DOI: 10.1111/infa.12571

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

Comparing apples to manzanas and oranges to naranjas: A new measure of English-Spanish vocabulary for dual language learners

Catherine S. Tamis-LeMonda[1] | George Kachergis[2] | Lillian R. Masek[1] | Sandy L. Gonzalez[1] | Kasey C. Soska[1] | Orit Herzberg[1] | Melody Xu[1] | Karen E. Adolph[1] | Rick O. Gilmore[3] | Marc H. Bornstein[4,5,6] | Marianella Casasola[7] | Caitlin M. Fausey[8] | Michael C. Frank[2] | Susan Goldin-Meadow[9] | Julie Gros-Louis[10] | Kathy Hirsh-Pasek[11] | Jana Iverson[12] | Casey Lew-Williams[13] | Brian MacWhinney[14] | Virginia A. Marchman[2] | Letitia Naigles[15] | Laura Namy[16] | Lynn K. Perry[17] | Meredith Rowe[18] | Adam Sheya[15] | Melanie Soderstrom[19] | Lulu Song[20] | Eric Walle[21] | Anne S. Warlaumont[22] | Hanako Yoshida[23] | Chen Yu[24] | Dan Yurovsky[14]

Correspondence
Catherine S. Tamis-LeMonda.
Email: catherine.tamis-lemonda@nyu.edu

Funding information
Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Number: R01HD094830; National Institute of Mental Health, Grant/Award Number: #T32 MH019524-28

Abstract

The valid assessment of vocabulary development in dual-language-learning infants is critical to developmental science. We developed the *Dual Language Learners English-Spanish (DLL-ES) Inventories* to measure vocabularies of U.S. English-Spanish DLLs. The inventories provide translation equivalents for all Spanish and English items on Communicative Development Inventory (CDI) short forms; extended inventories based on CDI long forms; and Spanish language-variety options. Item-Response Theory analyses applied to Wordbank and Web-CDI data (n = 2603, 12–18 months; n = 6722, 16–36 months; half female; 1% Asian, 3% Black, 2% Hispanic, 30% White,

The editor of this article is Gavin Bremner.

64% unknown) showed near-perfect associations between DLL-ES and CDI long-form scores. Interviews with 10 Hispanic mothers of 18- to 24-month-olds (2 White, 1 Black, 7 multi-racial; 6 female) provide a proof of concept for the value of the DLL-ES for assessing the vocabularies of DLLs.

1 | INTRODUCTION

Every healthy child learns language, but each follows a unique path in the rate and content of their vocabulary development. Thus, measures of individual differences in the words that infants understand and produce are critical to research and practice. Developmental researchers use a variety of approaches to assess early language development (Bornstein & Haynes, 1998; Pace et al., 2021). Looking time and eye-tracking studies test the words and grammar that infants and toddlers understand (Fernald et al., 1998; Golinkoff et al., 2013; Naigles, 2021). Audio and video recordings allow in-depth transcription and analysis of children's semantic, syntactic, and pragmatic skills (e.g., MacWhinney, 2019). And caregiver report—the focus here—yields easy to obtain, valid data on infants' language skills given caregivers' familiarity with what infants say and understand (e.g., Bates et al., 1988; Dale, 1991; Floccia et al., 2018). Indeed, since their creation more than 30 years ago, the MacArthur-Bates Communicative Development Inventories (CDIs) are widely used instruments for the reporting of infant and toddler language development, with adaptations available in multiple languages.

In the United States, Spanish is the most widely-used language after English. The large number of dual-language-learning (DLL) U.S. Hispanic children calls for instruments to assess vocabulary in English and Spanish (Hoff et al., 2012; Pearson & Fernández, 1994). However, the original CDI forms were created and normed on monolingual samples, with the goal of developing valid word lists for each language separately. Thus, they do not contain fully-matched items (i.e., translation equivalents) across languages. Examining children's receptive and productive language on a complete list of translation equivalents has advantages for addressing longstanding questions about DLLs' language development.

We aimed to expand opportunities to address core theoretical questions about the vocabulary development of Spanish-English DLLs through the creation of the Dual Language Learners English-Spanish (DLL-ES) Inventories. These inventories include fully-matched translation equivalents for all items on English and Spanish existing CDI short forms, offer additional matched items drawn from CDI long forms, incorporate options for different language varieties in Spanish-speaking populations, and show near-perfect alignment with estimated scores based on existing CDI long forms. The valid assessment of DLL children's early language skills has enormous practical significance: Children's early language skills predict later language development, which, in turn, cascades to social and academic skills in childhood and adolescence (Pace et al., 2019).

1.1 | Caregiver-report of child language

Caregiver report of child language became a formal scientific tradition in the 1970s. Researchers initially assessed children's vocabulary through open-ended interviews with a few caregivers (Bates

et al., 1975) and later administered structured checklists of common early-learned words to large samples of parents, who reported on the words that their children understood and produced (Bates et al., 1988). Structured checklists, in turn, evolved into the MacArthur-Bates CDIs, one of the most widely used language report instruments in developmental science (Frank et al., 2021).

After its publication, the CDI became a widely accepted measure for assessing infant vocabulary to address a variety of questions (e.g., Bauer et al., 2002; Fernald & Marchman, 2012; Hoff et al., 2012; Tsao et al., 2004; Walle & Campos, 2014; West et al., 2019). Widespread use of the CDI resulted in Wordbank, an open science repository for vocabulary data (Frank et al., 2021).

Multiple versions of the CDI are available for children of different ages. The CDI Words and Gestures form (8–18 months of age) assesses children's word comprehension and production of a list of words and includes a section on children's gestures. The Words and Sentences form (16–30 months) includes productive vocabulary items and a section on grammar. Both inventories come in short forms with fewer items. The CDI short forms were normed on a cross-sectional sample of 1379 English-speaking children (Fenson, et al., 2000), and the CDI long forms were normed on a cross-sectional sample of 4000+ children (Fenson et al., 2007).

Communicative Development Inventory short and long forms provide summary scores of the total number of words children produce (and understand at younger ages) and allow researchers to calculate age percentiles relative to the norming samples. Importantly, CDI forms demonstrate moderate to high concurrent, convergent, and predictive associations to other measures of children's expressive vocabulary and language comprehension, and distinguish between typically-developing children and children with language impairments and other disabilities (Dale, 1991; Friend & Keplinger, 2008; Miller et al., 1995; Ring & Fenson, 2000; Thal et al., 1999).

1.1.1 | Development and use of the Spanish Communicative Development Inventory

The CDI has been adapted for more than 100 languages (e.g., Spanish, American Sign Language, Korean) and language varieties (e.g., British and Australian English, Mexican and European Spanish) beyond the original U.S.-based American English inventories (e.g., Anderson & Reilly, 2002; Frank et al., 2021; Hamilton et al., 2000; Jackson-Maldonaldo et al., 1993; Kalashnikova et al., 2016; Pae, 2003). A major tenet behind the creation of CDI forms in languages other than English is the focus on adaptation, not translation. Consequently, CDI adaptations include culturally relevant words rather than translations of U.S. English words, with percentiles based on samples of monolingual children in target countries.

Spanish adaptations of the CDI have been available since the early 1990s. Mexican Spanish adaptations are the most widely used forms in the United States, are available in short and long forms, and have versions for younger and older infants. The Mexican Spanish long-form and short-form adaptations were developed and normed on 1872 and 3135 monolingual Spanish-speaking children in Mexico, respectively (Jackson-Maldonaldo et al., 1993, 2013).

1.1.2 | Use of Communicative Development Inventories in dual-language learners

Accurate estimates of DLL's vocabularies require administration of both language forms, because focus on only one language may underestimate vocabulary size by ignoring words that children

understand or produce in their other language (e.g., Hoff et al., 2012). Thus, researchers often use both English and Spanish versions of the CDI to estimate vocabulary in U.S. English-Spanish DLL infants and toddlers (e.g., Core et al., 2013; de Houwer et al., 2014; Mancilla-Martínez et al., 2011; Pearson & Fernández, 1994; Song et al., 2012). The combined use of long-form CDIs in two or more languages allows researchers to generate total vocabulary and conceptual vocabulary scores, with each offering unique information about children's language development (e.g., Core et al., 2013; Marchman & Martínez-Sussmann, 2002; Pearson & Fernández, 1994).

