matched on the Matrices subtest of the K-BIT.
Shin (2002)	Parents of 204 Korean-English bilingual children 4 yrs. to 18 yrs.	To examine the effect of birth order on English and Korean knowledge and use in Korean-English DLLs.	Only a parent questionnaire was used in this study.	Children's heritage language acquisition was assessed using a parent questionnaire developed by the researchers (Korean).
Silva-Corvalan & Montanari (2008)	1 Spanish-English bilingual child Initial age = 1 yr. 6 mos.; age at end = 2 yrs. 11 mos.	To examine the influence of adult input on the development of copula usage and the conceptual and syntactic contexts in which copulas occur in Spanish-English DLLs.	Spontaneous speech samples were collected in Spanish and English.	Child's copula and auxiliary uses of "ser," "estar," and "be" were assessed using diary notes (Spanish and/or English) and spontaneous speech samples (Spanish and/or English; MLUw). Data were collected every two to four days per week.
Tabors, Páez, & López (2003)	344 Spanish-English bilingual children in the U.S. (ECS) and 152 Spanish monolingual children in Puerto Rico (PRC) 4 yrs. of age	To examine and compare the language and emergent literacy skills of Spanish-English DLLs with that of Spanish monolingual children.	ECS group: Children were assessed in English and Spanish. Children who were tested first in Spanish (63%) were tested in English an average of 12 days later; children who were tested first in English were tested in Spanish an average of 15 days later. PRC group: Children were assessed only in Spanish.	Children's phonological awareness was assessed using the Phonological Awareness Task developed by the researchers (Spanish or Spanish/English; rhyme recognition, rhyme production, initial phoneme recognition, sentence segmenting, and syllable segmenting subtests). Four subtests of the WLPB-R were administered to children in Spanish or Spanish/English: Letter-Word Identification, Dictation, Picture Vocabulary, and Memory for Sentences.

Table A.1 *(continued)*

Study	Sample	Study Purpose	Approach	Assessment Tools Used
Vagh, Pan, & Mancilla-Martinez (2009)	29 English monolingual children and 56 Spanish monolingual or Spanish-English bilingual children Initial age: 2 yrs. to 3 yrs. (n = 34 age 2 yrs., n = 24 age 2 yrs. 6 mos., and n = 27 age 3 yrs.)	To compare growth in English productive vocabulary of bilingual and monolingual children, and examine the utility and validity of supplementing parent reports with teacher reports of child language.	Parents and teachers reported on children's English vocabulary knowledge. Direct child assessments were conducted only in English.	Children's vocabulary was assessed using the CDI (English, parent and teacher report), the WLPB-R (English; Picture Vocabulary), and the PPVT-III (English). Parents and teachers completed the CDI checklist when children were approximately 24, 27, 30, 33, and 36 months of age. Direct child assessments were obtained when children were 30 and 36 months of age.
Yaden & Tardibuono (2004)	47 Spanish-speaking preschoolers 4 yrs. of age	To examine the early writing development of 4-year-old Spanish-speaking preschoolers.	All children were assessed in Spanish.	Children's written language development was assessed using a General Writing Task (GW) and a Name-Writing Task (NW) developed by the researchers (Spanish). Children were administered the GW tasks in February and June and completed the NW task in April.
Yavas & Core (2001)	24 bilingual Spanish-English speaking first grade children were recruited and compared to sample of 38 English monolingual first graders in Yavas and Gogate (1999) Mean age or range not specified	To examine whether Spanish-English bilingual children differ from monolingual English-speakers in their ability to segment phonemes.	All children were assessed in English.	Children's phonemic awareness was assessed using a coda consonant phoneme deletion task developed by the researchers (English).

Note: L1 = first language learned; L2 = second language learned

Table A.2. Approach to Assessing Dual Language Learners in Government Reports

Study	Sample	Study Purpose	Approach	Language and Literacy Assessment Tools Used
Administration for Children and Families (DHHS), Office of Planning (2006)	2,800 children and families in 43 Head Start programs Children were 3 and 4 years old at program entry in fall 2000	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how children, families, and programs are performing currently and over time. This report provides information about the knowledge and skills that children have when they enter the Head Start program in fall 2000 and gains they made during the Head Start year and through the first year of elementary school.	Spanish-speaking children in the FACES sample were assessed in Spanish unless their teachers reported they had sufficient command of English to be assessed in English. When determined to be a primarily Spanish-speaker, children received the entire battery in Spanish. When tested one or two years later, they received the battery in English and were also administered the TVIP and Woodcock Munoz Letter-Word Identification in Spanish for comparison. Children who had been assessed in Spanish and English in fall 2000 with some Spanish sections were administered the entire assessment in English during the spring of their kindergarten year (either spring 2002 or 2003). DLLs from other language backgrounds were assessed in English in the spring, and subsequent years.	PPVT-III, WJ-R/WM-R Tests of Achievement, Story and Print Concepts, Color names
Aikens, Hulsey, Moiduddin, Kopack, Takyi-Laryea, Tarullo, & West (2011)	3,349 children who were newly enrolled in Head Start Children were 3 and 4 years old at program entry in fall 2009	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the children, families, and programs are performing currently and over time. These data tables provide information about the knowledge and skills that children have when they enter the Head Start program in fall 2009.	Language screening tool was used in combination with parent report of home language to determine whether children from households where English was not the primary spoken language should be assessed in English, in Spanish, or administered only the PPVT-4 and EOWPVT.	PreLAS, PPVT-4, EOWPVT/EOWPVT-SBE, TVIP, WJ-III/WM-III Tests of Achievement

Table A.2 *(continued)*

Study	Sample	Study Purpose	Approach	Language and Literacy Assessment Tools Used
Chernoff, Flanagan, McPhee, Park, & National Center for Educational Statistics (2007)	3,940 children; 53% white non-Hispanic, 13.8% Black non-Hispanic, 25.1% Hispanic, 2.6% Asian non-Hispanic, 0.5% American Indian and Alaska Native non-Hispanic, 4.2% other non-Hispanic Children ages birth to 4 years nationally	Selected descriptive data about the preschool experiences and development of a nationally representative sample of children born in 2001	If a child's family spoke a language other than English or Spanish, interviewers used an interpreter or family member (if an interpreter was not available).	ECLS-B Language and Literacy, Color knowledge, PreLAS
Hulsey, Aikens, Xue, Tarullo, West, Administration for Children and Families (DHHS), Office of Planning, et al. (2010)	3,315 children participated in the study in the fall of 2006. A total of 3,296 children were eligible for the spring 2007 follow-up and 88% of the eligible children participated. 62.8% of children entering Head Start in fall 2006 were 3 years old or younger, 37.2% were 4 years old or older.	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the children, families, and programs are performing currently and over time. These data tables provide information about the knowledge and skills that children have when they enter the Head Start program in fall 2006 and gains made during the Head Start year.	Language screening tool used in combination with parent report of home language to determine whether children from households where English was not the primary spoken language should be assessed in English, in Spanish, or administered only the PPVT-4. If a child had been assessed in English in one of the prior rounds, he or she was assessed in English in subsequent rounds.	PreLAS, PPVT-4, TVIP, WJ-III/WM-III Tests of Achievement, Story and Print Concepts (English and Spanish)

Table A.2 *(continued)*

Study	Sample	Study Purpose	Approach	Language and Literacy Assessment Tools Used
Malone, Hulsey, Aikens, West, Tarullo, Administration for Children and Families (DHHS), Office of Planning, et al. (2010)	2,096 children who entered Head Start as a 3- or 4-year-old, completed one or two years in the program, and then attended kindergarten the year after graduating from Head Start	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. These data tables provide information about the knowledge and skills that children have when they enter the Head Start program in fall 2006 and gains made during Head Start and through the first year of elementary school.	Language screening tool used in combination with parent report of home language to determine whether children from households where English was not the primary spoken language should be assessed in English, in Spanish, or administered only the PPVT-4. If a child had been assessed in English in one of the prior rounds, he or she was assessed in English in subsequent rounds.	PreLAS, PPVT-4, TVIP, WJ-III/WM-III Tests of Achievement, Story and Print Concepts (English and Spanish)
Moiduddin, Aikens, Tarullo, & West (2010)	1,203 3-year-old children newly enrolled in Head Start in fall of 2006 and still attending in spring 2008	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. These data tables provide information about the knowledge and skills of children who enter the program at age 3 and gains made during Head Start.	Language screening tool used in combination with parent report of home language to determine whether children from households in which English was not the primary language should be assessed in English, in Spanish, or should be administered only the PPVT-4. If a child had been assessed in English in one of the prior rounds, he or she was assessed in English in subsequent rounds.	PreLAS, PPVT-4, TVIP, WJ-III/WM-III Tests of Achievement, Story and Print Concepts (English and Spanish)

Table A.2 *(continued)*

Study	Sample	Study Purpose	Approach	Language and Literacy Assessment Tools Used
Najarian, Snow, Lennon, & Kinsey (2010)	1st wave: 10,200; 2nd wave: 9, 200; 3rd wave: 3,940 wave 1: 9 mos.; wave 2: 2 yrs.; wave 3: 4 yrs.	To describe the experiences and development of a representative sample of children born in 2001	Language of assessment was determined by parent report of child language for younger children and by PreLAS for older children (only one correct response needed to receive the English assessment). Spanish child assessment scores are not included in the file. Too few children took the Spanish cognitive assessment to meet sample size requirements in IRT analysis, so it was not scored.	ECLS-B abbreviated form for the Bayley Scales of Infant Development; English and Spanish-language versions of the PreLAS; TVIP; ECLS-B Language and Literacy Assessment
Puma, Bell, Cook, Heid, Shapiro, Broene, et al. (2010)	In the Puerto Rico sample, there are fewer than 200 children with completed assessments and a parent interview each spring. This sample size was too small to reliably estimate regression models containing baseline child covariates. U.S. sample of 4,667 newly entering children, including 2,559 in the 3-year-old group and 2,108 in the 4-year-old group.	Randomized controlled study of Head Start. Child outcomes provide measures of how well Head Start and non-Head Start preschool programs, or other child care programs, are achieving the goal of assisting children to be physically, socially, and educationally ready for success in school	Main care provider was asked 3 questions regarding a child's language ability. If two or more of the responses were English or Spanish, the child was tested in that language. When children received the assessment in Spanish, it was a bilingual assessment. Tests included the complete fall 2002 Spanish assessment battery and two English tests (PPVT and WJ-III Tests of Letter-Word Identification). In spring 2003 the bilingual assessment included the complete English battery and the TVIP and WM-III Tests of Letter-Word Identification). All children from Puerto Rico were assessed in Spanish at all data collection times. If a child's primary language was anything other than Spanish or English, a teacher decided if they could understand the assessment in English. If they could not, four tests (McCarthy Draw-a-Design, Color Names and Counting, Leiter-R-adapted and Story and Print concepts) were translated for the children.	PPVT-III, CTOPPP, Color names, Story and Print Concepts, Letter Naming, Writing name task, WJ-III Tests of Achievement

Table A.2 *(continued)*

Study	Sample	Study Purpose	Approach	Language and Literacy Assessment Tools Used
Vogel, Boller, Xue, Blair, Aikens, Burwick, et al. (2011)	Representative of the population of Early Head Start programs at the national level in spring 2009. Selected all children receiving center- and/or home-based services. Expected sample sizes: 1,987 newborn and 1-year-olds at study recruitment. Total study-eligible DLLs in sampled programs = 441. Effective sample size of DLL = 274. 1 -year-old cohort included children who were between 10 and 15 months at the spring 2009 visit; the newborn cohort included babies up to age 2 months at the spring 2009 visit, plus any pregnancies likely to result in a baby between 10 and 14 months at the time of the spring 2010 visit.	To describe the experiences and development of children and families in Early Head Start nationally in spring 2009.	Child assessments were all parent reports in the first year and were completed in the parent's preferred language.	Ages and Stages Questionnaires (ASQ-3) Communication; CDI, Inventarios
West, Tarullo, Aikens, & Hulsey (2008)	3,315 children participated in the study in the fall of 2006 62.8% of children entering Head Start in fall 2006 were 3 years old or younger, 37.2% were 4 years old or older.	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. These data tables provide information about the knowledge and skills that children have when they enter the Head Start program in fall 2006.	Language screening tool used in combination with parent report of home language to determine whether children from households in which English was not the primary language should be assessed in English, in Spanish, or should be administered only the PPVT-4.	PreLAS, PPVT-4, TVIP, WJ-III/WM-III Tests of Achievement, Story and Print Concepts (English and Spanish)

Table A.2 *(continued)*

Study	Sample	Study Purpose	Approach	Language and Literacy Assessment Tools Used
Zill, Sorongon, Kim, Clark, & Woolverton (2006)	2,400 children who entered Head Start when they were 3 or 4 years old and completed one year of the program.	FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This report provides information about the knowledge and skills that children have when they enter the Head Start program in 2003 and at the end of the first year.	Language screening tool used in combination with parent report of home language.	PreLAS, PPVT-III (48-item adapted version), TVIP, WJ-R/WJ-III Tests of Achievement, Batería WM-R, Pre-CTOPP, Story and Print Concepts

GOVERNMENT STUDIES

The studies included in this section are longitudinal and so have multiple reports across different years of the study. This section is arranged by study name.

Baby FACES: Early Head Start Family and Child Experiences Survey

Vogel, C. A., Boller, K., Xue, Y., Blair, R., Aikens, N., Burwick, A., . . . Stein, J. (2011). Learning as we go: A first snapshot of Early Head Start programs, staff, families, and children report and technical appendices (OPRE report #2011-7). Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- The purpose of this study was to provide information about the knowledge and skills that children have in the Early Head Start program and the gains they made in the program, as well as to provide information about their families, programs, and staff. These reports focused on the data for 1-year-olds.
- The sample included 1,987 newborn and 1-year-olds receiving center-based and/or home-based Early Head Start services.
- Parents completed the MacArthur-Bates Communicative Development Inventory (CDI; Fenson et al., 2000) and the Ages and Stages Questionnaires-3 (ASQ-3; Squires, Twombly, Bricker, & Potter, 2009) with items from more than one age form to avoid ceiling problems. The ASQ-3 is a screening assessment and so each form is designed to assess whether children are exhibiting age-expected skills; items from the next age form were added to each administration to describe the range of skills children exhibit. The Communication section of the ASQ-3 is the section that addresses language development.
- If children received any services in Spanish, at least one bilingual member was assigned to the team visiting the program. Parents, teachers, and home visitors who reported they spoke Spanish completed the Spanish version of CDI for children identified as understanding Spanish.
- The staff completed Spanish short form CDIs for 137 children.
- Internal consistency reliability for the full sample was very strong for the English CDI (= .95 to .97) and slightly weaker, though still very strong, for the Spanish CDI with a smaller sample size (= .87 to .98). The ASQ-3 Communication scale had weak to moderate internal consistency (= .65 to .73), although the total ASQ-3 had stronger reliability (= .78 to .84).
- Reliability estimates for the 1-year-old cohort were also presented by DLL status:
  - CDI reliability was strong for each group, particularly for the English speakers (= .95 to .98), with similar estimates for Spanish and other language DLLs (= .90 to .98). In each case, reliability was stronger for vocabulary comprehension than for vocabulary production.

  - ASQ-3 Communication reliability ranged from .58 to .77 for English speakers and from .64 to .71 for Spanish DLLs. The sample for other language DLLs was too small to estimate reliability. The ASQ-3 total score reliability was .79 to .86 for English speakers and .77 to .80 for Spanish DLLs.

- At Baby FACES baseline (spring 2009), the correlations of CDI scores with ASQ-3 Communication were very low, indicating little or no relationship between them and suggesting that the Communication scale is measuring areas unrelated to receptive and expressive vocabulary. The ASQ-3 Communication scores correlated with English CDI Vocabulary Comprehension at 0.08 and with English CDI Vocabulary Production at 0.14. The reported correlations with ASQ-3 Communication were 0.02 for Spanish Vocabulary Comprehension and 0.12 for Spanish Vocabulary Production.

Early Childhood Longitudinal Study–Birth Cohort (ECLS-B)

Andreassen, C., and Fletcher, P. (2005). Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), psychometric characteristics. Volume 1 of the ECLS-B methodology report for the nine-month data collection (NCES 2005-100). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Flanagan, K., and West, J. (2004). Children born in 2001: First results from the base year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (NCES 2005–036). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

- The purpose of these reports was to provide information on a wide variety of children's experiences nationally for the 9-month data collection period of the ECLS-B. Seventy-two percent of the children were between 8 and 10 months old, but the full range of ages at the first data collection was from 6 months to 22 months.

- Approximately 10,200 9-month-old children are included in the analysis. Fifty-four percent were white, non-Hispanic; 14% were Black, non-Hispanic; 26% were Hispanic, 3% were Asian/Pacific Islanders, non-Hispanic; 1% were American Indian, non-Hispanic; and 4% were multiracial, non-Hispanic.