Total vocabulary scores sum the number of words children understand or produce across their languages, and can be used to estimate a child's placement relative to other children (e.g., Mayor & Plunkett, 2011). However, total vocabulary may overestimate DLLs' vocabulary size relative to that of monolingual children because caregivers report on more words from two non-matching forms than they would for either form alone. For example, a caregiver of a DLL child filling out both English and Spanish CDI shortforms may indicate that a child says "hat" which appears on the American English (but not Mexican Spanish) form and "falda (skirt)" which appears on the Mexican Spanish (but not American English) form, potentially crediting the child with two words. In contrast, a caregiver of a monolingual English-speaking child filling out just the English form could only indicate if the child said "hat". Similarly, a caregiver of a monolingual Spanish-speaking child could only indicate if the child said "falda". In both monolingual cases, the maximum credit would be one word. Moreover, lack of item matches prevents analysis of whether learning a word in one language increases the likelihood of learning a translation equivalent in the other (e.g., Bilson et al., 2015), as seen for phonologically-similar languages (Floccia et al., 2018).

In contrast, *conceptual vocabulary* credits a word based on a child's understanding or production of the word for the single concept whether spoken in one or both languages (e.g., *"dog"* and *"perro"* would qualify as one concept, as would either *"dog"* or *"perro"*). Thus, conceptual vocabulary scores may underestimate DLLs' vocabulary size because a child may have two words for a single concept, whereas a monolingual child has only one (Core et al., 2013; Thordardottir et al., 2006). Thus, fully-matched inventories would allow researchers to assess both total and conceptual vocabulary across languages and test factors that relate to both. For example, whether a child learns the word for a concept in two languages may be driven by the child's specific learning context rather than overall language skills or vocabulary.

1.2 | Limitations on use of the Communicative Development Inventory forms with dual-language learners

Despite the common practice of using both English and Spanish CDIs in research with DLL children, the inventories were originally designed for use with monolingual children. Thus, combining forms may pose problems when used with U.S. English-Spanish DLLs.

1.2.1 | Lack of item overlap

The goals in creating English and Spanish CDI short forms resulted in numerous non-overlapping items. Creators sought to identify a subset of 100 words from the long forms that correlated most strongly with total long form scores; discriminated among children at all ages and ability levels; and

balanced word representation across semantic categories (e.g., clothing words, food words). Percentiles for item-response frequencies were based on monolingual children, and word selection was not conducted with translation equivalence in mind. Consequently, English and Spanish short forms have minimal item overlap.

Incomplete matching across English and Spanish CDI short forms requires researchers to rely on a smaller subset of words when calculating conceptual vocabulary and to sum across different words for total vocabulary. Restrictions incurred by mismatched items make it difficult to determine the value of conceptual vocabulary or total vocabulary scores in DLLs' language development (Core et al., 2013; Pearson & Fernández, 1994; Thordardottir et al., 2006). In contrast, full item overlap allows researchers to calculate total and conceptual vocabulary scores and address questions about mutual exclusivity and phonological similarity (for example) by delving into characteristics associated with item overlap in the lexicons of bilinguals (e.g., Bialystok et al., 2010; Byers-Heinlein & Werker, 2013; DeAnda et al., 2016; Frank & Poulin-Dubois, 2002; Genesee, 2001; Houston-Price et al., 2010).

1.2.2 | Lack of attention to Spanish language varieties

Adaptations normed on Mexican children are the most widely used Spanish CDI forms in the United States (Jackson-Maldonaldo et al., 1993). However, each Spanish-speaking country has unique customs, traditions, and language varieties. Consequently, caregivers from different nationalities may not recognize, or use, some of the words on the Mexican Spanish CDI forms and leading to underreporting of the concepts that children understand or produce (Gonzalez & Nelson, 2018). Inclusivity across Spanish-English DLL populations calls for options for word variants. Wide availability of inclusive word sets offers greater transparency and reproducibility than modifications conducted by individual investigators. Although the CDI recommends to report acceptable cultural word variants, there are no consistent varieties across administration manuals. Including options directly on forms would improve administration consistency.

1.2.3 | Addressing the burden of administering two Communicative Development Inventory long forms

Researchers who wish to assess a large inventory of words may choose to administer CDI long forms in Spanish and English. However, this would require caregivers to report on an unwieldy number of items—824 words across English and Spanish CDI Words and Gestures forms, and 1360 words in English and Spanish CDI Words and Sentences forms. One way to reduce burden is to target words that already contain matches on the other form by focusing on concrete nouns and action verbs. For example, although caregivers may confidently report that their child understands "*bread*" and its Spanish translation equivalent "*pan*", they may be uncertain whether their child understands more abstract concepts such as "*yesterday*" and "*ayer*". The level of a word's concreteness relates to children's comprehension and production of the word (Bornstein et al., 2004; Braginsky et al., 2019; Hirsh-Pasek & Golinkoff, 1996), and concrete nouns and action verbs words dominate children's early vocabularies.

1.3 | Toward a new approach: Dual Language Learners English-Spanish inventories

We developed the Dual Language Learners English-Spanish (DLL-ES) Inventories to improve the assessment of DLL infants' receptive and expressive vocabularies. The DLL-ES Inventories contain fully matched items across English and Spanish forms and word options for Spanish-language varieties, allowing researchers to address core questions on the vocabulary development of Spanish-English bilingual infants and toddlers. Here, we sought to confirm the convergent validity and feasibility of the DLL-ES inventories. In Study 1, we documented the measurement properties of simulated DLL-ES scores relative to CDI vocabulary data shared on Wordbank (Frank et al., 2021) and Web-CDI (de Mayo et al., 2021). In Study 2, we report on a test case of mothers of U.S. Dual-language-learning infants to assess feasibility and administration time, and to illustrate the value of fully matched inventories for describing individual differences in children's Spanish and English vocabularies.

2 | STUDY 1: MEASUREMENT PROPERTIES

2.1 | Method

2.1.1 | Participants

Measurement properties of the DLL-ES Matched and Extended Inventories were evaluated through analysis of data on 2603 CDI Words and Gesture administrations of children at 12–18 months (1250 female, 1330 male) in English ($n = 2057$) and Spanish ($n = 546$) and 6722 CDI Words and Sentences administrations of children at 16–36 months (2598 female, 2709 male) in English ($n = 5576$) and Spanish ($n = 1146$). The English and Spanish administrations were independent (i.e., not from the same participants). Wordbank contains large amounts of missing data on participant demographics. Sex was unknown for 1.1% of the Words and Gestures sample and 25.6% of the Words and Sentences sample, all from English administrations. Race and ethnicity were not available for any of the Spanish administrations and most (54.6%) of the English; available demographics were 2.4% Asian, 8.3% Black, 5.2% Hispanic, 79.6% White, and 4.4% other, based on respondents to either CDI version. Education information was not available for 53.8% of the sample; for the remainder, mother's education was 21.8% high school or below, 52.8% completed some or all of college, and 24.9% some or all of a graduate degree.

2.1.2 | Creating the matched inventories

We created DLL-ES Matched Inventories from Spanish and English CDI short forms. Four adult English-Spanish bilinguals with 0–6 years of experience administering English and Spanish CDIs compared English and Spanish CDI Words and Gestures short forms (89 and 104 words, respectively) and CDI Words and Sentences short forms (100 words each). Two versions exist for English CDI Word and Sentences short forms (Fenson, et al., 2000), and so we used Version A. As shown in the top panel of Figure 1, percentages of overlapping items were small (23%–28% of words).

Researchers then identified translation equivalents for non-overlapping items to create two DLL-ES Matched Inventories—one based on the CDI Words and Gestures short forms (DLL-

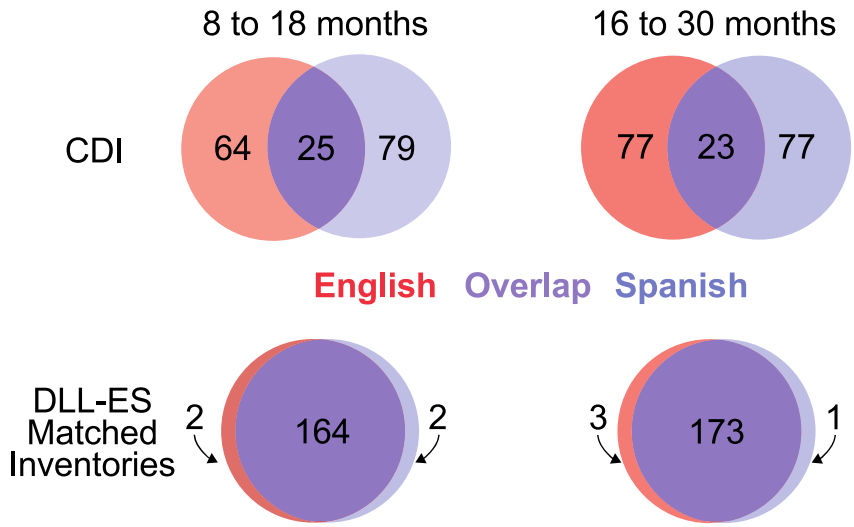

FIGURE 1 Overlap of words on Communicative Development Inventory (CDI) short forms and Dual Language Learners English-Spanish (DLL-ES) matched inventories. Top panel: Word overlap across the English and Spanish CDI Words and Gestures (8–18 months) and the CDI Words and Sentences (16–30 months) short forms. Bottom panel: Word overlap on the DLL-ES1 and DLL-ES2 Matched Inventories. Numbers in purple overlap sections indicate the number of translation equivalents, numbers in red and blue denote single words (e.g., unpaired). non-overlapping numbers (red and blue portions) in the bottom panel indicate loanwords.

ES1, 8–18 months) and one based on the CDI Words and Sentences short forms (DLL-ES2, 16–30 months). The DLL-ES1 and DLL-ES2 Matched Inventories contain translation equivalents for all non-overlapping CDI short-form items drawn from the other language forms (with the exception of 1–3 loanwords per form, e.g., "*tortilla*", "*uh oh*"). The DLL-ES1 Matched Inventory contains words from the original Spanish and English CDI Words and Gestures short forms, supplemented with Spanish translation equivalents or English translation equivalents for words that appear exclusively on the English short form or the Spanish short form, respectively. Similarly, the DLL-ES2 Matched Inventory includes words from the original Spanish and English CDI Words and Sentences short forms words, supplemented with Spanish translation equivalents for words that appear exclusively on the English short form and English translation equivalents for words that appear exclusively on the Spanish short form.

Communicative Development Inventory long forms served as a starting point for identifying translation equivalents. For example, "*dog*" appears on the English CDI Words and Sentences short form but not Spanish short form; however, its Spanish equivalent "*perro*" appears on the Spanish CDI Words and Sentences long form, and so "*perro*" was added to the DLL-ES2 Matched Inventory. When a translation equivalent did not appear on the long form of the other language, English-Spanish bilingual researchers discussed how to translate the word (e.g., "*toast*" was translated as "*tostada*"). The bottom panel of Figure 1 depicts overlap for items on the DLL-ES Matched Inventories. The resulting DLL-ES1 Matched Inventory includes 332 items (166 English, 166 Spanish), and the DLL-ES2 Matched Inventory includes 350 items (176 English, 174 Spanish). Both DLL-ES1 and DLL-ES2 Matched Inventories contain representative words form all semantic categories tested on the long-form CDI English and Spanish forms (see Supplementary Material C Tables S5-S8).

2.1.3 | Creating the extended inventories

Researchers interested in assessing a larger set of words might wish to augment DLL-ES Matched forms with additional translation equivalents. We took a multistep approach to expanding the list of words already included in the DLL-ES Matched Inventories without imposing undue burden on respondents (Figure 2). To create the DLL-ES Extended Inventories, we matched items across English and Spanish long forms and then winnowed down candidate words based on predetermined criteria (Words already on DLL-ES Matched forms were excluded from this process.)

First, we identified words on one long form that contained translation equivalents on the other long form. For example, *"train"* and *"tren"* appear on CDI long forms, but not on either short form, and so were considered for inclusion in the DLL-ES Extended Inventories. By using words that appear on the long forms, we were able to leverage the rich data on Word Bank to assess convergent validity for the DLL-ES in our analyses and ensure that selected words were developmentally appropriate for infants in the target age range in both languages. Most matches yielded a single translation equivalent for inclusion, but some words had multiple options (e.g., *"shirt"* could match with *"camisa"* or *"playera"* for the CDI Words and Sentences forms). In such instances, bilingual researchers determined which word to include by retrieving candidate items from Wordbank and selecting the closest matched word. Items that appeared on one long form only, with no translation equivalent on the other language form, were not included (e.g., "foca" appeared on the Spanish long form, but the translation equivalent of "seal" was not on the English long form, and therefore was excluded as an option). This matching

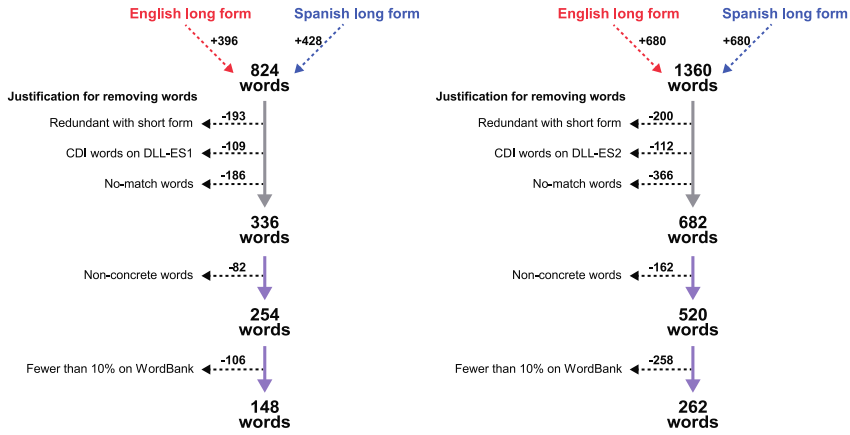

FIGURE 2 Decision flow chart for word inclusion on the Dual Language Learners English-Spanish (DLL-ES) extended inventories. Left panel: Process of word selection based on the English and Spanish Communicative Development Inventory (CDI) Words and Gestures long forms (8–18 months). Right panel: Process of word selection based on the English and Spanish CDI Words and Sentences long forms (16–30 months). The following justifications were used to exclude words from inclusion in the extended inventory. Redundant with short form means that the words included on the CDI long form were also included on the CDI short form, and are hence, represented in the DLL-ES Matched Inventories. CDI words on DLL-ES1/DLL-ES2 means that those words were already drawn from the CDI longform to provide translation equivalents for the CDI shortform and are hence, represented in the DLL-ES Matched Inventories. No-match words means that the word on the CDI longform in one language did not have a translation equivalent on the CDI longform in the other language. Non-concrete word means that the word fell into one of the excluded semantic categories (i.e., generic words, descriptive words such as adjectives, pronouns, question words, articles and prepositions, conjunctions, and words for people). Fewer than 10% on Wordbank means that for both the English and the Spanish, fewer than 10% of the sample in Word bank understood or produced the word for a concept.

process yielded 336 words as candidates for inclusion from the CDI Words and Gestures long form and 682 from the CDI Words and Sentences long form (Figure 2).

Because the DLL-ES Matched Inventories already included the full range of word classes from the CDI forms (e.g., nouns, verbs, prepositions, adjectives, etc), we only considered concrete nouns and action verbs for inclusion in the DLL-ES Extended Inventories. This process winnowed down candidates for inclusion to 254 words from the CDI Words and Gestures long form and 520 from the CDI Words and Sentences long form (Figure 2).

Finally, we pared down the resulting list of concrete nouns and action verbs to include words that 10% or more children understood or produced in English or Spanish. Percentiles were centered on data for 12-month-olds for the CDI Words and Gestures forms and 18-month-olds for the CDI Words and Sentences forms. This process resulted in 148 total words (74 English and 74 Spanish translation equivalents) from the CDI Words and Gestures forms; and 262 total words (131 English and 131 Spanish translation equivalents) from the CDI Words and Sentences forms to be used in conjunction with the DLL-ES Matched Inventories. The resulting DLL-ES1 Extended Inventory contains 480 items (240 English, 240 Spanish), and DLL-ES2 Extended Inventory includes 612 items (307 English, 305 Spanish), 42%–55% fewer words than the fully combined English and Spanish CDI long forms. Beyond the Extended Inventories, we created a supplementary list of words of math/spatial and technology concepts not included in existing CDIs (20 words with translation equivalents for English and Spanish); these added words expand options for abstract and contemporary concepts (e.g., circle, Google).