- All children were assessed with the Bayley Short Form Research Edition (BSF-R), a shortened version of the Bayley Scales of Infant Development-II (BSID-II; Bayley 1993) developed for the ECLS-B. The BSF-R includes a mental scale and a motor scale. A Spanish version was developed, but information about the psychometric characteristics of that version was not provided.

  - The reliability of the IRT-based theta for the mental scale was .79.

  - Five proficiency levels were created from the mental scale for the 9-month results based on the hierarchical nature of the data. Two of these levels (the third and fifth levels) assess language skills: babbles (including items about use of gestures and sounds to communicate) and uses words (including items about receptive and expressive vocabulary).

Mulligan, G. M., & Flanagan, K. D. (2006) Age 2: Findings from the 2-year-old follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (NCES 2006-043). U.S. Department of Education. Washington, DC: National Center for Educational Statistics.

- The purpose of this report was to provide information on children's development, health, and in- and out-of-home experiences nationally at the age 2 data collection of the ECLS-B.
- Approximately 8,950 children, ranging from 22 months to 25 months, are included in the analysis in the report.
- All children were assessed with the Bayley Short Form Research Edition (BSF-R), a shortened version of the Bayley Scales of Infant Development-II (BSID-II; Bayley 1993). The BSF-R includes a mental scale and a motor scale. No information was provided about the Spanish version. The ECLS-B data file contains proficiency probability scores related to language and literacy that are derived from the mental scale of the BSF-R (Mulligan et al., 2006, pp. 38–39):
  - **Jabbers Expressively [X2MTL_C].** This proficiency assesses communication through diverse nonverbal sounds and gestures, for example, vowel and vowel-consonant sounds, gesturing for an object, and jabbering expressively (e.g., jabbering with inflection and change in tone of voice).
  - **Names Object [X2MTL_E].** This proficiency measures a series of early communication skills, such as saying simple words like "mama," or "dada," knowing an object by its name (e.g., pointing to his or her foot when asked), and saying the name of an object (e.g., seeing something such as a toy car and saying the word "car").
  - **Receptive Vocabulary [X2MTL_F].** This proficiency can be characterized as the ability to recognize and understand spoken words or to indicate a named object by pointing. For example, when asked to point to a picture of a "shoe," the child points to the correct picture.
  - **Expressive Vocabulary [X2MTL_G].** This proficiency refers to children's verbal expressiveness using gestures, words, and sentences. For example, the child may name objects, name pictures of objects, use possessive pronouns (e.g., "mine," "my," "yours"), or combine two or more words when talking (e.g., "more milk").
  - **Listening/Comprehension [X2MTL_H].** This proficiency refers to children's ability to understand actions depicted by a story, in pictures, or by verbal instruction. For example, the child attends to a story when read to and displays verbal comprehension of the story (e.g., within the story, child can point to the corresponding picture when asked); the child understands the use of prepositions (e.g., when asked to put a stuffed animal on top of a blanket, the child does as asked and understands "on top of"); or the child spontaneously generates words to describe a picture (e.g., "doggie sleep" in reference to a picture of a dog asleep.)
- Expected increases in skill by age were evident in the data. Across the 22- to 25-months age groups, the percentage of children demonstrating proficiency

increased for all the language proficiency scores except "jabbers expressively," and item difficulty varied in expected ways. All children demonstrated proficiency on "jabbers expressively" and more children demonstrated proficiency in receptive vocabulary than in expressive vocabulary and listening comprehension.

Jacobson Chernoff, J., Flanagan, K. D., McPhee, C., & Park, J. (2007). Preschool: First findings from the preschool follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) (NCES 2008-025). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

- The purpose of this study was to describe the development and experiences of preschool children who were born in 2001.
- Approximately 8,750 children participated in the study beginning at age 9 months and ending data collection at preschool age (4 years). The racial/ethnic distribution was 53% white non-Hispanic, 13.8% Black non-Hispanic, 25.1% Hispanic, 2.6% Asian non-Hispanic, 0.5% American Indian and Alaska Native non-Hispanic, 4.2% other non-Hispanic.
- For the 9- and 24-month-old assessments, if a child's family spoke a language other than English or Spanish, interviewers used an interpreter, or a family member if an interpreter was not available.
- For the preschool-year assessment, the PreLAS was administered, along with parent report of home language, to determine language of administration. If the child failed all of the English language items after the practice items, the child did not receive the cognitive assessments (language and literacy, mathematics, and color knowledge) in English. Although some children did take a Spanish assessment, the sample sizes were too small to include scores on the file.
- For the English assessments (that included DLLs), the reliability was strong:
  - Reading, Preschool IRT-based scores, n=8,350, = .84
  - Math, Preschool IRT-based scores n=8,300, = .89
  - Color knowledge test, n=8400, =.82

Najarian, M., Snow, K., Lennon, J., & Kinsey, S. (2010). Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Preschool–Kindergarten 2007 Psychometric Report (NCES 2010-009). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

- The purpose of this study was to provide a comprehensive and reliable set of data that may be used to better understand and describe children's early development.
- 10,200 9-month-old children were assessed in the first wave; 9,200 2-year-old children were assessed in the second wave; 8,750 children were assessed in the preschool wave.
- Language of assessment was determined by both child performance on the PreLAS and parent report of child language. If the child failed all of the

English language items after the practice items, the child did not receive the cognitive assessments (early reading,[13] mathematics, and color knowledge) in English. Once determined, the language of assessment was constant and all components were conducted in the determined language. Spanish child assessment scores are not included in the file. Too few children took the Spanish cognitive assessment to meet sample size requirements in IRT analysis, so it was not scored.

- 8,550 children were found to be capable of taking the assessments in English and 200 children had insufficient English fluency.
- Acceptable reliability was found for the language and literacy assessments:
  - Early reading assessment preschool IRT-based scores: n = 8,350, r = .84
  - Color knowledge test: n = 8,400, r = .82
- Strong correlations of the early reading IRT with the mathematics IRT score were found across preschool (r = .76) and kindergarten (r = .77 to .81).
- Stability of the early reading assessment from preschool to kindergarten (r = .58 to .65) was weaker than for the mathematics assessment (r = .64 to .72).

Najarian and colleagues (2010) explain a change from language and literacy scale scores to an early reading scale score in the longitudinal files:

> While data from the preschool wave alone supported the development of unique scores for the dimensions of language and literacy (i.e., performance on the language-based items varied uniquely from performance on the literacy-based items), once the preschool data were pooled with the kindergarten 2006 data, it was determined that separate language and literacy scores were no longer appropriate. The longitudinal model for the preschool and kindergarten 2006 and 2007 waves supported a unidimensional early reading domain, which reflects children's performance on certain language-based items (receptive language/PPVT items) and literacy items (e.g., conventions of print, letter recognition, understanding of letter-sound relationships, phonological awareness, sight word recognition, understanding words in the context of simple sentences). As a result, the re-estimated preschool IRT thetas and resulting scale scores available in the 9-month kindergarten 2007 data file replace the preschool scores in the 9-month–preschool file previously released. That is, there is now a single early reading score for the preschool data and there are no longer separate language and literacy scores. Finally, in consultation with psychometricians, IRT-based subscale scores presented on the preschool data file have been dropped from the data set. (pp. 76–77)

[13] The language and literacy assessments formed a single scale in kindergarten, so the longitudinal file has a single score that includes both language and literacy.

Head Start Family and Child Experiences Survey 1997 (FACES 1997)

Administration for Children and Families (DHHS), Office of Planning. (2001). Head Start FACES: Longitudinal Findings on Program Performance Third Progress Report. Washington, DC: Administration for Children & Families, author.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This report provides information about the knowledge and skills that children had when they entered the Head Start programs in fall 1997 and the gains they made during the Head Start year as well as through their kindergarten/first-grade year of elementary school.
- A total of 3,200 children and families in 40 Head Start programs from a nationally stratified random sample participated. The children were studied from entry in the program in fall 1997 through spring of the kindergarten or first grade years.
- Language of assessment was determined by whether the child was Spanish-speaking. Spanish-speaking children took the assessment in Spanish unless their teachers reported they were able to take it in English. Children participating in a Head Start program in Puerto Rico with classroom instruction conducted in Spanish were tested in Spanish.
- 345 children were assessed in Spanish in fall 1997 and 120 children were assessed in Spanish in spring 1998. Mean scores were reported but no evidence of psychometric properties was available in this report.
- Information about the predictive validity of the measures used in FACES 1997 was included in the FACES 2000 report (Zill et al., 2003). The sample was limited to children who took the assessments in English at all timepoints and so did not include many DLLs. However, these estimates provide a point of comparison for validity estimates for low-income populations on these measures:
  - Spring Head Start WJ-R Letter-Word Identification score correlated with end of kindergarten ECLS-K Reading scale score (r = .55; beta = .32) and with the ECLS-K General Knowledge scale score (r = .39; n.s. in the multiple regression)
  - Spring Head Start WJ-R Dictation score correlated with end of kindergarten ECLS-K Reading scale score (r = .48; beta = .14) and with the ECLS-K General Knowledge scale score (r = .46; beta = .11)

Head Start Family and Child Experiences Survey 2000 (FACES 2000)

Administration for Children and Families (DHHS), Office of Planning. (2006). Head Start performance measures center, family and child experiences survey (FACES 2000) technical report. Washington DC: Administration for Children and Families, author.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This report provides

information about the knowledge and skills that children had when they entered the Head Start program in fall 2000 and the gains they made during the Head Start year as well as through their kindergarten year of elementary school.

- 2,800 children (ages 3 and 4 at program entry) and families in 43 Head Start programs participated.
- A screener using information from parents and teachers determined the language of administration. Children were assessed in English if teachers reported children had sufficient command of English. Spanish-speaking children in the FACES sample without sufficient command of English were assessed in Spanish. When a child was determined to be a primarily Spanish speaker, he or she received the entire battery in Spanish. When tested one or two years later these children received the battery in English but were also administered the TVIP and Woodcock-Muñoz Letter-Word Identification in Spanish for comparison.
- Children who had been assessed in Spanish and English in fall 2000 with some Spanish sections were administered the entire assessment in English during the spring of their kindergarten year (either spring 2002 or 2003).
- DLLs whose home language was something other than Spanish were not assessed in the fall. They received the full assessment battery in English in the spring.
- The number of children tested in English ranged from 803 to 1,908 on different language assessments. The number of children tested in Spanish at the first data collection time and in English at the second data collection time ranged from 172 to 356 across the different assessments.
- Reported reliability (internal consistency) was acceptable to strong (> .70) at each of the timepoints with the exception of the Story and Print Concepts test, which had weaker reliability. Internal consistency was stronger for the English tests than for the Spanish tests:
  - Peabody Picture Vocabulary Test-III (PPVT-III) scores: = .96 to .97.
  - Test de Vocabulario en Imagenes Peabody (TVIP), = .92 to .94.
  - Woodcock-Johnson (WJ) Letter-Word Identification subtest = .84 to .86. The Spanish version of this subtest (Woodcock-Muñoz Letter-Word Identification) had weaker reliability estimates: .75 for fall 2000, .78 for spring 2001, and .83 for spring 2002 (Head Start).
  - WJ Dictation subtest with FACES children averaged =.77 for fall 2000, spring 2001, and spring 2002/2003 (kindergarten), and =.71 for spring 2002 (Head Start). The Spanish version of this subtest (Woodcock-Muñoz) was =.77 for fall 2000.
  - Letter Naming Task reliability was not reported
  - Color Naming task: = .95 for fall 2000, =.94 for spring 2001, and =.90 for spring 2002 (Head Start). The Spanish version of the Color Naming task was =.92 for fall 2000.

- The Phonemic Analysis test from the Test of Language Development supplemental subtest with FACES children: = .96 for spring 2002/2003 (kindergarten).
- Early Childhood Longitudinal Study-Kindergarten cohort (ECLS-K) measures, with FACES sample in spring kindergarten:
  - English Reading—spring 2002/2003 (kindergarten): Routing = .87, Red (low form) = .95, Yellow (middle form) and Blue (high form) = .94
  - English General Knowledge subtest: = .77 for spring 2002/2003 (kindergarten)

- Most reliability estimates for Story and Print Concepts were low, with those for the Spanish version being even weaker:
  - Book Knowledge: = .57 for fall 2000, = .59 for spring 2001, = .61 for spring 2002
  - Print Conventions: =.73 for fall 2000, =.75 for spring 2001,and = .84 for spring 2002
  - Comprehension: = .43 for fall 2000, =.42 for spring 2001, and = .40 for spring 2002
  - For the Spanish version, Book Knowledge =.43 for fall 2000, Print Conventions =.59 for fall 2000, Comprehension =.39 for fall 2000.

Head Start Family and Child Experiences Survey 2003 (FACES 2003)

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: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This brief provides information about the knowledge and skills that children had when they entered the Head Start programs in fall 2003 and the gains they made during the first year in Head Start.
- No information was reported separately for DLLs and no information about the reliability of scores was provided in this brief.

Head Start Family and Child Experiences Survey 2006 (FACES 2006)

Hulsey, L., Aikens, N., Xue, Y., Tarullo, L., West, J., Administration for Children and Families (DHHS) Office of Planning, et al. (2010). Data tables for FACES 2006: A year in Head Start report. ACF-OPRE report. Administration for Children & Families.

West, J., Tarullo, L. Aikens, N., & Hulsey, L. (2008). Study design and data tables for FACES 2006 baseline report. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. The report by West et al. (2008) provides information about the knowledge and skills that children had when they entered the Head Start program in fall 2006, while the report by Hulsey et al. (2010) describes children's skills and the gains they made during the first year in Head Start.
- A sample of Head Start programs was selected from the 2004–2005 Head Start Program Information Report (PIR), and 3,315 children participated in the study in the fall of 2006. A total of 3,296 children were eligible for the spring 2007 follow-up and 88% of the eligible children participated. Of all children entering Head Start in fall 2006, 62.8% were 3 years old or younger and 37% were 4 years old or older.
- The PreLAS language screener was used in combination with parent report of home language to determine whether children from households where English was not the primary spoken language should be assessed in English, assessed in Spanish, or not assessed except for the PPVT-4. The PPVT-4 was administered to all children in order to obtain an estimate of progress in learning English. The TVIP was also administered to children who spoke Spanish at home. Children who spoke Spanish at home and made five consecutive errors on both Simon Says and Art Show were administered the Spanish assessment battery (along with the PPVT-4). Children who spoke a language other than English or Spanish who made five consecutive errors were not administered the PPVT-4. If a child had been assessed in English in the fall, he or she was assessed in English in the spring regardless of his or her spring performance on the language screener. Only the Simon Says from the PreLAS was administered to these children and was used as a warm-up rather than a decision tool.
- The number of children tested in English in both fall 2006 and spring 2007 ranged from 2,648 to 2,680 on different direct child assessments (Hulsey et al., 2010). The number of children with a valid score in Spanish at both waves ranged from 106 to 162 on different direct child assessments (Hulsey et al., 2010). At baseline, 405 to 411 were assessed in Spanish using the subtests of the Woodcock-Muñoz-III, with 716 children taking the TVIP (West et al. 2008).
- Reliability of English and Spanish directly administered language and literacy assessments in spring 2007 with the FACES 2006 data indicated strong reliability for all measures except Story and Print Concepts (= .70) and the Woodcock-Muñoz-III Ortografía (Spanish version of Spelling; = .67):

Table A.3. Psychometric Properties of Assessments in FACES 2006

Assessments	Cronbach's Alpha	
	(Hulsey et al., 2010)	(West et al., 2008)

PPVT-4	0.95	0.97
TVIP	0.94	0.93
WJ-III: Letter-Word Identification	0.86	0.81
WJ-III: Spelling	0.81	0.79
WJ-III: Applied Problems	0.87	0.88
Story and Print Concepts	0.70	0.70
WM-III: Letter-Word Identification	0.82	0.66
WM-III: Spelling	0.67	0.69
WM-III: Applied Problems	0.88	0.84

The indirect assessments of literacy had adequate reliability for the teacher reports but not for the parent reports (teacher-reported child literacy skills, = 0.84; parent-reported emergent literacy scale, = 0.48).