2.1.4 | Words from different language varieties

A team of four researchers from different Spanish-speaking nationalities (Chile, Venezuela, Cuba, Nicaragua, and Spain) reviewed all words in the DLL-ES Matched and Extended Inventories and discussed language variety options. This process yielded 26 items on the DLL-ES1 Inventories with possible words from different Spanish varieties (DLL-ES1 Matched Inventory: 8 items; DLL-ES1 Extended Inventory: 18 items) and 45 items on the DLL-ES2 Inventories (DLL-ES2 Matched Inventory: 14 items; DLL-ES2 Extended Inventory: 31 items). Sometimes, multiple substitutions were identified for a single word. For example, *"pastel"* (*"cake"*) has options of *"tarta", "torta"*, *"bizcocho",* and "*queque"*. In such cases, all substitutions were provided to cover a range of language varieties, with the original CDI word bolded on forms (e.g., Spanish options for the term "*blanket"*: **cobija**/manta/frazada/colcha).

2.1.5 | Analysis plan

After identifying candidate items for the DLL-ES Inventories, we used Item-Response Theory (IRT) analyses to test measurement properties (Embretson & Reise, 2013) based on CDI data from Wordbank (Frank et al., 2021) and Web-CDI (de Mayo et al., 2021). These psychometric analyses, explained below, assessed the convergent validity of the Matched and Extended Inventories for both English and Spanish and for each level/age group on the basis of a large database of long form CDI administrations.

Item-Response Theory overview

Item-Response Theory refers to a family of psychometric models that estimate, in their simplest form, the latent ability of individuals and the latent "difficulty" of individual items on an instrument, both assumed to be normally distributed. Thus, IRT allows researchers to test associations between a latent trait of unobservable abilities, such as vocabulary development and performance on an instrument—a language assessment like the DLL-ES (McCartney et al., 2006; Millsap, 2010). Item-Response Theory models are widely used in the creation of standardized tests and assessments in education and psychology, including CDI short forms.

Based on analyses reported elsewhere (Kachergis et al., 2022), 2-parameter logistic IRT models yielded information on the difficulty and discrimination of each Spanish and English word on the DLL-ES inventories. Our analyses compare the IRT-estimated difficulty and discrimination parameters for the translation equivalent Spanish and English "items" (i.e., words), much as psychometric comparisons are done across populations or tests (McCartney et al., 2006). To illustrate, the probability of a specific word being produced can be plotted against an overall "language ability score" using Item Characteristic Curves (ICC), with language ability reflecting a child's total vocabulary, for example, As depicted in the ICC of Figure 3a, difficult items (words) are shifted to the right (representing a later median age of onset), and easier items (words) to the left of the ability parameter (representing an earlier median age of onset). Item discrimination refers to an item's ability to distinguish among individuals who differ on the ability. For example, an item (word) with high discrimination will have a steeper slope than an item (word) with low discrimination (Figure 3b).

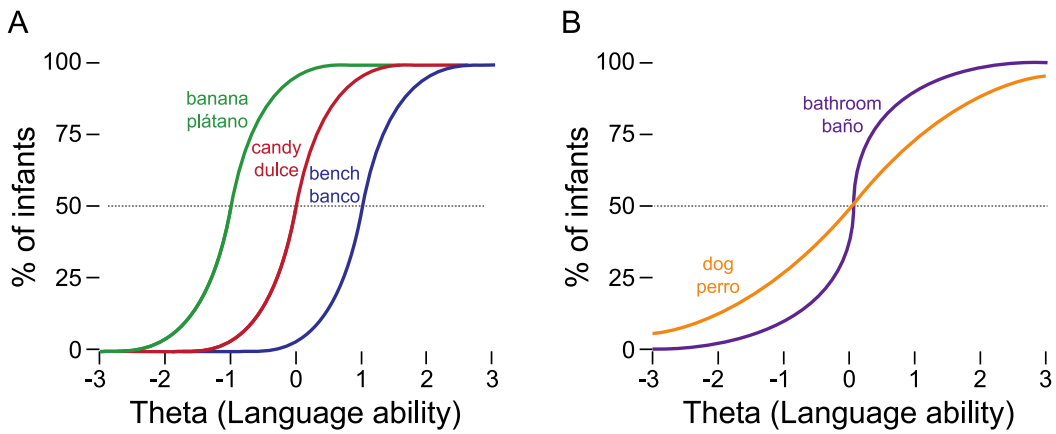

FIGURE 3 Hypothetical item characteristic curves (ICC) depicting item difficulty and discrimination based on Item-Response Theory (IRT) analyses: In an IRT framework, like a *z* score, the mean for the latent trait (theta) is set at 0 and the standard deviation at 1. (a) Difficulty parameter. Example ICC curves for items (words) that differ in difficulty (ICCs are illustrative rather than precise, but the difficulty ordering of example words is accurate). The *X*-axis represents the latent trait of language ability (theta) which is centered at 0. The *Y*-axis represents the percent of the population who passes an item (e.g., has a word in their vocabulary). As items shift to the right, difficulty increases, because only children with high ability (e.g., overall language skill) pass the item. (b) Discrimination parameter. ICC curve items (words) that differ in discrimination. (ICCs are illustrative, rather than precise, but discrimination differences between example words are accurate). As the slope of an item's characteristic curve increases, the item shows greater discrimination, because it distinguishes between children who can or cannot pass the item within a narrow window of ability.

Item-level analyses

To ensure comparability in English and Spanish forms, we examined the relative difficulty of words in each translation equivalent pair on the DLL-ES Inventories. To do so, we calculated the squared difference between the English and Spanish difficulty parameters of each item, represented as standardized z-scores. Higher values indicate greater mismatch on difficulty level within translation equivalent pairs. For example, "sombrero" is much more difficult (dSP = 2.14) than "hat" (dEN = −1.37), resulting in a large squared difficulty difference ($dSP - dEN)^2 = 12.33$). By aggregating across difficulty parameters of items, we can then compare the overall (average) difficulty of DLL-ES English forms to that of DLL-ES Spanish forms to assess equivalence in aggregate item difficulty between the two forms.

For the Matched Inventories, the median squared difference in difficulty level of English and Spanish words was small (i.e., English difficulty - Spanish difficulty = 0.99). Similarly, the DLL-ES2 Matched Inventory median squared difference was not excessive (i.e., English difficulty - Spanish difficulty = 1.36), although in both cases a small number of items showed large difficulty differences. In light of the comparability of English and Spanish word difficulty, and to further allow researchers to calculate percentile scores based on original CDI short-form norms, we chose to retain all words in the DLL-ES Matched forms.

However, for words in the DLL-ES Extended Inventories, we identified 15 pairs with the largest squared difference scores for replacement. New word pairs that matched the lexical class of the dropped pair (e.g., nouns, verbs) were drawn from previously omitted words on the CDI long forms to minimize differences. Replacement of the 15 mismatched word pairs reduced the average squared difference score on the DLL-ES1 Extended Inventory from 1.12 to 0.60, and from 1.26 to 0.88 for the DLL-ES2 Extended Inventory, reflecting minimal differences in difficulty. Moreover, the average difference in difficulty level of English and Spanish words (i.e., English difficulty - Spanish difficulty) approached zero (final value of 0.01 for the mean English - Spanish ease on the DLL-ES1). Similarly, the DLL-ES2 Extended Inventory final difference was minimal (i.e., English difficulty - Spanish difficulty = 0.05).

The complete list of DLL-ES1 and DLL-ES2 Extended Inventory word pairs, squared difference scores for items before replacement, and values for the 15 replacement pairs appear in Appendix S1 (Supplementary Material A Tables S1 and S2). We also provide information on difficulty parameters for all DLL-ES1 and DLL-ES2 Matched Inventory word pairs (with squared difference scores on item discrepancy), and a list of 13 word pairs on the DLL-ES1 and 10 word pairs on the DLL-ES2 that showed substantial divergence. Researchers have the option to exclude these items when calculating conceptual overlap (Appendix S2, Supplementary Material B Tables S1 and S2).

Content analysis

We ensured that words on the final DLL-ES Inventories (1) varied in difficulty (i.e., the full set of words in each language, based on monolingual data from Wordbank, showed a spread such that "less difficult" words are acquired at younger ages than "more difficult" words), (2) included frequent words in infants' early English and Spanish vocabularies, and (3) covered the range of parts of speech (e.g., nouns, verbs, prepositions, articles) and semantic categories (e.g., foods, clothing) on original CDI long forms. Importantly, focus on concrete nouns and action verbs for the DLL-ES Extended inventories did not compromise the content validity of words relative to CDI forms because Extended Inventories subsumed all words from the DLL-ES Matched Inventories that already contained these parts of speech (and even added to them when supplementary words were included). Therefore, resulting

forms yield similar word properties and distributions compared to CDI long forms (see Appendix S3, Supplemental Material C). Printable versions of the final DLL-ES Matched and Extended Inventories, sortable spreadsheets with all words, translations and provenance, along with lists of excluded words, are publicly available on OSF (https://osf.io/qx7h8/) and Databrary (https://nyu.databrary.org/volume/1168).

2.2 | Results study 1: Measurement properties

Using IRT analyses, we tested whether: (1) the DLL-ES Matched and Extended Inventories correspond to, underestimate, or overestimate vocabulary relative to scores on the CDI long forms, and (2) simulated scores on DLL-ES Short and Extended Inventories correlated with scores of individual children on CDI long forms.

Six separate models compared DLL-ES inventories to CDI long-form scores. Two of the models analyzed vocabulary comprehension for DLL-ES Matched and DLL-ES Extended Inventories for the younger age group only (i.e., DLL-ES1). The remaining four analyzed vocabulary production based on DLL-ES Matched Inventories and DLL-ES Extended Inventories at younger and older ages (i.e., DLL-ES1 and DLL-ES2). To further aid interpretation of findings, we conducted IRT models for estimates from original CDI short forms to CDI long form scores to compare with results from the DLL-ES Inventory models.

As shown in Figure 4, the DLL-ES inventories showed strong accuracy in predicting estimated scores from CDI long forms. Moreover, estimation accuracy of the DLL-ES Inventories to CDI long forms was as good as the estimation accuracy of CDI short forms to CDI long forms. Receptive and productive scores for English and Spanish DLL-ES1 Matched and Extended Inventories yielded only slight underestimation of CDI Words and Gestures long form scores. An exception was a small overestimation of English DLL-ES1 Matched and Extended Inventories at higher productive vocabulary scores on the CDI Words and Gesture long forms, as seen when blue estimated lines in Figure 4b rise above dashed lines. Similarly, the DLL-ES2 Matched and Extended Inventories slightly underestimated CDI Words and Sentences vocabulary scores in both English and Spanish, as seen when blue estimated lines drop below the dashed lines of Figures 4c and 4f. However, CDI short forms likewise yielded slight underestimations overall, and a small overestimation for production when compared to estimates of performance based on the CDI Words and Gestures long form (see Figure 4 right panels).

We next tested whether the subset of items on the DLL-ES1 and DLL-ES2 Matched and Extended Inventories in English and Spanish accurately predicted children's overall vocabulary size on CDI long forms (for production and comprehension where relevant). Simulated DLL-ES Inventory scores of children showed strong associations with CDI long-form scores on Wordbank (all rs = 0.98–0.99; Table 1). Thus, individual differences among children in scores generated by the DLL-ES Inventories show near-perfect alignment with scores generated by CDI long forms.

2.3 | Discussion study 1

The DLL-ES1 and DLL-ES2 Matched and Extended Inventories accurately estimate children's CDI long form scores, and estimates are as strong as existing CDI short forms. Moreover, the strong measurement properties of the DLL-ES Inventories come with the added benefit of fully matched items that diverge minimally in difficulty at the item level and overall.

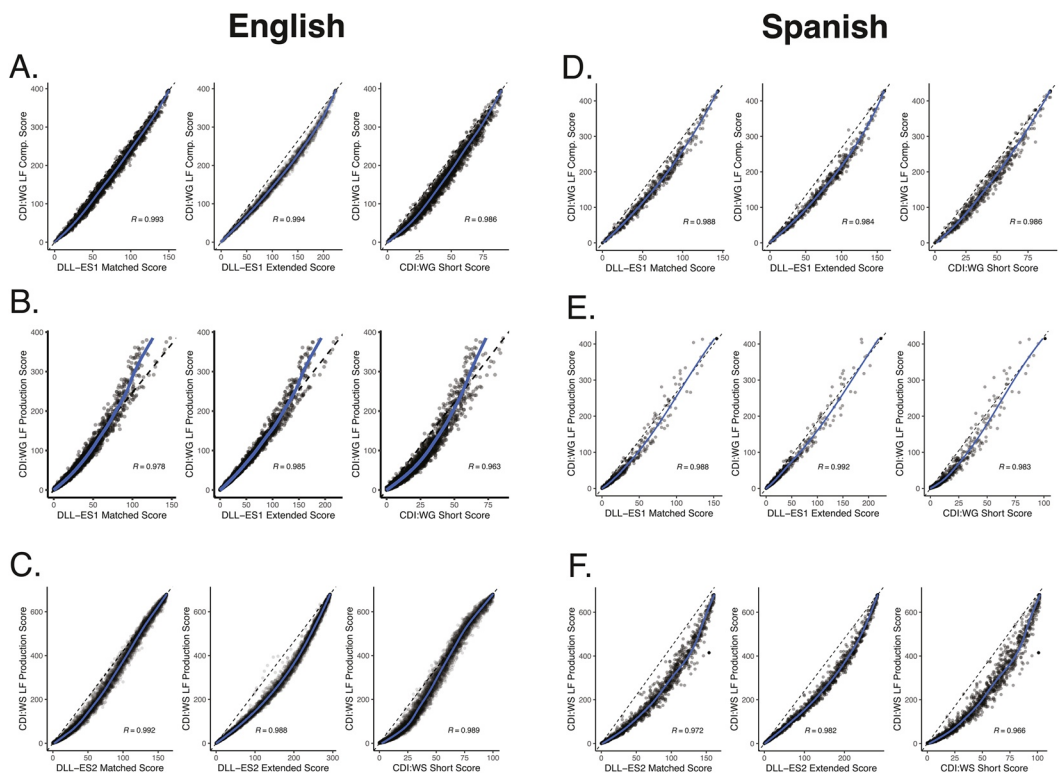

FIGURE 4 Item-Response Theory (IRT) estimations of Dual Language Learners English-Spanish (DLL-ES) inventory and Communicative Development Inventory (CDI) short form scores from CDI long form WordBank and Web-CDI data. (a) Estimated English comprehension scores for DLL-ES1 Matched, DLL-ES1 Extended, and CDI Words & Gestures short form. (b) Estimated English production scores for DLL-ES1 Matched, DLL-ES1 Extended, and CDI Words & Gestures short form. (c) Estimated English production scores for DLL-ES2 Matched, DLL-ES2 Extended, and CDI Words & Sentences short form. (d) Estimated Spanish comprehension scores for DLL-ES1 Matched, DLL-ES1 Extended, and CDI Words & Gestures short form. (e) Estimated Spanish production scores for DLL-ES1 Matched, DLL-ES1 Extended, and CDI Words & Gestures short form. (f) Estimated Spanish production scores for DLL-ES2 Matched, DLL-ES2 Extended, and CDI Words & Sentences short form.

TABLE 1 Correlations between Communicative Development Inventory (CDI) long form scores and estimated Dual Language Learners English-Spanish (DLL-ES) inventory scores.