Malone, L., Hulsey, L., Aikens, N., West, J., Tarullo, L., Administration for Children and Families (DHHS), Office, of Planning, et al. (2010). Data tables for FACES 2006: Head Start children go to kindergarten report. ACF-OPRE report. Administration for Children & Families.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This report describes the performance of children across two different intervals: (1) between Head Start entry and the end of kindergarten and (2) as they go from preschool to kindergarten.
- The report of child outcomes focuses on (1) the 1,250–1,404 cases assessed in English at Head Start entry, exit, and end of kindergarten and (2) the 1,611–1,647 cases assessed in English at Head Start exit and end of kindergarten.
- The PreLAS language screener was used in combination with parent report of home language to determine whether children from households where English was not the primary spoken language should be assessed in English, assessed in Spanish, or not assessed except for the PPVT-4. The PPVT-4 was administered to all children in order to obtain an estimate of progress in learning English. The TVIP was also administered to children who spoke Spanish at home. Children who spoke Spanish at home and made five consecutive errors on both Simon Says and Art Show were administered the Spanish assessment battery (along with the PPVT-4). Children who spoke a language other than English or Spanish and made five consecutive errors were not administered the PPVT-4. Once a child was assessed in English, he or she was assessed in English in all subsequent rounds of data collection.
- The number of children in this sample who had test scores in English at all three timepoints ranged from 1,250 to 1,392 on different direct child assessments. The number assessed at baseline ranged from 2,592 to 2,645. The number of children tested in Spanish at baseline ranged from 405 to 716.
- The report presents information (pp. 93–94) about the distribution of standard scores on the PPVT-4 and TVIP at Head Start exit and spring kindergarten for

Spanish-speaking DLLs relative to their baseline performance on the same measure. Some children (18%) did not achieve a basal on the PPVT-4 and are not included in the distribution. Overall, the distribution suggests that DLLs increase English vocabulary over time and that their Spanish vocabulary lags relative to Spanish-speaking children from the normative sample. Given that the TVIP was standardized on children from outside the continental United States and is more than 25 years old, it is difficult to interpret the meaning of these data. While they could be indicative of diminishing growth in Spanish, they may also be due to comparing DLLs in the U.S. with monolingual Spanish-speaking children receiving instruction only in Spanish from areas outside of the U.S.

Moiduddin, E., N. Aikens, L. Tarullo, & J. West. (2010). Data tables for FACES 2006: A second year in Head Start report. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This report describes the performance of children who entered the program at age 3 in their first and second year in Head Start.
- This report focuses on the 802–968 children who entered at age 3 and were assessed in English in fall 2006, spring 2007, and spring 2008. (The full sample of children who entered Head Start at 3 or 4 years old in the fall of 2006 and completed one or two years in the program and attended kindergarten the year after graduating from Head Start was 3,315 children.)
- The PreLAS language screener was used in combination with parent report of home language to determine whether children from households where English was not the primary spoken language should be assessed in English, assessed in Spanish, or not assessed except for the PPVT-4. The PPVT-4 was administered to all children in the sample. The TVIP was also administered to children who spoke Spanish at home. Children who spoke Spanish at home and made five consecutive errors on both Simon Says and Art Show were administered the Spanish assessment battery (along with the PPVT-4). Children who spoke a language other than English or Spanish that made five consecutive errors were not administered the PPVT-4. If a child had been assessed in English in one of the prior rounds, he or she was assessed in English in subsequent rounds.
- The number of children tested only in English ranged from 802 to 968 on different direct child assessments. The number of children tested in Spanish was not reported. Similar to Malone et al. (2010), the report presents information (Tables B.9 and B.10) about the distribution of standard scores on the PPVT-4 and TVIP in spring 2007 and spring 2008 for Spanish-speaking DLLs relative to their baseline performance on the same measure.

Head Start Family and Child Experiences Survey 2009 (FACES 2009)

Aikens, N., Hulsey, L. K., Moiduddin, E., Kopack, A., Takyi-Laryea, A., Tarullo, L., and West, J. (2011). Data tables for FACES 2009 Head Start children, families, and programs: Present and past data from FACES report. (OPRE Report 2011-33b). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

- FACES is a tool for measuring Head Start program performance at the national level. This recurring data collection provides the means to assess how the program is performing currently and over time. This report describes the performance of children entering Head Start in fall 2009.
- 3,349 children who were newly enrolled to Head Start at age 3 or 4 participated in fall 2009.
- Language of assessment in fall 2009 was determined by a screening made up of two subtests from the Pre-LAS 2000 (Simon Says, Art Show). All children were also administered the PPVT-4 and EOWPVT or EOWPVT-SBE. The TVIP was also administered to children who spoke Spanish at home. Children who spoke Spanish at home and made five consecutive errors on both Simon Says and Art Show were administered the Spanish assessment battery (along with the PPVT-4 and EOWPVT-SBE). Children who spoke a language other than English or Spanish and who made five consecutive errors were not administered a cognitive assessment; they were given a short vocabulary assessment (PPVT-4 and EOWPVT).
- In fall 2009, 2,605 children were assessed in English, 512 were assessed in Spanish, and 33 were given a short vocabulary assessment. In spring 2010, 2,616 children were assessed in English, 251 were assessed in Spanish, and 12 were given a short vocabulary assessment.
- This report included information about the administration of assessments to DLLs from 2000 to 2009 (Table B.25). In FACES 2000, all children were assessed in either English (84%) or Spanish (16%). In FACES 2003, DLLs who were assessed received assessments in English (87.7%) or Spanish (12.3%). In FACES 2006 and FACES 2009, 11.7% and 14.7% of the children, respectively, were assessed in Spanish for at least one assessment. Fewer than 1% (0.5% to 0.6%) of the children in FACES 2006 and FACES 2009 were DLLs with a language other than Spanish.
- The reliability of English and Spanish assessments of language and literacy in the fall FACES 2009 are listed below. The estimates for the English assessments include the DLLs.

Table A.4. Psychometric Properties of Language and Literacy Assessments in FACES 2009

Language and Literacy Assessments	Cronbach's Alpha
PPVT-4	0.97
TVIP	0.93
WJ-III: Letter-Word Identification	0.85
WJ-III: Spelling	0.79

WM-III: Letter-Word Identification	0.67
WM-III: Spelling	0.66

Mean standard scores were reported for the PPVT-4, the EOWPVT, EOSPVT-SBE, and TVIP by subgroups: all children, children with English home language, all DLLs, DLLs passing the language screener with Spanish home language, DLLs passing the language screener with another home language, DLLs failing the language screener with Spanish home language, and DLLs failing the language screener with another home language (Table B.10). Across all these measures except the TVIP, DLLs who did not pass the language screener and spoke Spanish had the lowest mean standard scores. On the TVIP, all subgroups of DLLs with a Spanish home language had similar mean scores (83.1 to 84.8).

Head Start Impact Study (HSIS)

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., et al. (2010). Head Start impact study. Technical report. Administration for Children & Families.

- The purpose of this study was to evaluate the effect of Head Start using a rigorous design that compared children who were assigned to Head Start to children who did not receive Head Start.
- A total of 4,667 newly entering children participated, including 2,559 in the 3-year-old group and 2,108 in the 4-year-old group. The participants consisted of a nationally representative sample of Head Start programs (U.S. and Puerto Rico) and of entering 3- and 4-year-old children who had not previously attended Head Start. Children were randomly assigned either to a Head Start group that had access to Head Start services in the initial year or to a control group that could receive any other non-Head Start services available in the community, chosen by their parents. In addition, all children in the 3-year-old cohort could receive Head Start services in the second year.
- To determine language of assessment, the main care provider was asked three questions regarding the child's language ability. If two or more of the responses were English or Spanish, the child was tested in that language. Children who were assessed in Spanish in fall 2002 completed the full Spanish assessment battery and two English tests (PPVT and WJ-III Letter-Word Identification). In spring 2003, these children were assessed with the complete English battery and the Spanish equivalents of vocabulary and literacy assessments (TVIP and the WM-III Letter-Word Identification, respectively) if they received services in the mainland U.S. In Puerto Rico, all children were given Spanish assessments at all data collection timepoints.
- If a child's primary language was anything other than Spanish or English, the child's teacher assessed whether he or she could understand the assessment in English. For children who could not understand, four tests (McCarthy Draw-a-Design, Color Names and Counting, Leiter-R-adapted and Story and Print concepts) were translated for them.
- 54 children were tested in a language other than Spanish or English. Information was not reported on how many children received assessments in English or Spanish.

- Reliability was reported for assessments from publisher manuals and/or prior use in FACES.

APPENDIX B

PSYCHOMETRIC EVIDENCE BY ASSESSMENT

VOCABULARY MEASURES

Communicative Development Inventories (CDI)

See MacArthur-Bates Communicative Development Inventories

Expressive One-Word Picture Vocabulary Test (EOWPVT)

Expressive One-Word Picture Vocabulary Test-Spanish Bilingual Edition (EOWPVT-SBE)

See also Receptive One-Word Picture Vocabulary Test (ROWPVT)

Description/Purpose of Assessment

This measure of expressive vocabulary for children ages 2 to 18 years has an English version (Expressive One-Word Picture Vocabulary Test [EOWPVT]; Brownell, 2000) and a conceptually scored Spanish version (Expressive One-Word Picture Vocabulary Test Spanish-Bilingual Edition [EOWPVT-SBE]; Brownell, 2001a). The EOWPVT was standardized with the same sample as the ROWPVT, and similarly, the EOWPVT-SBE was standardized with the same sample as the ROWPVT-SBE.

Evidence of Reliability

One study (Anthony et al., 2009) reported only the published reliability for the EOWPVT and EOWPVT-SBE. Aikens et al. (2011) reported acceptable reliability for both the EOWPVT and EOWPVT-SBE in FACES 2009 (= 0.86) with a sample that included DLLs.

Evidence of Validity

Rather than scoring conceptually, Anthony et al. (2009) did a Spanish-only administration of the EOWPVT-SBE in order to compare it to an English-only administration of the EOWPVT.

The composite[14] correlation of the expressive and receptive vocabulary tests in English (EOWPVT/ROWPVT) was more strongly correlated with an English literacy measure of phonological awareness (PCTOPPP = .69) than with a Spanish literacy measure of phonological awareness (SPCTOPPP = .26). The composite correlation of the expressive and receptive literacy tests in English (EOWPVT/ROWPVT) was moderately correlated with the Spanish composite (ROWPVT-SBE/EOWPVT-SBE = .23).

[14] Composite scores were calculated as the average of the standardized residuals from the two measures that were used to assess the same construct: English vocabulary at pretest, Spanish vocabulary at pretest, English phonological awareness at pretest, Spanish phonological awareness at pretest, English letter knowledge at pretest, Spanish letter knowledge at pretest, English phonological awareness at posttest, and Spanish phonological awareness at posttest. Variables used to create each composite were moderately correlated (r = .42 to .75).

The composite correlation of the expressive and receptive vocabulary tests in Spanish (EOWPVT-SBE/ROWPVT-SBE) was more weakly correlated with an English literacy measure of phonological awareness (PCTOPPP = .35) than with a Spanish measure (SPCTOPPP = .64). The composite correlation of the expressive and receptive vocabulary tests in Spanish (EOWPVT-SBE/ROWPVT-SBE) was moderately correlated with the English vocabulary composite (ROWPVT/EOWPVT = .23).

The composite English vocabulary tests (EOWPVT/ROWPVT) more strongly predicted future English phonological awareness (PCTOPPP = .52) than Spanish phonological awareness (SPCTOPPP = .38). Composite Spanish vocabulary tests (EOWPVT-SBE/ROWPVT-SBE) more weakly predicted future English phonological awareness (PCTOPPP = .46) than Spanish phonological awareness (SPCTOPPP = .52).

Macarthur-Bates Communicative Development Inventories

Macarthur-Bates Inventarios Del Desarrolo De Habilidades Comunicativas (Inventarios)

Description/Purpose of Assessment

This parent-reported measure of receptive and expressive communication for children from 8 months to 3 years old has both an English edition (The MacArthur-Bates Communicative Development Inventories [CDI]; Fenson et al., 1993) and a Spanish edition (MacArthur-Bates Inventarios del Desarrollo de Habilidades Comunicativas [Inventarios]; Jackson-Maldonado, Bates, & Thal, 2003). In some studies, teachers, rather than parents, complete the measure.

Evidence of Reliability

Seven studies used the CDI. Only Baby FACES (Vogel et al., 2011) reported study-specific reliability estimates for the full sample (= .95 to .98 for English; and = .90 to .98 for Spanish) for the CDI raw score. One study reported the internal consistency estimates reported in the manual.

Evidence of Validity

Studies with bilingual samples and dual language administration of the CDI (Parra, Hoff, & Core, 2011; Vagh, Pan, & Mancilla-Martinez, 2009) reported moderate correlations between the English and the Spanish versions (r = -.57). In addition, researchers note support for the developmental trajectory of scores and relation between language exposure and children's scores on the relevant version of the CDI. The Composite CDI/Inventario II score correlated with age (r = .57). Significant bivariate correlations were also noted for CDI and Inventario II vocabulary scores, grammatical complexity, and M3L-words with age (r = .25 to .33) and Spanish:English exposure ratio (r = .25 to .37).

Parent-reported CDI scores (but not teacher-reported scores) were more strongly correlated with children's picture vocabulary (WLPB-R; r = .46 and .52 at 30 and 36 months, respectively) than with their receptive vocabulary (PPVT-III; r = .38 and .47 at 30 and 36 months, respectively). The magnitude of the correlation between parent- and

teacher-reported scores was stronger among monolingual (English or Spanish) than among bilingual children at 30 months (r = .74 versus .48; Vagh et al., 2009).

Bivariate correlations and hierarchical regression provided evidence of validity for the English CDI with DLLs indicating a relation between percentage of language exposure in English at 22 months and English vocabulary size at 25 months (r = .72), and English grammatical complexity at 25 months (r = .58). All children were simultaneous Spanish-English-language learners (Parra et al., 2011).

Peabody Picture Vocabulary Test (PPVT)

Description/Purpose of Assessment

The Peabody Picture Vocabulary Test (PPVT-4: Dunn & Dunn, 2005) is one of the most widely used assessments of English receptive vocabulary of preschoolers. The FACES 2006 (Hulsey et al., 2010) and FACES 2009 (Aikens et al., 2011) studies used the current fourth edition, PPVT-4. Study-specific psychometric information about this measure is reported about two earlier editions, the Peabody Picture Vocabulary Test III (PPVT-III; Dunn & Dunn, 1997) and the Peabody Picture Vocabulary Test–Revised (PPVT-R; Dunn & Dunn, 1981), used in the studies of DLLs' language and literacy development as well as in prior rounds of FACES. A Spanish version of the PPVT-R, Test de Vocabulario en Imágenes Peabody (TVIP; Dunn, Lugo, Padilla, & Dunn, 1986), has not been updated alongside subsequent versions of the PPVT. It differs greatly from the PPVT and thus is a very different measure. It is discussed separately.

Evidence of Reliability

PPVT-4

Reports from FACES 2006 and FACES 2009 (Aikens et al., 2011; Hulsey et al., 2010; West et al. 2008) included study-specific reliability for the PPVT-4 for the full sample, including DLLs (= .94–.97).

PPVT-III

Seven studies (Bialystok, Luk, Peets, & Yang, 2010; Davidson, Raschke, & Pervez, 2010; Dickinson, McCabe, Clark-Chiarelli, & Wolf, 2004; Hammer, Lawrence, & Miccio, 2007; Hammer, Lawrence, & Miccio, 2008a; Marinova-Todd, Zhao, & Bernhardt, 2010; Vagh et al., 2009) reported only the published reliability for the PPVT-III.

One study (Hammer, Davison, Lawrence, & Miccio, 2009) reported stability of PPVT-III scores over the three-year period ranging from .48 to .75.

When reported, strong internal consistency reliability (= .96 to .97) typically is noted for the PPVT-III with a Head Start sample that included children who are DLLs (Administration for Children and Families, 2006).

One study (Puma et al., 2010) reported weaker reliability, ranging from .61 to .80 over a four-year period.

PPVT-R

One study (Farver, Xu, Eppe, & Lonigan, 2006) reported only the published reliability measures.

PPVT-R (Mandarin)

One study (Marinova-Todd et al., 2010) used a Mandarin version of the PPVT-R but did not provide any evidence of reliability.

Evidence of Validity

PPVT-4

There is no additional evidence of validity for the PPVT-4 available in these reports.

PPVT-III

Four studies (Bialystok et al., 2010; Hammer et al., 2007; Hammer et al., 2008a; Vagh et al., 2009) reported monolingual English speakers scored higher than DLLs, and these differences maintained over time. Monolingual English speakers outperformed DLLs at every age comparison. Overall, monolingual English speakers averaged standard scores of 107 (SD = 12.3) and DLLs averaged 96 (SD = 13.0) (Bialystok et al., 2010). One study (Hammer et al., 2009) shows moderate correlation with use of English in the home within time (r = .28 to .38). This measure also positively correlated with use of English in the home across assessments (r = .28 to .41).