Long form score	Language	*N*	*r* with CDI short form	*r* with DLL-ES matched inventory	*r* with DLL-ES extended inventory
CDI: Words & gestures comprehension	English	3696	0.99	0.99	0.99
CDI: Words & gestures production	English	3696	0.96	0.98	0.99
CDI: Words & sentences production	English	6411	0.99	0.99	0.99
CDI: Words & gestures comprehension	Spanish	731	0.99	0.99	0.99
CDI: Words & gestures production	Spanish	731	0.98	0.99	0.99
CDI: Words & sentences production	Spanish	1383	0.98	0.99	0.99

3 | STUDY 2: PROOF OF CONCEPT

We administered the DLL-ES Extended Inventory to mothers of DLL infants as a proof of concept to estimate administration time and generate descriptive data on infants' Spanish and English vocabularies. In this proof-of-concept sample, we sought to demonstrate the value of fully matched forms for quantifying individual differences among children in total vocabulary size, conceptual vocabulary size, and the distribution of words across the two languages. Additionally, we explored whether age of acquisition (based on Wordbank data) and phonological similarity (e.g., "sun" and "sol" vs. "dog" and "perro") affected the likelihood of a child having doublets of specific words. These exploratory analyses illustrate the type of questions that researchers can ask of larger samples when using the DLL-ES Inventories.

3.1 | Method

3.1.1 | Participants

Participants were 10 mothers of 18–24-month-old DLL infants (6 female) (M age = 21.3 months), recruited from an existing database, who were interviewed in person (n = 2) or remotely (n = 8). Mothers reported speaking to their infants in Spanish at home (with or without speaking English), and no other home language exposure. Infants were typically developing and full term. Mothers averaged 36.4 years (range = 29–47 years), with 9 self-identifying as Hispanic (1 White, 1 Black, 7 more than one race), and 1 identifying the father as Hispanic. Four mothers had been born outside the United States, and averaged U.S. residency from 12 to 36 years. Six mothers had graduate or professional degrees, and the remaining 4 had some college or bachelor's degree. Mothers received digital gift cards for their participation. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Institutional Review Board at New York University.

3.1.2 | Procedures

A fluent English-Spanish bilingual researcher interviewed mothers. The researcher gave examples of what was meant by "understanding a word" and/or "producing a word", and explained that the child's pronunciation of the word did not have to be standard (e.g., "raf" for giraffe). The researcher read each item on the DLL-ES Matched Forms, beginning with words from the original CDI short forms followed by the added Spanish and English translation equivalents, and starting with words in the child's dominant language. The researcher then asked about words on the DLL-ES Extended Inventories.

An illustrative video of a bilingual experimenter administering the DLL-ES is shared with authorized investigators on Databrary (https://nyu.databrary.org/volume/1168/slot/56100/). Videos from all sessions are shared on Databrary (https://nyu.databrary.org/volume/1442).

Calculation of age of acquisition

Age of acquisition was determined using monolingual Wordbank data (http://wordbank.stanford.edu) for the English American and the Spanish Mexican forms. If age of acquisition data were not available

for a word in the English American form or the Spanish Mexican form, age of acquisition was calculated based on other language varieties of English or Spanish. Age of acquisition score was the earliest age in months that at least 50% of the cross-sectional sample in Wordbank produced the word. If more than 50% of the sample produced the word at 16-months—the youngest age for the Words and Sentences form—or fewer than 50% of the sample produced the word at 30-months—the oldest age for the Words and Sentences form—an age of acquisition score of 16 or 30 was given, respectively (see Appendix S4 Supplementary Material Table D1 for age of acquisition in English and Spanish for all doublets).

Calculation of phonological similarity

Phonological similarity of doublets was scored using the Crosslinguistic Overlap Scale for Phonology (COSP; Kohnert et al., 2004), a common measure of phonological overlap between languages (e.g., Kambanaros et al., 2017; Lindgren & Bohnacker, 2020; Simpson Baird et al., 2016). The COSP assesses four phonological features: initial sound, number of syllables, consonants, and vowels. Each feature is rated on a scale of 0–3 (initial sound and consonant overlap) or 0–2 (number of syllables and vowel overlap), for a total score of 0–10. A score of zero indicates no phonological overlap between words (e.g., "bye" and "adios") and a score of 10 indicates complete phonological overlap (e.g., "photo" and "foto"). We retained 3 word pairs with complete phonological/semantic overlap that appeared on either the English or the Spanish CDI short forms, allowing researchers to generate percentile scores for each language.

Three researchers (two native English speakers and one native Spanish speaker) scored the doublets included in the DLL-ES. Disagreements were rare, and resolved through discussion (see Appendix S4 Supplementary Material Table D1 for COSP for all doublets).

3.2 | Results study 2: Illustrative cases

3.2.1 | Administration time

Administration of the DLL-ES indicated low burden. Administration of items on the original CDI short forms averaged 3.78 min (range = 1.28–7.78) for the English and 3.75 min (range = 1.35–5.22) for the Spanish forms. The DLL-ES supplemented translation equivalents added an average of 2.69 min (range = 0.74–4.73) for English words and 2.84 min (range = 1.43–3.92) for Spanish words. In total, the DLL-ES Matched Inventory averaged less than 10 min to administer. Items from the Extended Inventory added 7.79 min on average (range = 4.00–11.81). Thus, administration of the full DLL-ES Inventory (Matched and Extended) averaged less than 20 min. This time burden is substantially lower than administration of two long-form CDIs, which can take up to 60 min.

3.2.2 | Spanish and English vocabulary sizes and conceptual overlap

Even in this small test-case, individual differences were striking. Infants produced $M = 158.4$ ($SD = 88.8$) total words (i.e., crediting each doublet as two words) and $M = 124.5$ ($SD = 70.5$) total concepts (i.e., crediting each doublet as one concept), but varied on each. Infants' total words, ranged from 18 to 274 (Figure 5 top). Similarly, total concepts ranged from 15 to 219.

Infants also differed in the percentages of words they used in English and in Spanish (Figure 5 bottom). In terms of the distribution of words across the two languages, infants ranged from 23.4% to

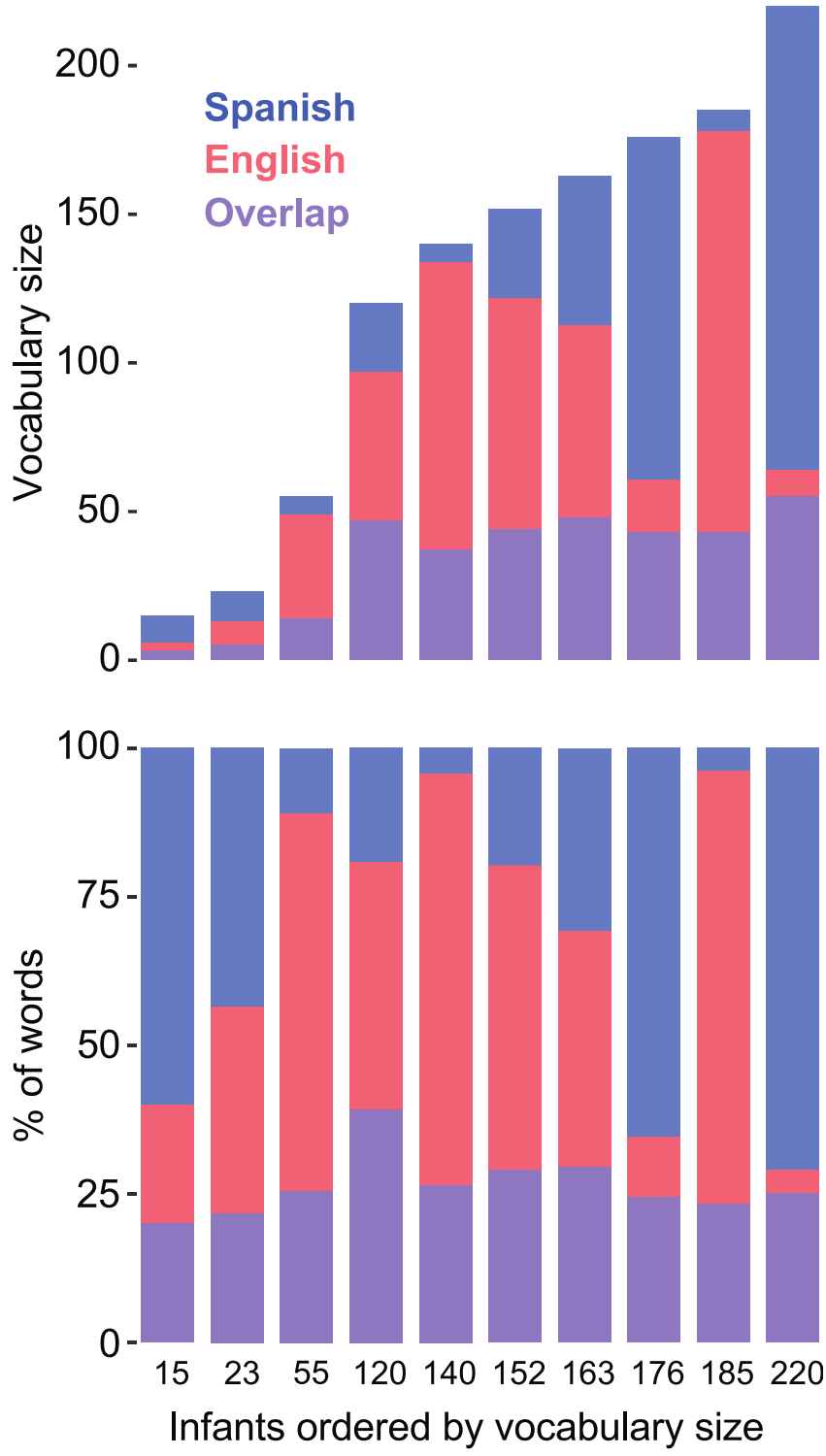

FIGURE 5 Vocabulary size for children in Study 2 whose mothers completed the DLL-ES2; each bar represents a child ordered from smallest to largest total vocabulary size. (Top) Number of words produced in English (red), Spanish (blue) and both languages (purple) scaled by total combined vocabulary size. (Bottom) Percentage of words produced in English, Spanish, and both languages, with children ordered from left to right by total vocabulary size.

78.0% of their words being in English (when using total vocabulary as the denominator). Similarly, infants ranged from 22.0% to 76.6% of their words being in Spanish.

Overlapping concepts, relative to the total number of concepts each infant produced, ranged from 20.0% to 39.5%. The substantial number of concepts for which children had two words (Figure 5) aligns with research showing that bilingual infants are more likely than monolingual infants to assign two labels to a single referent (e.g., Bosch & Ramon-Casas, 2014; Core et al., 2013; Mancilla-Martinez et al., 2011).

Interpreting contributors to overlap: Age of acquisition

Across the 301 concepts tested in both English and Spanish, 142 were produced in both English and Spanish by at least one child. Close inspection of the 142 doublets allowed us to examine associations between age of acquisition and phonological similarity on the concepts that children were most likely to produce in both languages. Of the 142 concepts produced by at least one child in both English and Spanish, 139 had age of acquisition data in English and 141 had age of acquisition data in Spanish. Of the other 159 words (produced by no child in both English and Spanish), 151 had age of acquisition data in English and 152 had age of acquisition data in Spanish.

Age of acquisition was associated with the concepts that infants produced in both languages. For doublets produced by at least one child, the average age of acquisition was 22.0 months for the English and 22.8 months for the Spanish, within the age range this study examined. In contrast, for doublets not produced by any children, the average age of acquisition was 25.9 months for the English and 26.0 months for the Spanish, above the age range this study examined. Indeed, age of acquisition predicted which words were produced in *both* languages. It was more likely that at least one infant produced the concept both languages when the age of acquisition was lower for the English word, $z = 7.49$, $p < 0.001$, $OR = 1.40$, and the Spanish word, $z = 6.41$, $p < 0.001$, $OR = 1.27$, in the doublet.

Interpreting contributors to overlap: Phonology

Crosslinguistic Overlap Scale for Phonology scores were overall low in the sample of words tested by the DLL-ES, suggesting limited phonological overlap between early-learned English and Spanish words. Moreover, phonological similarity between words did not relate to production of the concept in both languages by any child, ($z = 1.43$, $p = 0.153$). Nonetheless, the COSP score for doublets that were produced by any child was higher ($M = 2.90$, $SD = 2.69$) than the average COSP score for doublets not produced by any child ($M = 2.50$, $SD = 2.20$). That is, doublets with relatively lower phonological similarities were less likely to be produced by any child than those with higher similarity, although we did not have power to detect differences in this proof-of-concept case.

3.3 | Discussion study 2

We aimed to determine administration time of the DLL-ES and illustrate the type of information that matched inventories can yield. Notably, the DLL-ES was easy to administer: The Extended Inventory with a full set of translation equivalents averaged under 20 min. Descriptive data illustrate the value of the DLL-ES in yielding information on a full set of matched items. Even on this test case sample, the DLL-ES yielded substantial differences among infants in total vocabulary, conceptual vocabulary, and the distribution of words between Spanish and English, which aligns with the wide variability reported by others (e.g., Bosch & Ramon-Casas, 2014).

The words that children produced in both English and Spanish had earlier ages of acquisition (based on Wordbank) compared to the words that children produced in one language only. Furthermore,

doublets indicated low phonological overlap, and in this test-case sample, children were not more likely to acquire phonologically similar than dissimilar words. Findings based on large samples across multiple languages (that by definition represent a wider a range of phonological similarity), show an advantage for word learning for phonologically similar languages (e.g., Bosch & Ramon-Casas, 2014; Floccia et al., 2018; Sheng et al., 2016).

4 | GENERAL DISCUSSION

We created the DLL-ES Inventories to measure the vocabularies of English-Spanish DLL infants. Despite the wealth of information on the vocabulary development of DLL children, a lack of item overlap between CDI language forms limits analytic options and the questions that researchers can pursue. Leveraging CDI forms as a starting point, we created cross-language matched forms that allow for full concept matching in English and Spanish and include different language variety options for Spanish words. Item-Response Theory confirmed that the DLL-ES Inventories accurately estimate infants' vocabularies relative to estimated scores on CDI long forms, yield near-perfect associations with CDI estimated scores from English and Spanish CDI long-forms, and show minimal differences in word difficulty across languages. Findings on a test-case sample of mothers of 18- to 24-month-olds confirmed low time burden and revealed wide variation among children in their vocabulary sizes and the distributions of their words and concepts across Spanish and English.

4.1 | Uses of the Dual Language Learners English-Spanish inventories

The DLL-ES Inventories offer flexibility in use. Researchers can sort the openly shared spreadsheet of words to address their specific goals. They can opt to use Matched Inventories only (sorting by "Matched") or augment with words from Extended Inventories (adding on "Extended"). They can choose to present words for each concept in English and Spanish jointly, or sort by Spanish or English to begin with a child's dominant language. Dual Language Learners English-Spanish Matched Inventories allow researchers to calculate percentile scores in each language while providing information on 164 (DLL-ES1) and 173 (DLL-ES2) overlapping items (relative to 25 and 23 overlapping items in original forms). Dual Language Learners English-Spanish Extended Inventories likewise yield percentile scores, and data on a large set of over 300 matched items (concepts) from a variety of semantic categories. Researchers can calculate total vocabulary size, by summing the number of words a child understands or produces across languages, or conceptual vocabulary, by summing unique concepts.

The DLL-ES Extended Inventories allow researchers to test hypotheses about linguistic, social, and developmental influences on DLL language development based on a large set of words. Linguistically, researchers can test whether concepts that are phonologically similar (e.g., sun and sol) are more likely to appear in a child's Spanish *and* English vocabularies than words that are phonologically distinct (e.g., hat and sombrero). In terms of social influences, researchers can compare words on the DLL-ES Extended Inventories to the words that children and caregivers produce during interactions. Finally, researchers can assess children's vocabularies at different ages to model developmental change in how and when the lexicons of DLLs begin to separate, if at all (e.g., DeAnda et al., 2016).

4.2 | Challenges and limitations

We recognize that challenges come with developing tools to study DLL language development. First, language variety differences go beyond the words that we added, and we did not include options for

English words. However, parents and other caregivers can volunteer information on the child's production of regionalisms or translations from different language varieties that can be easily noted via a comment section. Researchers should be knowledgeable of the populations they sample, and language variety options should accommodate the heterogeneity of U.S. Spanish-speaking communities.

Second, researchers should ideally select a caregiver with knowledge of the child's vocabulary in each language (Marchman & Martínez-Sussmann, 2002), although this is not always possible. In such cases, researchers should rely on multiple informants who may each bring unique insight to children's early emerging vocabulary (de Houwer et al., 2014). Indeed, challenges to accurate assessment of the full range of words that infants understand and produce may be compounded for parents of DLLs, who must estimate vocabulary across two languages (Marchman et al., 2017; Potter et al., 2019).