Bialystok et al. (2010) conducted item analysis of the words in item sets 1–10 for their Canadian sample. These items were categorized based on relevance to home and school context (inter-rater agreement = 91.7%, kappa = 0.73) and analysis of subset scores examined for 161 children ages 6 years to 6 years, 11 months. Mean percentage correct in home versus school context category was calculated for each child, and proportions were examined by monolingual and bilingual language groups. Monolingual English speakers scored higher than DLLs in the home category only; both groups scored the same for the school-only category.

One study (Dickinson et al., 2004) found this measure of receptive vocabulary weakly correlated with the Spanish language version of the TVVP-R, an earlier version of this measure (TVIP = .23). This measure was moderately correlated with English emergent literacy (ELP = .30). They also found a stronger correlation with the English literacy measure (EPAP = .48) than with the Spanish measure (Spanish EPAP = .32). The PPVT-III measure of receptive vocabulary predicted future English phonological awareness (EPAP = .29).

In one study (Hammer et al., 2007), principal components analysis was used to derive a criterion measure, which was combined with a language measure (PPVT-

III/TELD-3), to create a measure of overall receptive language ability in English[15] (full sample only). They reported mean scores by language group.

One study (Vagh et al., 2009) reported means by language group at 30 and 36 months. As expected, this measure is positively correlated with a parent-reported measure of vocabulary scores (CDI) at 30 months and 36 months (r = .38 to .47). Parent-teacher composite vocabulary measure scores (CDI) were significantly associated with scores on this vocabulary measure (r = .34) at 36 months.

PPVT-R

This measure of receptive vocabulary was moderately correlated with literacy (Title Recognition Task (TRT) = .24). As expected, there are weak to moderate correlations between this measure of receptive vocabulary and measures of literacy exposure, including parents' literacy habits subscale scores (HLEQ = .26), children's literacy interest subscale scores (r = .38), and parents' literacy involvement subscale scores (r = .25). Also as anticipated, this measure is moderately correlated with child age (r = .26) and maternal and paternal education (r = .31 and .28, respectively). As expected, this measure of receptive vocabulary is correlated with the socio-emotional measures: the Social Functioning subscale of the Behavior Assessment System for Children (BASC), = .89, r = .34, and the Parental Stress Index, (PSI), = .89, r = -.28 (Farver et al., 2006).

In the FACES 1997 sample, the spring Head Start PPVT-III score correlated with end of kindergarten ECLS-K General Knowledge scale score (r = .77; beta = .62) and with the ECLS-K Reading scale score (r =.42; ns for the multiple regression) (Zill et al. 2003). The sample was limited to children who took the assessments in English at all timepoints and so did not include many DLLs; however, these estimates provide a point of comparison for validity estimates for low-income populations on these measures.

PPVT-R (Mandarin)

In one study (Marinova-Todd et al., 2010), researchers used the Mandarin version of the PPVT-R (Lu & Liu, 1998). The mean scores were reported by age and language group. Scores of Mandarin monolinguals were similar at 5 years, but higher than the scores of bilinguals only for the 6-year-old group.

Receptive One-Word Picture Vocabulary Test (ROWPVT)

Receptive One-Word Picture Vocabulary Test–Spanish Bilingual Edition (ROWPVT-SBE)

See also Expressive One-Word Picture Vocabulary Test (EOWPVT).

[15] The assessments used in the study do not provide normative information for bilingual children. The PPVT–III, TELD–3, and TERA–2 were developed for monolingual English-speaking populations; the TVIP contains normative data for monolingual Puerto Rican children; and the PLS–3 was based on a monolingual Spanish-speaking sample.

Description/Purpose of Assessment

This measure of receptive vocabulary has both an English edition (Receptive One-Word Picture Vocabulary Test [ROWPVT]; Brownell, 2000) and a conceptually scored Spanish bilingual edition (Receptive One-Word Picture Vocabulary Test: Spanish Bilingual Edition [ROWPVT-SBE]; Brownell, 2001b). The EOWPVT was standardized with the same sample as the ROWPVT; similarly, the EOWPVT-SBE was standardized with the same sample as the ROWPVT-SBE.

Evidence of Reliability

One study (Anthony et al., 2009) reported only the published reliability for the ROWPVT and ROWPVT-SBE.

Evidence of Validity

Rather than scoring conceptually, Anthony et al. (2009) used a Spanish-only administration of the ROWPVT-SBE in order to compare it to an English-only administration of the ROWPVT.

The composite[16] correlation of the expressive and receptive vocabulary tests in English (EOWPVT/ROWPVT) was more strongly correlated with an English literacy measure of phonological awareness (PCTOPPP = .69) than with a Spanish literacy measure of phonological awareness (SPCTOPPP = .26). The composite correlation of the expressive and receptive literacy tests in English (EOWPVT/ROWPVT) was moderately correlated with the Spanish composite (ROWPVT-SBE/EOWPVT-SBE = .23).

The composite correlation of the expressive and receptive vocabulary tests in Spanish (EOWPVT-SBE/ROWPVT-SBE) was more weakly correlated with an English literacy measure of phonological awareness (PCTOPPP = .35) than with a Spanish measure (SPCTOPPP = .64). The composite correlation of the expressive and receptive vocabulary tests in Spanish (EOWPVT-SBE/ROWPVT-SBE) was moderately correlated with the English vocabulary composite (ROWPVT/EOWPVT = .23).

The composite English vocabulary tests (EOWPVT/ROWPVT) more strongly predicted future English phonological awareness (PCTOPPP = .52) than Spanish phonological awareness (SPCTOPPP = .38). Composite Spanish vocabulary tests (EOWPVT-SBE/ROWPVT-SBE) more weakly predicted future English phonological awareness (PCTOPPP = .46) than Spanish phonological awareness (SPCTOPPP = .52).

[16] Composite scores were calculated as the average of the standardized residuals from the two measures that were used to assess the same construct: English vocabulary at pretest, Spanish vocabulary at pretest, English phonological awareness at pretest, Spanish phonological awareness at pretest, English letter knowledge at pretest, Spanish letter knowledge at pretest, English phonological awareness at posttest, and Spanish phonological awareness at posttest. Variables used to create each composite were moderately correlated (r = .42 to .75).

Spanish-English Vocabulary Checklist (SEVC)

Description/Purpose of Assessment

The Spanish-English Vocabulary Checklist (SEVC; Patterson, 1998) is a parent-reported measure and is an adaptation of Rescorla's (1989) Language Development Survey. It consists of 564 items, half in English and half in Spanish. English and Spanish directions are offered, as well as the option for the instructions to be read aloud to parents in the language of their choice or in both languages.

Evidence of Reliability

One study (Patterson, 2002) reported only the previously published test-retest reliability for the SEVC (Patterson, 1998).

Evidence of Validity

Patterson, 2000 noted the number of words reported by parents on the SEVC was significantly correlated with the number of different words children used in language samples with the reporting parent (r = .66). When examined by language, the number of Spanish words reported by parents on the SEVC was strongly correlated with the number of different Spanish words children use in language samples with the reporting parent (r = .92).

Child age and reading frequency uniquely contributed to vocabulary size in both English and Spanish. Similar relations were found for amount of input by language: English input was correlated with English vocabulary (r = .33) and Spanish input was correlated with Spanish vocabulary (r = .32). Reading in the same language had slightly stronger correlations with reported vocabulary on the SEVC (English with English reading = .40; Spanish with Spanish reading = .35) (Patterson, 2002).

Test De Vocabulario En Imágenes Peabody (TVIP)

Description/Purpose of Assessment

The Test de Vocabulario en Imágenes Peabody (TVIP; Dunn, Padilla, Lugo, & Dunn, 1986) is a measure of Spanish receptive vocabulary. It was originally designed to complement the English PPVT. Unlike the PPVT, the TVIP has not been updated. The standardization sample for the TVIP is drawn from samples of children in Mexico and Puerto Rico combined (Dunn et al., 1986).

Evidence of Reliability

Five studies (Dickinson et al., 2004; Farver et al., 2006; Hammer et al., 2009; Hammer et al., 2007; Hammer et al., 2008a) reported only the published reliability for the TVIP.

Acceptable reliability is reported for study samples in FACES 2000, 2006, and 2009 (ranges from .92 to .94) (Administration for Children and Families, 2006; Aikens et al. 2011; Hulsey et al. 2010; West et al. 2008).

One study (Puma et al., 2010) reported only published reliability estimates.

One study (Hammer et al., 2009) reported a stability range of TVIP scores over the three-year period from .50 to .71.

Evidence of Validity

Two studies (Hammer et al., 2008a; Hammer et al., 2007) found children who spoke English only at school had significantly higher scores on this Spanish vocabulary measure than children who spoke English at home and at school.

As expected, one study (Hammer et al., 2009) found this Spanish vocabulary measure is negatively correlated with use of English in the home within time (r = -.30 to -.40). Similarly, this study also presented moderate evidence of predictive validity, reporting a negative correlation between this measure and the use of English in the home during children's second year of Head Start (r = -.25 to -.36) and the spring of kindergarten (r = -.40 to -.45).

One study (Hammer et al., 2007) used Principal Components Analysis to derive criterion measures, by combining this measure with a language measure (TVIP/PLS-3). This measure was used as overall receptive language ability in Spanish[17] (full sample only).

This measure of receptive vocabulary was moderately correlated with literacy (Title Recognition Task (TRT) = .24). As expected, there are weak to moderate correlations between this measure of receptive vocabulary and measures of literacy exposure, including parents' literacy habits subscale scores (HLEQ = .26), children's literacy interest subscale scores (r = .38), and parents' literacy involvement subscale scores (r = .25). Also as anticipated, this measure is moderately correlated with child age (r = .26) and maternal and paternal education (r = .31 and .28, respectively). As expected, this measure of receptive vocabulary is correlated with the socio-emotional measures (Social Functioning subscale of the Behavior Assessment System for Children, BASC (= .89, r = .34) and inversely correlated with parental stress (Parental Stress Index, PSI (= .89, r = -.28) (Farver et al., 2006).

[17] The assessments used in the study do not provide normative information for bilingual children. The PPVT–III, TELD–3, and TERA–2 were developed for monolingual English-speaking populations; the TVIP contains normative data for monolingual Puerto Rican children; and the PLS–3 was based on a monolingual Spanish-speaking sample.

LANGUAGE MEASURES

Ages & Stages Questionnaires: Communication Subtest (ASQ-3)

Description/Purpose of Assessment

The Ages & Stages Questionnaires (Third Edition) (ASQ-3; Squires, Twombly, Bricker, & Potter, 2009) is a parent-report measure of children's development used to screen for developmental difficulties in communication, problem solving, personal-social, and gross and fine motor skills. This section will discuss information specific to the communication subtest. Baby FACES included questions from two or more of the age-appropriate questionnaires included in the ASQ-3 (a total of 21 questionnaires are available for ages 1 month to 5.5 years).

Evidence of Reliability

Baby FACES (Vogel et al., 2011) included information from the publishers as well as study-specific information about reliability. The internal consistency of the communication items for the full sample (including DLLs) ranged from .65 to .73, although the total ASQ-3 (with more items) had stronger reliability (=.78 to .84).

Evidence of Validity

No studies presented any additional evidence of validity beyond the publisher information.

Bayley Short Form Research Edition (BSF-R)

Description/Purpose of Assessment

The Bayley Short Form–Research Editions is a shortened version of the Bayley Scales of Infant Development–II (BSID-II; Bayley, 1993) developed for use in ECLS-B. The BSF-R includes a mental scale and a motor scale as well as measures of communication and expressive and receptive vocabulary. A Spanish version of this measure was developed.

Proficiency scores were developed from the mental scale based on the hierarchical nature of the data. Across the nine-month to two-year data collection periods, six of the proficiency scores are relevant language outcomes.

Evidence of Reliability

One study (Flanagan & West, 2004) reported strong reliability for the mental scale (= .79).

Evidence of Validity

- Expected increases in skill by age were evident in the data. Across the 22-months to 25-months age groups, the percentage of children demonstrating proficiency increased for all the language proficiency scores except "jabbers expressively," and item difficulty varied in expected ways. All children demonstrated proficiency on "jabbers expressively" and more children

demonstrated proficiency in receptive vocabulary than in expressive vocabulary and listening comprehension.

Bilingual English Spanish Assessment (BESA)

Description/Purpose of Assessment

The Bilingual English Spanish Assessment (BESA; Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, in preparation) was designed as a tool for identifying typical and atypical language development among DLLs. The BESA allows scores in English, in Spanish, and conceptual scores. Two subtests on the BESA include multiple tasks:

1. *Semantics subtest* (48 items in three difficulty levels) assesses children's understanding and use of words and concepts in relation to a story. The semantic subtests include both Spanish and English items about characteristic properties, functions (e.g., what do you do with a spoon), analogies, linguistic concepts (e.g., spatial and time concepts such as "under the table," "before you go outside"), similarities and differences, and comprehension of passage.

2. *Morphosyntax subtest* (51 items) assesses children's understanding and use of the parts of words that change to convey meaning (morphemes such as the use of –s for regular plurals and –ed for regular past tense) and the grammar of a language (syntax). The BESA includes a Sentence Imitation task (12 English sentences and 11 Spanish sentences that increase in difficulty and length) as well as a Morphosyntax Cloze task in both Spanish (including assessment of articles, direct object clitic, preterit, and subjunctives) and English (to assess children's understanding of plural nouns, possessives, third-person singular, regular and irregular past, present and past auxiliaries plus progressive, passives, and auxiliaries plus negation) to assess the salient features of each language.

Evidence of Reliability

Bedore and colleagues (2005) reported inter-rater reliability for the BESA ranging from 95 to 99 percent agreement. No estimates of internal consistency were presented, though analysis of correlations between the English and Spanish items on the subtests were moderately correlated even nine months later (r ranged from .41 to .69), suggesting that the English and Spanish items on a subtest were measuring the same construct (Castilla, Restrepo, & Perez-Leroux, 2009). However, the correlation was sometimes stronger between subtests in different areas across language. For example, the English Semantic test was more strongly correlated with the Spanish morphosyntax sentence repetition task (r = .69), than with the Spanish Semantic task (r = .69) (Castilla et al., 2009).

Evidence of Validity

Semantic subtest

Characteristic property items. The authors reported that the English and Spanish characteristic properties items targeted the same concept (e.g., size, shape, function, color), but different questions were used for concepts in each language (that is, items were not direct translations). The selection of vocabulary and themes was based on a

literature review of language development and cultural relevance in each language. Vocabulary that children would be exposed to at home was used for the Spanish version, while the English version included vocabulary that children would be more likely to learn at school. An examination of performance on these characteristic property items using repeated measures ANOVA indicated no significant score differences for test language (Spanish or English) or language group (monolingual or bilingual), suggesting that each language included items of similar difficulty levels. Scores were calculated for the correct responses to each item: monolingual score in English or Spanish, total response score, and conceptual score. Mean scores were reported by language group (primarily English [PE], bilingual English [BE], bilingual Spanish [BS], primarily Spanish [PS]) with no group differences found on Spanish scores. The mean English scores for children in the bilingual Spanish group were significantly lower than those for the primarily English-speaking group (Bedore et al., 2005).

Semantic subtest. According to the authors, the Semantic subtest items for English and Spanish versions were selected based on difficulty levels and discrimination values from a larger set of items. Item difficulty levels ranged from 0.30 to 0.80 for English and Spanish (mean item difficulty level for English items = 0.61, for Spanish items = 0.56) (Bedore et al., 2005).

In one study, the difference in the monolingual scores between children in the BS and BE groups was greater on the English subtest than on the Spanish subtest (Bedore et al., 2005). This is interesting because the semantic subtest includes many academic concepts and so you might expect that children in both groups would know more of these in English than in Spanish.

With a sample of children who reportedly were typically developing, the authors conducted classification analysis to examine the role of conceptual scoring in classifying children from different language dominance groups. Using the primarily English and primarily Spanish groups to set the cut-off scores, classification as typically developing was good for BE children using both the English monolingual and English conceptual scores (correctly identifying 100% as typically developing in English). For the children in the BS group, the Spanish monolingual score resulted in correctly identifying only 70% of the sample as typically developing and the conceptual score resulted in a change from poor to fair classification (correctly identifying 80%). In both languages, there were fewer typically developing bilingual children with scores greater than -1.5 SD below the mean when the conceptual score was used than when the monolingual score was used (this difference was more pronounced for the BE children on the Spanish scores). For the Spanish subtest items, with conceptual scoring, classification accuracy increased from 50% to 80% for the BE group and from 70% to 80% for the BS group, suggesting that the conceptual scores will be more valid indicators of disability than the monolingual scores (Bedore et al., 2005).

None of the BESA subtests was significantly related to the child's mean length of utterance (MLU) in Spanish or to the Spanish diversity of words (D measure) (Castilla et al., 2009). With a sample size of only 49 DLLs, power to detect relations was limited. All correlations of the BESA subtests with MLU and the D measure were less than .10. Nonsignificant correlations emerged between English and Spanish sentence repetition and Spanish MLU and Spanish D (r ranged from .19 to .30), as well as between the Spanish Semantic test and Spanish D (r = .23).

The Spanish BESA was used in a descriptive study in the fall of prekindergarten to predict spring English-BESA (Castilla et al., 2009). As noted above, the Spanish and English subtests were moderately to strongly correlated even with nine months between administrations. The strongest correlation was found between the fall Spanish-MT Sentence Repetition scores and the spring English-MT Cloze Test scores = .69; and English-ST scores = .69.

Bilingual English Spanish Oral Language Screener (BESOS)

See also the Bilingual English Spanish Assessment (BESA; Peña et al., in preparation)

Description/Purpose of Assessment

The Bilingual English Spanish Oral Language Screener (BESOS; Peña et al., in preparation) comprises a semantic subtest as well as a morphosyntax[18] subtest in both Spanish and English. The BESOS developed from the experimental item pool of the Bilingual English Spanish Assessment (BESA; Peña et al., in preparation).

Evidence of Reliability

- Semantics subtest: Test-retest (n = 20): r = .70 for Spanish semantics, r = .64 for English semantics (Bohman, Bedore, Peña, Mendez-Perez, & Gillam, 2010).
- Morphosyntax subtest: Test-retest (n = 20): r = .86 for Spanish morphosyntax, r = .75 for English morphosyntax (Bohman et al., 2010).

Evidence of Validity

Correlations of the BESOS subtests with the full set of items on the parent BESA were examined using the data from the normative sample and indicated that the BESOS subtests measure the same construct as the BESA subtests (r = .85 for Spanish Semantics, r = .89 for English semantics; r = .83; for Spanish morphosyntax, r = .89; for English morphosyntax) (Bohman et al., 2010).

With a sample of 757 pre-kindergarten and kindergarten-age children, Bohman and colleagues (2010) used zero-inflated Poisson (ZIP) and zero-inflated binomial (ZINB) logistic regression analyses to examine factors associated with scores on the BESOS Spanish and English morphosyntax and semantic outcomes. The age and percentage of language output was significant for both Spanish and English and for both morphosyntax and semantics, but language input was only significant for Spanish outcomes. Maternal (but not paternal) education was related to English semantics and morphosyntax. Gender was significant in predicting Spanish, but not English, morphosyntax. Free lunch status significantly predicted Spanish morphosyntax and semantics as well as English semantics.

Spanish Preschool Language Scale 3 (PLS-3)

Description/Purpose of Assessment

The Preschool Language Scale 3 (PLS-3; Zimmerman, Steiner, & Pons, 1992) can be utilized to measure children's (birth to 6 years) language skills, specifically receptive vocabulary, comprehension of concepts, and understanding of grammatical structures in both English and Spanish. The auditory component of the subtest comprises 48 items. Hammer et al. (2007, 2008a, 2008b) used the Spanish PLS-3 to test children's receptive

[18] The glossary in Appendix E provides definitions.

language abilities in Spanish. There is a recently released fourth edition of this Spanish measure (Spanish PLS-4) that includes standard scores based on conceptual scoring.

Evidence of Reliability

Three studies (Hammer et al., 2007, 2008a, 2008b) reported only the published reliability estimates for the Spanish PLS-3.

Evidence of Validity

Spanish PLS-3 Auditory Comprehension Scale

Principal Components Analysis was used to derive criterion measures; the combined TVIP/PLS-3 for measure of overall receptive language ability in Spanish (full sample only) and the first component captured an average of 95% of the variance across four measurement occasions (range of 93% to 97%), suggesting that performance on TVIP was well correlated with performance on the Spanish PLS-3 (Hammer et al., 2007).

Mean scores were higher for children in the school English communication (SEC) group than in the home English communication (HEC) group. These are children who were not exposed to English until they began Head Start (Hammer et al., 2007, 2008a, 2008b). Child age was positively associated with scores (Hammer et al., 2008a); a random intercept model indicated children's scores increased across the four timepoints in one study (Hammer et al., 2008a), while in a subsequent study the mean standard scores of children in the HEC and SEC groups increased over the first three timepoints, followed by a decline (Hammer et al., 2008b).The variance in the scores for the children in the SEC group in the latter study was smaller than for the HEC group.

Test of Early Language Development–3 (TELD-3)

Description/Purpose of Assessment

The Test of Early Language Development-3 (TELD-3; Hresko, Reid, & Hammill, 1999) is used to assess children ages 2 to 7. The full TELD-3 includes an overall spoken language score as well as receptive and expressive language subtests. The receptive language subtest of the TELD-3 was utilized in Hammer et al. (2007, 2008a, 2008b) to examine children's understanding of language.

Evidence of Reliability

Strong internal consistency was found for the TELD-3 (r = .90 to .95) (Hammer et al., 2007, 2008a, 2008b).

Evidence of Validity

Principal Components Analysis was used to derive criterion measure; combined PPVT-III/TELD-3 Receptive Language subtest for measure of overall receptive language ability in English (full sample only) (Hammer et al., 2007). The first component captured an average of 97.5% of the variance in responses with a range from 96% to 99% across four measurement occasions.

Across multiple timepoints, mean scores reported by language group were higher among children in the HEC group than for children in the SEC who were not introduced to English until they began Head Start (Hammer et al., 2007, 2008a, 2008b). Child age and passage of time was positively associated with scores (Hammer et al., 2008a, 2008b).

Woodcock Language Proficiency Battery–Revised (WLPB-R)

Woodcock Language Proficiency Battery–Revised, Spanish Form

Description/Purpose of Assessment

This measure of overall language proficiency has both an English (Woodcock Language Proficiency Battery–Revised, English Form [WLPB-R]; Woodcock, 1995) and a Spanish edition (WLPB–R, Spanish Form; Woodcock & Muñoz-Sandoval, 1995). It includes 13 assessments divided into oral language, reading, and written language subtests.

Evidence of Reliability

Five studies (Hammer et al., 2007; Páez, Tabors, & Lopez, 2007; Rinaldi & Páez, 2008; Tabors, Páez, & López, 2003; Vagh et al., 2009) reported only the published reliability for WLPB-R. Hammer et al., 2007 reported strong internal consistency (r = .96) for the Letter-Word Identification.

Evidence of Validity

Overall

One study (Vagh et al., 2009) reported moderate correlation with a parent-reported vocabulary measure (CDI; r = .46 to .52) at 30 months and 36 months. Parent-teacher composite vocabulary measure (CDI) scores were also significantly associated with this language measure (r = .38) at 36 months.

Picture Vocabulary (English Edition)

One study (Rinaldi & Páez, 2008) reported that preschool English Vocabulary predicted English Letter-Word Identification scores in first grade. As expected, one study (Vagh et al., 2009) found bilingual children had lower scores than monolingual English children.

Two studies (Páez et al., 2007; Rinaldi & Páez, 2008) reported nonsignificant growth in scores over time (Spanish-English bilinguals [ECS]: fall = 68.1 [SD = 19.2]; spring = 70.5 [SD = 18.5]; Time 1 to Time 3: PreK = 70.3 [SD = 18.8], K = 72.3 [SD = 19.6], 1st = 79.4 [SD = 19.5]) suggesting that studies may need to be powered properly to detect the level of change that is observed across this time period with this assessment. Further examination of the sensitivity of this measure among bilingual children may be warranted.

Picture Vocabulary, Vocabulario Sobre Dibujos (Spanish Edition)

One study (Tabors et al., 2003) reported a negative correlated with English Picture Vocabulary scores (r = -.28).

Two studies (Páez et al., 2007; Tabors et al., 2003) reported Spanish-speaking monolinguals scored significantly higher than Spanish-English bilinguals (as would be expected), and this change did not decrease over time (bilinguals: fall = 65.2, spring = 62; monolinguals: fall = 84, spring = 86.9)

Memory for Sentences (English Edition)

One study (Tabors et al., 2003) reported a low correlation with Spanish Memory for Sentences scores (r = .25). Another study (Rinaldi & Páez, 2008) reported that preschool English Memory for Sentences predicted English Letter-Word Identification scores in first grade.

Two studies (Páez et al., 2007; Rinaldi & Páez, 2008) reported that scores did not improve significantly over time for DLLs (fall = 73.1, spring = 77.2 [Páez et al., 2007]; PreK = 78.0, K = 77.3, 1st = 85.1 [Rinaldi & Páez, 2008]).

Memory for Sentences, Memoria para Frases (Spanish Edition)

One study (Tabors et al., 2003) reported a low correlation with English Memory for Sentences scores (r = .25). This is not unexpected since memory for sentences is a measure of children's understanding of syntax in a language. Because English and Spanish have different syntax, the strength of the relation will depend on proficiency in both languages. Two studies (Páez et al., 2007; Tabors et al., 2003) noted that Spanish-English bilinguals (with potentially less frequent exposure to Spanish) scored lower than Spanish-speaking monolinguals; that difference remained over time (bilinguals = 70.1 to 72; monolinguals = 83.8 to 88.6).

Letter-Word Identification (English Edition)

One study (Tabors et al., 2003) reported moderate correlation with the Spanish Letter-Word Identification scores (r = .51).

One study (Hammer et al., 2007) reported nonsignificant correlations with English vocabulary and language measure (PPVT-III/TELD-3 component scores) and a moderately negative correlation with Spanish vocabulary and language measure (TVIP/PLS-3 component score, r = -.33 to -.46), as expected. Additionally, they found moderate evidence of validity, reporting this English language measure positively correlated with a measure of English literacy (TERA-2, r = .32 to .72).

One study (Páez et al., 2007) found that scores did not improve significantly between fall and spring for Spanish-English bilinguals (fall = 90.8, spring = 91.1). Another study (Rinaldi & Páez, 2008) found evidence of validity over time, with standard scores improving as the child progressed from pre-kindergarten to kindergarten to first grade, as expected (PreK = 91.2, K = 96.8, 1st = 105.6). This study further reported that preschool English Vocabulary and Memory for Sentences predicted Letter-Word Identification scores in first grade. In a separate model, preschool English Vocabulary, Spanish Vocabulary, English Memory for Sentences, and Spanish Letter-Word Identification predicted English Letter-Word Identification skills in first grade.

Letter-Word Identification, Identificación de Letras y Palabras (Spanish Edition)

One study (Tabors et al., 2003) reported a moderate correlation with the English Letter-Word Identification subtest (r = .51). Another study (Rinaldi & Páez, 2008) reported preschool Spanish Letter-Word Identification predicted English Letter-Word Identification skills in first grade.

Tabors et al. (2003) found no significant difference for scores between Spanish-English bilinguals and Spanish-speaking monolinguals (bilinguals = 88.8; monolinguals = 87.9).

Dictation/Spelling (English Edition)

One study (Páez et al., 2007) found that scores did not improve significantly between fall and spring for Spanish-English bilinguals (fall = 88.7, spring = 91.6).

One study (Tabors et al., 2003) found moderate evidence of validity, reporting this measure of English dictation is correlated with Spanish Dictation scores (r = .50)

Dictation/Spelling, Dictado/Ortografía (Spanish Edition)

One study (Tabors et al., 2003) reported a moderate correlation between this subtest of Spanish dictation and the scores on the English Dictation subtest (r = .50), offering support for the validity of the measure. This study further found no significant difference in standard scores between Spanish-English bilinguals and Spanish-speaking monolinguals (bilinguals = 90.3; monolinguals = 86.7). A later study (Páez et al., 2007) reported also finding no significant differences (bilinguals = 90.4; monolinguals = 86.7).

LITERACY MEASURES

Early Phonological Awareness Profile (EPAP)

Description/Purpose of Assessment

The Early Phonological Awareness Profile (EPAP; Dickinson & Chaney, 1997) measures phonological awareness in English and Spanish.[19] It consists of two tasks: deletion detection and rhyme recognition.

Evidence of Reliability

One study (Dickinson et al., 2004) reported strong evidence of reliability for EPAP with a diverse sample (=.93). Evidence was strong with a monolingual sample for English (=.94) and Spanish (=.93) as well. They found the Spanish version more strongly predicted future phonological awareness (=.42) than the English version did within the same language (=.35). Similarly, the Spanish measure more strongly predicted future English phonological awareness (=.41) than the English measure predicted future Spanish phonological awareness (=.32). Correlation between fall English and Spanish EPAP = .60; between spring English and Spanish EPAP = .78.

Evidence of Validity

As expected, one study reported the Spanish version of this phonological awareness measure was moderately correlated with age (EPAP = .25). The English version of this phonological awareness measure moderately predicted future English vocabulary (PPVT = .29). The Spanish version more strongly predicted future literacy (ELP = .42) than did the English measure (ELP = .32) (Dickinson et al., 2004).

Emergent Literacy Profile (ELP)

Description/Purpose of Assessment

The Emergent Literacy Profile (Emergent Literacy Profile (ELP); Dickinson & Chaney, 1997) is used to measure children's emergent literacy and comprises four subtasks. Dickenson et al. (2004) created a Spanish version of this assessment as well.

Evidence of Reliability

One study (Dickinson et al., 2004) reported acceptable reliability (= 0.82) for a large sample of English-speaking children.

Evidence of Validity

As expected, given the ELP's print-based nature, language versions are extremely similar; pilot work indicated that children's performance when assessed in English was

[19] Spanish version of the instrument is not a direct translation, but retains the same conceptual structure. Authors make no a priori claims that the two instruments are of equivalent difficulty.

closely related to performance when assessed in Spanish. One study (Dickinson et al., 2004) reported evidence of convergent validity, showing correlations with Spanish and English literacy measures (Spanish EPAP = .48; English EPAP = .48). This study also presents moderate evidence of discriminate validity, reporting ELP measure correlation with future Spanish literacy (EPAP = .42) and English literacy (EPAP = .32), as expected. They further showed moderate correlations between this literacy measure and measures of English vocabulary (PPVT-III = .30) and Spanish vocabulary (TVIP = .33).

Preschool Comprehensive Test Of Phonological And Print Processing (PCTOPPP)

Description/Purpose of Assessment

This measure of phonological awareness and print concepts has both an English edition (Preschool Comprehensive Test of Phonological and Print Processing [PCTOPPP]; Lonigan et al., 2002) and a Spanish edition (Spanish Preschool Comprehensive Test of Phonological and Print Processing [SPCTOPPP]; Lonigan & Farver, 2002).

Evidence of Reliability

One study (Anthony et al., 2009) reported only the published reliability for PCTOPPP and SPCTOPPP. It reported a higher stability for the English version (PCTOPPP = .70) than for the Spanish version (SPCTOPPP = .50).

Evidence of Validity

One study (Anthony et al., 2009) reported evidence of convergent validity between the elision and blending subtests of this literacy measure and their corresponding versions in the other language (SPCTOPPP = .43; PCTOPPP = .43). As expected, this study also reported a stronger correlation between this English measure and a measure of expressive and receptive English vocabulary (ROWPVT/EOWPVT = .69) than with expressive and receptive Spanish vocabulary (ROWPVT-SBE/EOWPVT-SBE = .35). Similarly, this study also reported a stronger correlation between this Spanish measure and a measure of expressive and receptive Spanish vocabulary (ROWPVT-SBE/EOWPVT-SBE = .64) than with expressive and receptive English vocabulary (ROWPVT /EOWPVT = .26). Anthony and colleagues (2009) reported that in earlier studies the Elision and Blending subtests demonstrated good convergent and discriminant validity (Anthony et al., 2006; Anthony et al., 2007).

Pre-Language Assessment Scale (Pre-LAS)

Description/Purpose of Assessment

The Pre-Language Assessment Scale (Pre-LAS; DeAvila & Duncan, 2000) is a measure of language proficiency for children ages 4 to 6, administered in both an English and a Spanish version. This measure includes subtests of receptive comprehension (Simon Says), expressive vocabulary (Art Show), and literacy and language skills (Let's Tell Stories), which combine to form a measure of emergent literacy.

Evidence of Reliability

The ECLS-B study (Najarian et al., 2010) reported reliability range from .84 to .93 for early reading assessment.

Two studies (Gonzalez & Uhing, 2008; Lopez & Greenfield, 2004) reported only the published reliability for PreLAS 2000.

Evidence of Validity

One study (Lopez & Greenfield, 2004) reported subtests were all significantly correlated within language, but the correlation between the PreLAS English and Spanish versions was not significant in that sample. This study reported stronger correlations between the English PreLAS and a researcher-developed English Phonological Sensitivity test (r = .52) than between the Spanish versions of those measures (r = .33).

One study (Gonzalez & Uhing, 2008) reported the expected correlation between the Spanish version of this measure and family support, a measure of the levels of interaction within a family that includes grandparents and other relatives, created by the authors (r = .32). This study also reported the expected correlation between the English version of this measure and a measure of library use created by the authors (r = .39).

Story & Print Concepts

Description/Purpose of Assessment

This literacy measure developed for use in FACES 2000 is composed of several subtasks centered around the assessor reading a story book. Instructions were also translated into Spanish by the FACES Research Team (Administration for Children and Families [ACF], 2006).

Evidence of Reliability

FACES 2000 (ACF, 2006) reported reliability for three time periods: fall 2000, spring 2001, and spring 2002. The reported reliability was low for the English versions and very low for the Spanish version. The reliability for the English subtests of book knowledge and print conversations increased slightly over time (book knowledge = .57 to .61; print conversations = .73 to .85). The English comprehension subtest decreased slightly over time (comprehension = .43 to .40). The reliability estimates were similar for the Spanish version for the first time period, fall 2000 (book knowledge = .43; print conversations = .59; comprehension = .39).

For FACES 2006, researchers (Hulsey et al. 2010; West et al. 2008) reported moderate reliability (= 0.70) overall for the English version, but lower reliability for the Spanish.

Evidence of Validity

Teacher educational attainment showed a weak, though significant, correlation with spring children's scores on Story and Print Concepts in FACES 1997 (Zill et al. 2001); however, this was examined only for children who were assessed in English at every timepoint.

In the FACES 1997 sample, the spring Head Start scores on Book Knowledge correlated with end of kindergarten ECLS-K General Knowledge (r = .40), but was not a significant predictor in the multiple regression analysis (Zill et al., 2003)

Test of Early Reading Ability–2 (TERA-2)

Description/Purpose of Assessment

This measure of emergent literacy ability (Test of Early Reading Ability-2 (TERA-2); Reid, Hresko, & Hammill, 1991) provides insight into children's knowledge of the alphabet, print conventions, and ability to construct meaning from print. It was standardized with English-speaking children ages 3.5 to 8.5 years. There is a third edition of this measure (TERA-3) now available.

Evidence of Reliability

Two studies (Hammer et al., 2007, Hammer, Miccio, & Wagstaff, 2003) reported only the published reliability for TERA-2. One study (Hammer et al., 2009) reported the range of stability from .34 to .53 for the TERA-2 scores over the three-year period.

Evidence of Validity

Hammer et al. (2003) reported that mean scores did not significantly differ between children who were introduced to English at home and those who did not learn English until starting Head Start, though the scores of the latter group were approximately 5.3 standard score points lower.

Hammer et al. (2007) noted moderate correlations for Spanish bilingual children between TERA-2 in spring kindergarten and WLBP-R Letter-Word Identification scores (r = .32 to .72) with the standard scores more strongly correlated than raw scores.

Spring kindergarten performance on the TERA-2 correlated with mother to child use of English in the home during the first and second years of Head Start (r = .43), with weaker concurrent relations (during the spring of kindergarten r = .34; Hammer et al., 2009).

The TVIP/Spanish PLS-3 component score had moderate negative relations with the TERA-2 (r = -.33 to -.48). Associations of the spring kindergarten TERA-2 with the preschool PPVT-III/TELD-3 component scores were nonsignificant (Hammer et al., 2007).

Woodcock-Johnson III (WJ-III), English Form

Bateria III Woodcock-Munoz (WM-III), Spanish Form

See also Woodcock Language Proficiency Battery–Revised for similar subtests

Description/Purpose of Assessment

This measure of overall language proficiency has both English (Woodcock-Johnson III [WJ-III]; Woodcock, McGrew, and Mather, 2007) and a Spanish edition (Woodcock-Munoz III [WM-III]; Woodcock, Munoz-Sandoval, McGrew, and Mather, 2007). It includes 13 assessments divided into oral language, reading, and written language subtests. Studies of Head Start, including all the FACES studies and the Head Start Impact Study, administered two subtests to address literacy development: (1) Letter-Word Identification and (2) Dictation (WJ-R) or Spelling (WJ-III).[20]

Evidence of Reliability

Four studies (Head Start Impact Study and FACES 2006 and 2009: Puma et al., 2010; Aikens et al. 2011; Hulsey et al. 2010; West et al. 2008) reported strong reliability for the English WJ-III Spelling (= .70 to .89) and Letter-Word Identification (= .81 to .94) subtests. FACES 2000 (ACF, 2006) reported similar information for the WJ-R English Dictation (= .77, fall 2000, and .77, spring 2001) and Letter-Word Identification (= .84, fall 2000, and .86, spring 2001).

Three studies (FACES 2006 and 2009: Aikens et al. 2011; Hulsey et al. 2010; West et al. 2008) reported moderate evidence for reliability for the WM-III Spanish Spelling (= .66, fall 2000; .69, fall 2006; .67, spring 2007; .66, fall 2009) and Letter-Word Identification (= .75 to .83) as well. One study (FACES 2000: ACF, 2006) reported similar information for the WM-R Spanish Dictation (= .77, fall 2000 and .73, spring 2001) and Letter-Word Identification (= .75, fall 2000, and .78, spring 2001).

Evidence of Validity

Dictation

In the FACES 1997 sample, the spring Head Start WJ-R Dictation score correlated with end of kindergarten ECLS-K Reading scale score (r = .48; beta = .14) and with the ECLS-K General Knowledge scale score (r = .46; beta = .11) (Zill et al. 2003). The sample was limited to children who took the assessments in English at all timepoints and so did not include many DLLs; however, these estimates provide a point of comparison for validity estimates for low-income populations on these measures.

[20] With the third edition, the revision of the Dictation subtest was renamed Spelling.

Letter-Word Identification

In the FACES 1997 sample, the spring Head Start WJ–R Letter-Word Identification score correlated with end of kindergarten ECLS-K Reading scale score (r =.55; beta = .32) and with the ECLS-K General Knowledge scale score (r= .39; n.s. in the multiple regression) (Zill et al. 2003). The sample was limited to children who took the assessments in English at all timepoints and so did not include many DLLs; however, these estimates give a point of comparison for validity estimates for low-income populations on these measures.

APPENDIX C

SUMMARY TABLES: VOCABULARY, LANGUAGE, AND LITERACY ASSESSMENTS

Table C.1. Vocabulary Measures Used in Small-Scale Studies and Government Reports

Assessment	Studies	Reliability Evidence	Validity Evidence
The MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 1993)[21]	Administration for Children and Families (2011) Conboy & Mills (2006) Conboy & Thal (2006) Marchman, Fernald, & Hurtado (2010) Marchman, Martinez-Sussmann, & Dale (2004) Parra, Hoff, & Core (2011) Vagh, Pan, & Mancilla-Martinez (2009)	α = .98 for English, α = .98 for Spanish for the CDI raw score (Administration for Children and Families, 2011)	CONCURRENT: • Correlation with Inventario II = .43. (Conboy & Thal, 2006) • Mean scores were comparable to those on the Inventario II (but correlation was negative and nonsignificant). Children scored below age-based norms as compared to monolinguals, scoring on average below the 50th percentile. Correlation with Spanish:English exposure ratio = -.63; with mean reaction time in English = -.63. (Marchman, Fernald, & Hurtado, 2010) • Means scores reported for CDI and Inventario II. CDI vocabulary scores, grammatical complexity, and M3L-words correlated with age (r = .25 to .33) and Spanish:English exposure ratio (r = .25 to .37). Composite CDI/Inventario II score correlated with age (r = .57). Correlation between vocabulary and grammatical complexity (r = .74); and M3L-words and complexity (r = .78). (Marchman, Martinez-Sussmann, & Dale, 2004) • Correlation between English vocabulary and English grammatical complexity = .84; English vocabulary and Spanish grammatical complexity = -.29. (Parra, Hoff, & Core, 2011) • Means reported by language group at 24, 27, 30, 33 and 36 months (by parent versus teacher report). (Vagh, Pan, & Mancilla-Martinez, 2009) • Parent CDI scores were correlated with concurrent WLPB-R (r = .46 to .52) and PPVT-III (r = .38 to .47) scores at 30 months and 36 months. Parent-teacher composite CDI scores were significantly associated with WLPB-R (r = .38) and PPVT-III (r = .34) scores at 36 months. Teacher CDI scores were not associated with the WLPB-R or PPVT-III. (Vagh, Pan, & Mancilla-Martinez, 2009)

[21] See also: Macarthur-Bates Inventarios Del Desarrolo De Habilidades Comunicativas (Inventarios) Jackson-Maldonado, Bates, & Thal, 2003)

Table C.1 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
			PREDICTIVE: Correlation between percentage of language exposure in English at 22 months and English vocabulary size at 25 months = .72; and English grammatical complexity at 25 months = .58. Further supported by hierarchical regression analyses. (Parra, Hoff, & Core, 2011)
MacArthur-Bates Inventarios del Desarrollo de Habilidades Comunicativas (Inventarios; Jackson-Maldonado, Bates, & Thal, 2003)	n.a.	n.a.	n.a.
Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000)	Anthony et al. (2009)	n.a.	CONCURRENT: • Composite EOWPVT/ROWPVT correlation with PCTOPPP = .69; with SPCTOPPP = .26; with ROWPVT-SBE/EOWPVT-SBE = .23. As reported in Brownell (2000; 2001), correlations with standardized vocabulary measures range from .67 to .90. PREDICTIVE: • Composite EOWPVT/ROWPVT correlation with Time 2 PCTOPPP = .52; with Time 2 SPCTOPPP = .38.
Expressive One-Word Picture Vocabulary Test: Spanish Bilingual Edition (EOWPVT-SBE; Brownell, 2001)	Aikens et al. (2011) Anthony et al. (2009)	n.a.	CONCURRENT: • Composite EOWPVT-SBE/ROWPVT-SBE correlation with PCTOPPP = .35; with SPCTOPPP = .64; with ROWPVT/EOWPVT = .23. As reported in Brownell (2000; 2001), correlations with standardized vocabulary measures range from .67 to .90. PREDICTIVE: • Composite EOWPVT-SBE/ROWPVT-SBE correlation with Time 2 PCTOPPP = .46; with Time 2 SPCTOPPP = .52.
Peabody Picture Vocabulary Test–4 (PPVT-4; Dunn & Dunn, 2007)	Aikens et al., 2011 Hulsey et al., 2010 West et al., 2008	α = .97 (Aikens et al., 2011) α = .94 (Hulsey et al., 2010) α = .97 (West et al., 2008)	n.a.

Table C.1 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Peabody Picture Vocabulary Test–III (PPVT-III; Dunn & Dunn, 1997)	ACF (2006) Bialystok, Luk, Peets, & Yang (2010) Davidson, Raschke, & Pervez (2010) Dickinson, McCabe, Clark-Chiarelli, & Wolf (2004) Hammer, Lawrence, & Miccio (2008a) Hammer, Davison, Lawrence, & Miccio (2009) Hammer, Lawrence, & Miccio (2007) Marinova-Todd, Zhao, & Bernhardt (2010) Zill et al. 2003	Internal consistency estimates ranged from .96 to .97 (ACF, 2006; Zill et al., 2003)	CONCURRENT: • Mean scores reported by language group and age. Monolinguals outperformed bilinguals at every age comparison. Overall, monolinguals averaged 107 (SD = 12.3) and bilinguals averaged 96 (SD = 13.0). (Bialystok, Luk, Peets, & Yang, 2010) • Study 1: Mean scores reported by language group. No significant differences emerged between English monolingual and bilingual children. Study 2: Main effect of age, such that older children (5- and 6-year-olds) outperformed younger children (3- and 4-year-olds), irrespective of language group. (Davidson, Raschke, & Pervez, 2010) • Correlation with TVIP = .23; with Spanish EPAP = .32; with English EPAP = .48; with ELP = .30.(Dickinson, McCabe, Clark-Chiarelli, & Wolf, 2004) • Positively correlated with use of English in the home within time (*r* = .28 to .38). (Hammer, Davison, Lawrence, & Miccio, 2009) • Mean scores reported by language group. Average English scores were higher among children in the HEC group. (Hammer, Lawrence, & Miccio, 2007) • Means scores reported by age and language group. Scores of English monolinguals were higher than the scores of bilinguals (across both age groups). (Marinova-Todd, Zhao, & Bernhardt (2010) • Means reported by language group at 30 and 36 months. Bilingual children had lower scores than monolingual children. Vagh, Pan, & Mancilla-Martinez (2009) • Parent CDI scores were correlated with concurrent PPVT–III scores (*r* = .38 to .47) at 30 months and 36 months. Parent-teacher composite CDI scores were significantly associated with PPVT–III scores (*r* = .34) at 36 months. Vagh, Pan, & Mancilla-Martinez (2009) • Mean scores reported by language group. At each of four timepoints, average scores were higher among children in the HEC group than SEC group. (Hammer, Lawrence, & Miccio, 2008a)

Table C.1 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
			PREDICTIVE: • Correlation with fall English EPAP = .29. (Dickinson, McCabe, Clark-Chiarelli, & Wolf, 2004) • Positively correlated with use of English in the home across assessments (r = .28 to .41). Stability of PPVT-III scores over the 3-year period ranged from .48 to .75. (Hammer, Davison, Lawrence, & Miccio, 2009) • According to a random intercept model, children's scores increased across the four timepoints. At baseline, children in the SEC group had significantly lower scores than children in the HEC group, and these differences maintained over time. (Hammer, Lawrence, & Miccio, 2008a) • In the FACES 1997 sample, the spring Head Start PPVT-III score correlated with end of Kindergarten ECLS-K General Knowledge scale score (r= .77; beta = .62); and with the ECLS-K Reading scale score (r =.42; ns for the multiple regression) Zill et al. (2003)
Peabody Picture Vocabulary Test–Revised (PVVT-R; Dunn & Dunn, 1981) (Chinese version; Lu & Liu, 1998)	Farver, Xu, Eppe, & Lonigan (2006) Marinova-Todd, Zhao, & Bernhardt (2010)	n.a.	CONCURRENT: • PPVT-R/TVIP correlated with child age (r = .26); maternal education (r = .31); paternal education (r = .28); HLEQ parents' literacy habits subscale scores (r = .26), children's literacy interest subscale scores (r = .38), and parents' literacy involvement subscale scores (r = .25); BASC (r = .34), PSI (r = -.28), and TRT scores (r = .24). Farver, Xu, Eppe, & Lonigan (2006) • Means scores reported by age and language group. Scores of Mandarin monolinguals were higher than the scores of bilinguals (6-year-old group only). Marinova-Todd, Zhao, & Bernhardt (2010)
Receptive One-Word Picture Vocabulary Test (ROWPVT; Brownell 2000)	Anthony et al. (2009)	n.a.	CONCURRENT: • Composite EOWPVT/ROWPVT correlation with PCTOPPP = .69; with SPCTOPPP = .26; with ROWPVT-SBE/EOWPVT-SBE = .23. As reported in Brownell (2000; 2001), correlations with standardized vocabulary measures range from .71 to .97. PREDICTIVE: • Composite EOWPVT/ROWPVT correlation with Time 2 PCTOPPP = .52; with Time 2 SPCTOPPP = .38.

Table C.1 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Receptive One-Word Picture Vocabulary Test : Spanish Bilingual Edition (ROWPVT-SBE; Brownell 2001)	Anthony et al. (2009)	n.a.	CONCURRENT: • Composite EOWPVT-SBE/ROWPVT-SBE correlation with PCTOPPP = .35; with SPCTOPPP = .64; with ROWPVT/EOWPVT = .23. As reported in Brownell (2000; 2001), correlations with standardized vocabulary measures range from .71 to .97. PREDICTIVE: • Composite EOWPVT-SBE/ROWPVT-SBE correlation with Time 2 PCTOPPP = .46; with Time 2 SPCTOPPP = .52.
Spanish-English Vocabulary Checklist (SEVC; Patterson, 1998)	Patterson (2002)	Documented high short-term, test-retest reliability (as reported by Patterson, 1998)	CONCURRENT: • Significantly correlated with the number of different words children use in language samples with the reporting parent (as reported by Patterson, 2000). • Mean vocabulary size in English and Spanish reported by child gender. Males = 86 and 53 in English and Spanish, respectively; females = 110 and 48 in English and Spanish, respectively. Correlation between English vocabulary and English input = .33; English reading = .40. Correlation between Spanish vocabulary and Spanish input = .32; Spanish reading = .35. • Child age and reading frequency uniquely contributed to vocabulary size in both English and Spanish.

Table C.1 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Test de Vocabulario en Imágenes Peabody (TVIP; Dunn, Padilla, Lugo, & Dunn, 1986)	Administration for Children and Families [ACF] (2006) Aikens et al. (2011) Hammer, Davison, Lawrence, & Miccio (2009) Hammer, Lawrence, & Miccio (2008a) Dickinson, McCabe, Clark-Chiarelli, & Wolf (2004) Hammer, Lawrence, & Miccio (2007) Farver, Xu, Eppe, & Lonigan (2006) Hulsey et al. (2010)	Internal consistency = .92 to .94 (ACF, 2006); .94 (Hulsey et al., 2010); .93 (Aikens et al., 2011)	CONCURRENT: • Negatively correlated with use of English in the home within time (r = -.30 to -.40). (Hammer, Davison, Lawrence, & Miccio, 2009) • Mean scores reported by language group. At each of four timepoints, average scores were higher among children in the SEC group than HEC group. (Hammer, Lawrence, & Miccio, 2008a) • Correlation with PPVT-III = .23; with Spanish EPAP = .40; with English EPAP = .35; with ELP = .33. (Dickinson, McCabe, Clark-Chiarelli, & Wolf, 2004) • Mean scores reported by language group. Average Spanish scores were higher for children in the SEC group. (Hammer, Lawrence, & Miccio, 2007) • PPVT-R/TVIP correlated with child age (r = .26); maternal education (r = .31); paternal education (r = .28); HLEQ parents' literacy habits subscale scores (r = .26), children's literacy interest subscale scores (r = .38), and parents' literacy involvement subscale scores (r = .25); BASC (r = .34), PSI (r = -.28), and TRT scores (r = .24). (Farver, Xu, Eppe, & Lonigan, 2006) PREDICTIVE: • Negatively correlated with use of English in the home during children's second year of Head Start (r = -.25 to -.36) and the spring of kindergarten (r = -.40 to -.45). Stability of TVIP scores over the 3-year period ranged from .50 to .71. (Hammer, Davison, Lawrence, & Miccio, 2009) • According to a random slope model, children's scores increased across the four timepoints. At baseline, children in the SEC group had significantly higher scores than children in the HEC group. (Hammer, Lawrence, & Miccio, 2008a)

Table C.1 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Woodcock Language Proficiency Battery–Revised, English Form (WJ-III; Woodcock; 1995) Picture Vocabulary	Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003) Vagh, Pan, & Mancilla-Martinez (2009)	n.a.	CONCURRENT: • Spanish-English bilinguals (ECS): fall = 68.1 (SD = 19.2); spring = 70.5 (SD = 18.5) (Tabors, Páez, & López, 2003) • Time 1 to Time 3: PreK = 70.3 (SD = 18.8), K = 72.3 (SD = 19.6), 1st = 79.4 (SD = 19.5) Preschool English Vocabulary predicted English Letter-Word Identification scores in first grade (Vagh, Pan, & Mancilla-Martinez, 2009) PREDICTIVE: • Negatively correlated with Spanish Picture Vocabulary scores (r = -.28) (Páez, Tabors, & Lopez, 2007) • Means reported by language group at 30 and 36 months. Bilingual children had lower scores than monolingual children. Parent CDI scores were correlated with concurrent WLPB-R scores (r = .46 to .52) at 30 months and 36 months. Parent-teacher composite CDI scores were significantly associated with WLPB-R scores (r = .38) at 36 months. (Rinaldi & Páez, 2008)
Woodcock Language Proficiency Battery–Revised, Spanish Form (WM-III; Woodcock & Muñoz-Sandoval, 1995) Vocabulario Sobre Dibujos	Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Negatively correlated with English Picture Vocabulary scores (r = -.28). Spanish-English bilinguals (ECS): 65.1 (SD = 16.7). Spanish-speaking monolinguals (PRC): 84.0 (SD = 10.7) (Tabors, Páez, & López, 2003) PREDICTIVE: • Spanish-English bilinguals (ECS): fall = 65.2 (SD = 16.6); spring = 62.0 (SD = 19.0). Spanish-speaking monolinguals (PRC): fall = 84.0 (SD = 10.7); spring = 86.9 (SD = 13.4). Monolingual Spanish-speakers in the PRC sample performed higher than bilinguals in the ECS sample at Time 1 (d = 1.34) and Time 2 (d = 1.52). (Páez, Tabors, & Lopez, 2007) • Time 1 to Time 3: PreK = 63.4 (SD = 18.9), K= 53.2 (SD = 22.1), 1st = 48.8 (SD = 24.6). Preschool Spanish Vocabulary predicted English Letter-Word Identification skills in first grade. (Rinaldi & Páez, 2008)

Table C.2. Language Measures Used In Small-Scale Studies and Government Reports

Assessment	Studies	Reliability Evidence	Validity Evidence
Ages & Stages Questionnaires – Third Edition: Communication (ASQ-3 Communication; Squires, Twombly, Bricker, and Potter, 2009)	Vogel et al., 2011	α = .65 to .73(different age forms)	n.a.
Bilingual English Spanish Assessment (BESA; Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, 2009)	Bedore, Peña, García, & Cortez (2005) Castilla, Restrepo, & Perez-Leroux (2009)	Characteristic Properties subtest: Inter-rater reliability for scoring = 95%. (Bedore, Peña, García, & Cortez, 2005) Phase 2 Semantic subtest: Inter-rater reliability for scoring = 99% (Bedore, Peña, García, & Cortez, 2005)	CONCURRENT: • Characteristic Properties subtest: Mean scores reported by language group. No differences were found on Spanish scores. Significant differences emerged between the English scores of children in the bilingual Spanish (BS), bilingual English (BE), and primarily English-speaking (PE) groups. The scores for the BS group were lower than those for the PE group. (Bedore, Peña, García, & Cortez, 2005) • Phase 2 Semantic subtest: The difference between children in the bilingual Spanish (BS) and bilingual English (BE) groups was greater on the English subtest than on the Spanish subtest. (Bedore, Peña, García, & Cortez, 2005) PREDICTIVE: • Spanish Semantic Subtest, Spanish-ST: Correlated with English-MT Cloze Test scores (*r* = .50) and English-ST scores (*r* = .59). (Castilla, Restrepo, & Perez-Leroux, 2009) • English Semantic Subtest, English-ST: Correlated with Spanish-MT Cloze Test scores (*r* = .45), Spanish-MT Sentence Repetition (*r* = .69), and Spanish-ST scores (*r* = .59). (Castilla, Restrepo, & Perez-Leroux, 2009) • Spanish Morphosyntax Subtest, Spanish-MT: correlation between Spanish-MT and English-MT Cloze Test scores = .74; with English-ST scores = .45. Correlation between Spanish-MT Sentence Repetition scores English-MT Cloze Test scores = .69; with English-ST scores = .69. (Castilla, Restrepo, & Perez-Leroux, 2009) • English Morphosyntax Subtest, English-MT: Correlated with Spanish-MT Cloze Test scores (*r* = .74), Spanish-MT Sentence Repetition (*r* = .69), and Spanish-ST scores (*r* = .50). (Castilla, Restrepo, & Perez-Leroux, 2009)

Table C.2 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Bilingual English Spanish Oral Language Screener (BESOS; Peña, Bedore, Gutierrez-Clellen, Iglesias, & Goldstein, in preparation)	Bohman, Bedore, Peña, Mendez-Perez, & Gillam (2010)	Semantics subtest: Test-retest (n = 20): r = .70 for Spanish semantics, r = .64 for English semantics. Morphosyntax subtest: Test-retest (n = 20): r = .86 for Spanish morphosyntax, r = .75 for English morphosyntax.	CONCURRENT: • Semantics subtest: Examined factors associated with scores on Spanish and English semantics subtests, including percent language input/output, age, gender, free/reduced lunch status, site, and parent education. • Morphosyntax subtest: Examined factors associated with scores on Spanish and English morphosyntax subtests, including percent language input/output, age, gender, free/reduced lunch status, site, and parent education.
Pre-Language Assessment Scale 2000 (Pre-LAS; DeAvila and Duncan, 2000)	Gonzalez & Uhing (2008) Lopez & Greenfield (2004) ECLS-B FACES 2006 FACES 2009 HSIS	Cronbach's alpha across subtests: English version = .86 to .90; Spanish version = .66 to .88 (as reported by Duncan & DeAvila, 1998) (Lopez & Greenfield, 2004)	• CONCURRENT: • Correlation between Spanish PreLAS scores and Extended Family subscale scores (r = .32); between English PreLAS scores and Library Use subscale scores (r = .39). In standardization sample: Scores for children from minority-language backgrounds were significantly lower than for children from English-only backgrounds. Younger children scored consistently lower than older children and were less proficient across test sections. Scores and proficiency levels increased significantly by grade level. (Gonzalez & Uhing, 2008) • Subtests were all significantly correlated within language. The correlation between the PreLAS English and Spanish versions were not significant. Correlation between English PreLAS and English Phonological Sensitivity Test (researcher-developed) = .52; between Spanish PreLAS and Spanish Phonological Sensitivity Test (researcher-developed) = .33. English PreLAS, Spanish PreLAS, and Spanish PST scores uniquely contributed to English PST scores in concurrent models. (Lopez & Greenfield, 2004)

Table C.2 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Preschool Language Scale 3 (PLS-3; Zimmerman, Steiner, & Pond, 1992)	Hammer, Lawrence, & Miccio (2007) Hammer, Lawrence, & Miccio (2008a) Hammer, Lawrence, & Miccio (2008b)	Auditory Comprehension: Internal consistency = .81	CONCURRENT: • Auditory Comprehension: Mean scores reported by language group. Average Spanish scores were higher for children in the SEC group. (Hammer, Lawrence, & Miccio, 2007) • Spanish version; Auditory Comprehension subtest: Mean scores reported by language group. At each of four timepoints, average scores were higher among children in the SEC group than HEC group. Child age positively associated with scores. (Hammer, Lawrence, & Miccio, 2008a) • Spanish version; Auditory Comprehension subtest: Mean scores reported by language group. At each of four timepoints, average scores were higher (and SDs smaller) among children in the SEC group than HEC group. (Hammer, Lawrence, & Miccio, 2008b) PREDICTIVE: • According to a random intercept model, children's scores increased across the four timepoints. At baseline, children in the SEC group had significantly higher scores than children in the HEC group, and these differences maintained over time. (Hammer, Lawrence, & Miccio, 2008a) • Spanish version; Auditory Comprehension subtest Mean scores of children in the HEC and SEC groups increased over the first three timepoints, followed by a decline. (Hammer, Lawrence, & Miccio, 2008b)

Table C.2 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Test of Early Language Development–3 (TELD-3; Hresko, Reid & Hammill, 1999)	Hammer, Lawrence, & Miccio (2007) Hammer, Lawrence, & Miccio (2008a) Hammer, Lawrence, & Miccio (2008b)	Receptive Language: Internal consistency = .91 (Hammer, Lawrence, & Miccio, 2007)	CONCURRENT: Receptive Language: • Mean scores reported by language group. Average English scores were higher among children in the HEC group. (Hammer, Lawrence, & Miccio, 2007) • Mean scores reported by language group. At each of four timepoints, average scores were higher among children in the HEC group than SEC group. Child age positively associated with scores. (Hammer, Lawrence, & Miccio, 2008a) • Mean scores reported by language group. At each of four timepoints, average scores were higher (and SDs larger) among children in the HEC group than SEC group. (Hammer, Lawrence, & Miccio, 2008b) PREDICTIVE: • According to a random intercept model, children's scores increased across the four timepoints. At baseline, children in the SEC group had significantly lower scores than children in the HEC group. (Hammer, Lawrence, & Miccio, 2008a) • Mean scores increased over time for children in the HEC and SEC groups. (Hammer, Lawrence, & Miccio, 2008b)
Woodcock Language Proficiency Battery–Revised, English Form (WJ-III; Woodcock; 1995) Memory for Sentences	Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Correlated with Spanish Memory for Sentences scores (r = .25) (Tabors, Páez, & López, 2003) PREDICTIVE: • Spanish-English bilinguals (ECS): fall = 73.1 (SD = 19.0); spring = 77.2 (SD = 14.7) (Páez, Tabors, & Lopez, 2007) • Time 1 to Time 3: PreK = 78.0 (SD = 14.9), K = 77.3 (SD = 15.6), 1st = 85.1 (SD = 15.5). Preschool English Memory for Sentences predicted English Letter-Word Identification scores in first grade. (Rinaldi & Páez, 2008)

Table C.2 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Woodcock Language Proficiency Battery–Revised, Spanish Form (WM-III; Woodcock & Muñoz-Sandoval, 1995) Memoria para Frases	Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Correlated with English Memory for Sentences scores (r = .25). Spanish-English bilinguals (ECS): 69.7 (SD = 17.5). Spanish-speaking monolinguals (PRC): 83.8 (SD = 17.5) (Tabors, Páez, & López, 2003) PREDICTIVE: • Spanish-English bilinguals (ECS): fall = 70.1 (SD = 16.9); spring = 72.0 (SD = 16.6). Spanish-speaking monolinguals (PRC): fall = 83.8 (SD = 17.5); spring = 88.6 (SD = 12.7). Monolingual Spanish-speakers in the PRC sample performed higher than bilinguals in the ECS sample at Time 1 (d = 0.82) and Time 2 (d = 1.12). (Páez, Tabors, & Lopez, 2007) • Time 1 to Time 3: PreK = 72.5 (SD = 15.8), K = 66.8 (SD = 16.6), 1st = 69.2 (SD = 13.9) (Rinaldi & Páez, 2008)

[a]Monolingual Spanish.

Table C.3. Literacy Measures Used In Small-Scale Studies and Government Reports

Assessment	Studies	Reliability Evidence	Validity Evidence
Early Phonological Awareness Profile (EPAP; Dickinson & Chaney, 1997)	Dickinson, McCabe, Clark-Chiarelli, & Wolf (2004)	LINGUISTICALLY DIVERSE: As reported in a prior study with English-speaking children (n = 984), Cronbach's alpha = .93. MONOLINGUAL: Cronbach's alpha = .94 for the Spanish version; .93 for the English version.	CONCURRENT: Correlation between fall English and Spanish EPAP = .60; between spring English and Spanish EPAP = .78. Age is correlated with Spanish EPAP scores (*r* = .25). PREDICTIVE: Correlation between fall and spring Spanish EPAP = .42; between fall and spring English EPAP = .35; between fall English EPAP and spring Spanish EPAP = .32; between spring English EPAP and fall Spanish EPAP = .41; between fall English EPAP and spring PPVT = .29; between fall Spanish EPAP and spring ELP = .42; between fall English EPAP and spring ELP = .32.
Emergent Literacy Profile (ELP; Dickinson & Chaney, 1997)	Dickinson, McCabe, Clark-Chiarelli, & Wolf (2004)	As reported in a prior study with English-speaking children (n = 578), Cronbach's alpha = .86. In the current study, Cronbach's alpha = .82.	CONCURRENT: Correlation with PPVT-III = .30; with TVIP = .33; Spanish EPAP = .48; with English EPAP = .48. PREDICTIVE: Correlation with fall Spanish EPAP = .42; with fall English EPAP = .32.
Preschool Comprehensive Test of Phonological and Print Processing (PCTOPPP; Lonigan et al., 2002)	Anthony et al. (2009)	Elision and Blending subtests demonstrate good internal consistency (as reported by Anthony et al., 2006; Anthony et al., 2007).	CONCURRENT: Correlation with SPCTOPPP = .43; with ROWPVT/EOWPVT = .69; with ROWPVT-SBE/EOWPVT-SBE = .35. Elision and Blending subtests demonstrate good convergent and discriminant validity (Anthony et al., 2006; Anthony et al., 2007). PREDICTIVE: Correlation with Time 2 PCTOPPP (stability) = .70; with Time 2 SPCTOPPP = .50.
Story & Print Concepts (FACES Research Team 1997)	Administration for Children and Families (ACF), 2006 Hulsey et al. 2010 West et al., 2008 Zill et al. 2001 Zill et al. 2003	Internal consistency: English Version Book Knowledge .57 to .61 Print Conventions .73 to .84 Comprehension .40 to .43 (ACF, 2006) Total score = .70 (Hulsey et al. 2010; West et al. 2008) Spanish version: Book Knowledge .43 Print Conventions .59 Comprehension .39 (ACF, 2006)	Weak, though significant, correlation found between teacher's educational attainment children's spring story and print concepts (r ranged from .09 to .14, p < .05), and this was only examined for children who were assessed in English at every timepoint (Zill et al. 2003)

Table C.3 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Test of Early Reading Ability-2 (TERA-2; Reid, Hresko, & Hammill, 1991)	Hammer, Davison, Lawrence, & Miccio (2009) Hammer, Lawrence, & Miccio (2007) Hammer, Miccio, & Wagstaff (2003)	REPORTED FROM MANUAL: • Median internal consistency = .91. (Hammer, Davison, Lawrence, & Miccio, 2009) • Internal consistency = .91 (Hammer, Lawrence, & Miccio, 2007) • Hammer, Miccio, & Wagstaff (2003)	CONCURRENT: • Correlated with use of English in the home only during the spring of kindergarten (r = .34) (Hammer, Davison, Lawrence, & Miccio, 2009) • Mean scores reported by language group (no significant differences). Correlation with WLBP-R (r = .32 to .72) (Hammer, Lawrence, & Miccio, 2007) • Mean scores reported by language group (not significantly different). (Hammer, Miccio, & Wagstaff, 2003) PREDICTIVE: • Spring kindergarten performance correlated with use of English in the home during the first and second years of Head Start (r = .43). Stability of TERA-2 scores over the 3-year period ranged from .34 to .53. (Hammer, Davison, Lawrence, & Miccio, 2009) • Correlations with PPVT-III/TELD-3 component scores are nonsignificant; with TVIP/PLS-3 component score (r = -.33 to -.48) (Hammer, Lawrence, & Miccio, 2007) • Overall, children's scores were significantly lower at Time 2 compared to Time 1 (language group differences not observed) (Hammer, Miccio, & Wagstaff, 2003)

Table C.3 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Woodcock Language Proficiency Battery–Revised, English Form (WJ-III; Woodcock; 1995) Letter-Word Identification	Hammer, Lawrence, & Miccio (2007) Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003)	Internal consistency = .96 (Hammer, Lawrence, & Miccio, 2007)	CONCURRENT: • Mean scores reported by language group (no significant differences). Correlation with TERA-2 (r = .32 to .72) (Hammer, Lawrence, & Miccio, 2007) • Correlated with Spanish Letter-Word Identification scores (r = .51) (Tabors, Páez, & López, 2003) PREDICTIVE: • Correlations with PPVT-III/TELD-3 component scores are nonsignificant; with TVIP/PLS-3 component score (r = -.33 to -.46) (Hammer, Lawrence, & Miccio, 2007) • Spanish-English bilinguals (ECS): fall = 90.8 (SD = 9.6); spring = 91.1 (SD = 12.6) (Páez, Tabors, & Lopez, 2007) • Time 1 to Time 3: PreK = 91.2 (SD = 12.5), K = 96.8 (SD = 14.4), 1st = 105.6 (SD = 15.8) Preschool English Vocabulary and Memory for Sentences predicted Letter-Word Identification scores in first grade. In a separate model, preschool English Vocabulary, Spanish Vocabulary, English Memory for Sentences, and Spanish Letter-Word Identification predicted English Letter-Word Identification skills in first grade. (Rinaldi & Páez, 2008)
Woodcock Language Proficiency Battery–Revised, English Form (WJ-III; Woodcock; 1995) Dictation/Spelling	Páez, Tabors, & Lopez (2007) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Correlated with Spanish Dictation scores (r = .50) (Tabors, Páez, & López, 2003) PREDICTIVE: • Spanish-English bilinguals (ECS): fall = 88.7 (SD = 13.8); spring = 91.6 (SD = 14.5) (Páez, Tabors, & Lopez, 2007)
Woodcock Language Proficiency Battery–Revised, Spanish Form (WM-III; Woodcock & Muñoz-Sandoval, 1995) Dictado	Páez, Tabors, & Lopez (2007) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Correlated with English Dictation scores (r = .50). Spanish-English bilinguals (ECS): 90.3 (SD = 12.8). Spanish-speaking monolinguals (PRC): 86.7 (SD = 15.8) (Tabors, Páez, & López, 2003) PREDICTIVE: • Spanish-English bilinguals (ECS): fall = 90.4 (SD = 13.0); spring = 90.9 (SD = 9.2). Spanish-speaking monolinguals (PRC): fall = 86.7 (SD = 15.8); spring = 90.6 (SD = 10.8) (Páez, Tabors, & Lopez, 2007)

Table C.3 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Woodcock Language Proficiency Battery–Revised, Spanish Form (WLPB=R Spanish; Woodcock & Muñoz-Sandoval, 1995) (Identificación de Letras y Palabras)	Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Correlated with English Letter-Word Identification scores (r = .51). Spanish-English bilinguals (ECS): 88.8 (SD = 7.4). Spanish-speaking monolinguals (PRC): 87.9 (SD = 6.3) (Tabors, Páez, & López, 2003) • PREDICTIVE: • Spanish-English bilinguals (ECS): fall = 88.9 (SD = 7.4); spring = 85.6 (SD = 9.3). Spanish-speaking monolinguals (PRC): fall = 87.9 (SD = 6.3); spring = 85.2 (SD = 11.1) (Páez, Tabors, & Lopez, 2007) • Time 1 to Time 3: PreK = 86.2 (SD = 8.5), K = 80.1 (SD = 19.5), 1st = 81.8 (SD = 29.9). Preschool Spanish Letter-Word Identification predicted English Letter-Word Identification skills in first grade. (Rinaldi & Páez, 2008)
Woodcock Language Proficiency Battery–Revised, English Form (WLPB-R; Woodcock; 1995) Letter-Word Identification	Hammer, Lawrence, & Miccio (2007) Páez, Tabors, & Lopez (2007) Rinaldi & Páez (2008) Tabors, Páez, & López (2003)	REPORTED FROM MANUAL: Internal consistency = .96 (Hammer, Lawrence, & Miccio, 2007)	CONCURRENT: • Mean scores reported by language group (no significant differences). Correlation with TERA-2 (r = .32 to .72) (Hammer, Lawrence, & Miccio, 2007) • Correlated with Spanish Letter-Word Identification scores (r = .51) (Tabors, Páez, & López, 2003) PREDICTIVE: • Correlations with PPVT-III/TELD-3 component scores are nonsignificant; with TVIP/PLS-3 component score (r = -.33 to -.46) (Hammer, Lawrence, & Miccio, 2007) • Spanish-English bilinguals (ECS): fall = 90.8 (SD = 9.6); spring = 91.1 (SD = 12.6) (Páez, Tabors, & Lopez, 2007) • Time 1 to Time 3: PreK = 91.2 (SD = 12.5), K = 96.8 (SD = 14.4), 1st = 105.6 (SD = 15.8) Preschool English Vocabulary and Memory for Sentences predicted Letter-Word Identification scores in first grade. In a separate model, preschool English Vocabulary, Spanish Vocabulary, English Memory for Sentences, and Spanish Letter-Word Identification predicted English Letter-Word Identification skills in first grade. (Rinaldi & Páez, 2008)

Table C.3 *(continued)*

Assessment	Studies	Reliability Evidence	Validity Evidence
Woodcock Language Proficiency Battery–Revised, English Form (WLPB-R; Woodcock; 1995) Dictation/Spelling	Páez, Tabors, & Lopez (2007) Tabors, Páez, & López (2003)	n.a.	CONCURRENT: • Correlated with Spanish Dictation scores (r = .50) (Tabors, Páez, & López, 2003) PREDICTIVE: • Increase in score across time: Spanish-English bilinguals (ECS): fall = 88.7 (SD = 13.8); spring = 91.6 (SD = 14.5) (Páez, Tabors, & Lopez, 2007)
Woodcock-Johnson Tests of Achievement -Revised and Third Edition (WJ-R and WJ- III; Mather & Woodcock (2001, 2007) Letter-Word Identification	Administration for Children and Families, (2006) Hulsey et al. (2010) West et al., (2008) Zill et al. (2001) Zill et al. (2003)	**WJ-R** Letter-Word Identification Internal consistency estimates ranged from α = .84 to α = .86 (ACF, 2006; Zill et al., 2003) **WJ-III** Letter-Word Identification α = .81 (Hulsey et al., 2010; West et al., 2008) α = .85 (Aikens et al., 2011)	PREDICTIVE: Spring Head Start score correlated with end of Kindergarten ECLS-K Reading score with the FACES 1997 sample (r =.55; beta = .32); spring Head Start score correlated with ECLS-K General Knowledge (r =.40), but was not a significant predictor in the multiple regression analysis (Zill et al. 2003)
Woodcock-Johnson Tests of Achievement -Revised and Third Edition (WJ-R and WJ- III; Mather & Woodcock (2001, 2007) Dictation/Spelling	Administration for Children and Families, (2006) Hulsey et al.(2010) West et al., (2008) Zill et al. (2001) Zill et al. (2003)	**WJ-R** Dictation internal consistency estimates ranged from α = .71 to α = .77 (ACF, 2006; Zill et al., 2003) **WJ-III** Spelling α = .81 (Hulsey et al., 2010; West et al., 2008) α = .79 (Aikens et al., 2011)	PREDICTIVE: With children taking the assessment in English in the FACES 1997 sample, the spring Head Start score correlated with end of Kindergarten ECLS-K Reading scale score (r =.48; beta = .14) and with the ECLS-K General Knowledge scale score (r= .46; beta = .11) (Zill et al. 2003)
Woodcock-Muñoz Batería Revised and Third Edition (WM-R and WM-III; Identificación de Letras y Palabras	Administration for Children and Families, 2006 Hulsey et al. 2010 West et al., 2008 Zill et al. 2001 Zill et al. 2003	**WM-R** Identificación de Letras y Palabras internal consistency estimates ranged from α = .75 to α = .83 (ACF, 2006; Zill et al., 2003) **WM-III** Identificación de Letras y Palabras α = .82 (Hulsey et al., 2010; West et al., 2008); α = .67 (Aikens et al., 2011);	n.a.
Woodcock-Muñoz Batería Revised and Third Edition (WM-R and WM-III; Dictado/ Ortografía	Administration for Children and Families, 2006 Hulsey et al. 2010 West et al., 2008 Zill et al. 2001 Zill et al. 2003	**WM-R** Dictado α = .75 to α = .83 (ACF, 2006; Zill et al., 2003) **WM-III** Ortografía α = .67 (Hulsey et al., 2010; West et al., 2008); α = .66 (Aikens et al., 2011)	n.a.

[a] Spanish monolingual sample.

APPENDIX D

MEASURE ACRONYMS

Table D.1. Measure Acronyms

Acronym	Full Name	Citation
ASQ-3	Ages and Stages Questionnaires–Third Edition	Squires, Twombly, Bricker, & Potter (2009)
BASC	Behavior Assessment System for Children	Reynolds & Kamphaus (1992)
BESA	Bilingual English Spanish Assessment	Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore (2009)
BESOS	Bilingual English Spanish Oral Language Screener	Peña, Bedore, Gutierrez-Clellen, Iglesias, & Goldstein (in preparation)
BFQ	Bilingual Language Proficiency Questionnaire	Modified version of Mattes & Santiagos (1985); in Guiberson, Barrett, Jancosek, & Itano (2006)
CASA-P	Comprehensive Assessment of Spanish Articulation–Phonology	Brice, Carson, & O'Brien (2008)
CDI	MacArthur-Bates Communicative Development Inventories	Fenson et al. (1993)
CTOPP	Comprehensive Test of Phonological Processing	Wagner, Torgesen, & Rashotte (1999)
ELP	Emergent Literacy Profile	Dickinson & Chaney (1997)
EOWPVT	Expressive One-Word Picture Vocabulary Test	Brownell (2000)
EOWPVT-SBE	Expressive One-Word Picture Vocabulary Test: Spanish Bilingual Edition	Brownell (2001)
EPAP	Early Phonological Awareness Profile	Dickinson & Chaney (1997)
GFTA-2	Goldman-Fristoe Test of Articulation 2	Goldman & Fristoe (2000)
GW	General Writing Task	Ferreiro & Teberosky (1982)
HLEQ	Home Literacy Environment Questionnaire	Hammer et al. (2003); Lonigan & Farver (2002); Payne, Whitehurst, & Angel (1994)
HLEQ	Home Language Environment Questionnaire	Parra, Hoff, & Core (2011); Marchman et al. (2004)
Inventario II	El Inventario del Desarrollo de Habilidades Comunicativas–II	Jackson-Maldonado et al. (1993, 2003)
K-BIT	Kaufman Brief Intelligence Test	Kaufman & Kaufman (1990)
KLPA-2	Khan-Lewis Phonological Analysis–Second Edition	Khan & Lewis (2002)
NW	Name-Writing Task	Ferreiro & Teberosky (1982)
PABA	Phonological and Articulatory Bilingual Assessment	Gildersleeve-Neumann (2010)
PCTOPPP	Preschool Comprehensive Test of Phonological and Print Processing	Lonigan et al. (2002)

Table D.1 *(continued)*

Acronym	Full Name	Citation
PLS-3	Preschool Language Scale 3	Zimmerman, Steiner, & Pond (1992)
PLS-4	Preschool Language Scale 4	Zimmerman, Steiner, & Pond (2002)
PMS	Parental Modernity Scale	Schaefer & Edgerton (1985)
PPVT-R	Peabody Picture Vocabulary Test-Revised	Dunn & Dunn (1981)
PPVT-III	Peabody Picture Vocabulary Test-III	Dunn & Dunn (1997)
PPVT-4	Peabody Picture Vocabulary Test-4	Dunn & Dunn (2007)
PreLAS; PreLAS 2000	Pre-Language Assessment Scale 2000	DeAvila & Duncan (2000)
PSI	Parenting Stress Index	Abidin (1995)
Real Word Reading Task	Real Word Reading Task	Durgunoglu et al. (1993)
ROWPVT	Receptive One-Word Picture Vocabulary Test	Brownell (2000)
ROWPVT-SBE	Receptive One-Word Picture Vocabulary Test: Spanish Bilingual Edition	Brownell (2001)
SEVC	Spanish-English Vocabulary Checklist	Patterson (1998)
SLAP	Spanish Language Assessment Procedure-Third Edition	Mattes (1995)
SPCTOPPP	Spanish Preschool Comprehensive Test of Phonological and Print Processing	Lonigan & Farver (2002)
TELD-3	Test of Early Language Development-3	Hresko, Reid & Hammill (1999)
TERA-2	Test of Early Reading Ability-2	Reid, Hresko, & Hammill (1991)
TRT	Title Recognition Task	Lonigan (2000)
TVIP	Test de Vocabulario en Imágenes Peabody	Dunn, Padilla, Lugo, & Dunn (1986)
VBS	Videotape Behavior Scale	Guiberson, Barrett, Jancosek, & Itano (2006)
WJ-III	Woodcock-Johnson Tests of Achievement	Mather & Woodcock (2001, 2007)
WLPB-R (English)	Woodcock Language Proficiency Battery-Revised, English Form	Woodcock (1995)
WLPB-R (Spanish)	Woodcock Language Proficiency Battery-Revised, Spanish Form	Woodcock & Muñoz-Sandoval (1995)
WM-III	Batería III Woodcock-Muñoz	Woodcock, Muñoz-Sandoval, McGrew, Mather, & Schrank (2004, 2007)

APPENDIX E

DEFINITIONS OF KEY DIMENSIONS OF VOCABULARY, LANGUAGE, AND LITERACY

Table E.1. Definitions of Key Dimensions of Vocabulary, Language, and Literacy

Term	Definition
Vocabulary Measure	Assesses knowledge of words and concepts
Receptive Vocabulary	A child's understanding of words spoken to him/her; usually assessed by having a child point to the picture that represents the word named by the assessor
Expressive Vocabulary	Ability to use words to name objects, actions, and concepts; usually assessed by having a child name a picture or representation of a concept (e.g., "What are the [OBJECTS] doing?" or "What is a name for all of these [OBJECTS]?")
Language Measure	Assesses broad knowledge of both expressive and receptive language
Literacy Measure	Assesses knowledge of reading and writing
Emergent Literacy	The early process of learning to read and write for children, beginning at birth and continuing until the child can read and write
Morphology	Knowledge of parts of a word that hold meaning (e.g., "s" on the end of a word to denote the plural form; "un" at the beginning of a word to denote an opposite action; base words)
Syntax	Knowledge of how the order of words in a sentence conveys meaning (e.g., "She put the dress on it" versus "She put it on the dress")
Morphosyntactic	Knowledge of both morphology and syntax
Semantics	The meaning of words in use (e.g., "The bride had a long *train*," "The man will *train* the dog," "The engineer stopped the *train*")
Phonology	Perception and knowledge of the sounds in words
Elision	The omission of one or more sounds (such as a vowel, a consonant, or a whole syllable) in a word or phrase
Blending	Forming words from parts of two or more other words. These parts are sometimes referred to as morphemes
Conceptually Scored	A method of scoring that gives credit for correct answers independent of the language in which the response was given
Principal Component Analysis	A mathematical model that examines the latent structure of the data that explains the variance

About CECER-DLL

CECER-DLL is a national center that is building capacity for research with dual language learners (DLLs) ages birth through five years. CECER-DLL aims to improve the state of knowledge and measurement in early childhood research on DLLs, identify and advance research on best practices for early care and education programming, and develop and disseminate products to improve research on DLLs. CECER-DLL is a cooperative agreement between the Frank Porter Graham (FPG) Child Development Institute at The University of North Carolina at Chapel Hill and the Office of Planning, Research, & Evaluation (OPRE) in the Administration for Children & Families (ACF), in collaboration with the Office of Head Start and the Office of Child Care.

UNC
FRANK PORTER GRAHAM
CHILD DEVELOPMENT INSTITUTE