Third, identifying translation equivalents is not straightforward. Matched inventories assume that a true translation equivalent exists for every word, which overlooks subtle differences in meaning among purportedly matched items. Some translation equivalents may have more conceptual overlap than others (e.g., "cat" and "gato" are more conceptually similar than "bread" and "pan") or may be more synonymous in one language variety than in another (e.g., how equivalent "bolsa" and "cartera" are to "bag"). Furthermore, some concepts have multiple possible translation equivalents (e.g., "there" in English has three translation equivalents in Spanish, "allá", "allí", and "ahí"). Lack of clarity in the words that constitute a concept can lead to overestimation or underestimation of DLLs' vocabulary.

Finally, we could have taken several approaches to development of the DLL-ES, such as by matching all items on the English and Spanish CDI long forms. Every approach has strengths and drawbacks. We elected to match words starting with the CDI short forms because such lists contain fewer items than long forms and allow researchers to generate percentile scores. Researchers interested in a large number of words can supplement Matched Forms with Extended Inventories at their discretion. Our approach, therefore, ensures flexibility for use across a variety of purposes.

4.3 | A living document

We view the DLL-ES Inventories as a living document. We encourage researchers to add new words from different language varieties based on caregiver feedback and to recommend updates to word lists as needed. For example, existing CDI forms do not capture many contemporary experiences of children (e.g., "Alexa" or "Google" are likely known by more children than CDI words of "radio" and "cassette"). We therefore compiled openly available lists of math and technology-based words for appended use with the DLL-ES Inventories to capture abstract spatial, numerical, and magnitude words that are generally acquired in the third year of life, and words for commonly used technologies. Researchers can access additional word lists on OSF (https://osf.io/qx7h8/) and Databrary (https://nyu.databrary.org/volume/1168).

Researchers can submit suggestions at https://forms.gle/TS4z6Au299D2JUxv6. Continual updates will be made to DLL-ES Inventories through vetting of new language-variety options by a subset of the 32 authors of this manuscript who created the inventories. Updates to the DLL-ES Inventories will be conducted with full transparency and include version numbers and dates for tracking across studies so that researchers can determine which version(s) to use (e.g., longitudinal research may best be served by consistent versions).

The power of norms is well illustrated by Wordbank (Fenson et al., 2007; Frank et al., 2021), which includes openly shared CDI vocabulary data from 65,000+ children across several languages.

Researchers can compare data across languages, investigate item trajectories across ages, and so on. However, a delicate balance exists between standardized assessment and flexibility in adjusting vocabulary items to align with change. We acknowledge the tension between establishing and retaining norms versus creating up-to-date vocabulary lists to capture the words that children understand and produce.

4.4 | Future directions

We aimed to develop the DLL-ES Inventories and test their convergent validity with existing CDI English and Spanish long forms. Future work might examine convergent validity for the same cohort of children; examine predictive validity between the DLL-ES Inventories and production data on repositories (e.g., TalkBank or HomeBank); and test predictive validity with other measures of language or cognitive development. Although our proof-of-concept study illustrated the feasibility and potential use cases of the DLL-ES Inventories, the full benefit of the DLL-ES will only be realized through work with large samples—ideally infants from Spanish-speaking families from across the United States. Use of the DLL-ES in diverse samples promises to advance an understanding of the characteristics and correlates of individual differences in DLLs total and conceptual vocabularies, the ages of acquisition of specific words and word classes, and other issues central to bilingual language development.

Finally, our open sharing of inventories, psychometric data, analytic code, and videos on test cases allows researchers to create matched inventories in other languages (See Abdelwahab et al., 2021). Researchers can conduct IRT analyses on the shared psychometric data and phonological similarity data to create short forms with fully matched items. Indeed, reliable language data can be obtained with as few as 25–30 words (e.g., Floccia et al., 2018; Mayor & Mani, 2019). Vocabulary data based on fully matched forms provides the ideal starting point for constructing short forms and extending inventories to other languages.

5 | CONCLUSION

Caregiver report is a valuable tool for studying infant language development. We capitalized on the rich data generated by CDI forms to construct instruments to quantify the vocabularies of Spanish-English DLLs in the United States. As U.S. demographics continue to shift, the need for inclusive measures will continue to grow. We encourage researchers to adopt measures that contain full translation equivalents when studying the words that young DLL children know and say.

AFFILIATIONS

[1]New York University, New York, New York, USA

[2]Stanford University, Stanford, California, USA

[3]The Pennsylvania State University, University Park, Pennsylvania, USA

[4]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Bethesda, Maryland, USA

[5]Institute for Fiscal Studies, London, UK

[6]UNICEF, New York, New York, USA

[7]Cornell University, Ithaca, New York, USA

[8]University of Oregon, Eugene, Oregon, USA

[9]University of Chicago, Chicago, Illinois, USA
[10]University of Iowa, Iowa City, Iowa, USA
[11]Temple University, Philadelphia, Pennsylvania, USA
[12]Boston University, Boston, Massachusetts, USA
[13]Princeton University, Princeton, New Jersey, USA
[14]Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[15]University of Connecticut, Storrs, Connecticut, USA
[16]Institute for Education Sciences, Washington, District of Columbia, USA
[17]University of Miami, Coral Gables, Florida, USA
[18]Harvard University, Cambridge, Massachusetts, USA
[19]University of Manitoba, Winnipeg, Manitoba, Canada
[20]Brooklyn College, CUNY, Brooklyn, New York, USA
[21]University of California Merced, Merced, California, USA
[22]University of California Los Angeles, Los Angeles, California, USA
[23]University of Houston, Houston, Texas, USA
[24]University of Texas at Austin, Austin, Texas, USA

ACKNOWLEDGMENTS

This work was supported by a grant from NICHD #R01 HD094830-01 to Karen E. Adolph, Catherine S. Tamis-LeMonda, & Rick O. Gilmore, and a training grant from NIMH #T32 MH019524-28 to Sandy L. Gonzalez. We thank Alexandra Mendelsohn for data collection, Jennifer Rachwani and Fernanda Robelo for feedback on Spanish translations, and Mark Blumberg for feedback on figures. The authors declare no conflicts of interest with regard to the funding source for this study. Correspondence concerning this article should be addressed to Catherine S. Tamis-LeMonda, Department of Applied Psychology, New York University, 246 Greene Street, Room 408W, New York, NY 10003. Electronic mail may be sent to catherine.tamis-lemonda@nyu.edu.

ORCID

Catherine S. Tamis-LeMonda https://orcid.org/0000-0002-9071-4028
Lillian R. Masek https://orcid.org/0000-0003-0448-3671
Kasey C. Soska https://orcid.org/0000-0002-8257-0511
Karen E. Adolph https://orcid.org/0000-0003-2819-134X
Marc H. Bornstein https://orcid.org/0000-0002-6810-8427
Caitlin M. Fausey https://orcid.org/0000-0003-1857-104X
Virginia A. Marchman https://orcid.org/0000-0001-7183-6743

REFERENCES

Abdelwahab, A. G. S., Forbes, S., Cattani, A., Goslin, J., & Floccia, C. (2021). An adaptation of the MacArthur-Bates CDI in 17 Arabic dialects for children aged 8 to 30 months. *Language Learning and Development*, *17*(4), 425–446. https://doi.org/10.1080/15475441.2021.1916502

Anderson, D., & Reilly, J. (2002). The MacArthur communicative development inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, *7*(2), 83–106. https://doi.org/10.1093/deafed/7.2.83

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*, 205–226.

Bates, E., Snyder, L., & Bretherton, I. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press.

Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied PsychoLinguistics*, *23*(3), 313–335. https://doi.org/10.1017/S0142716402003016

Bialystok, E., Luk, G., Peets, K., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134. https://doi.org/10.1016/j.cognition.2015.03.013

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children. *Child Development*, *75*(4), 1115–1139. https://doi.org/10.1111/j.1467-8624.2004.00729.x

Bornstein, M. H., & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development*, *69*(3), 654–671. https://doi.org/10.1111/j.1467-8624.1998.tb06235.x

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322. https://doi.org/10.1177/0165025414532559

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, *128*(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 Months: Implications for assessment. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1637–1649. https://doi.org/10.1044/1092-4388(2013/11-0044)

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech & Hearing Research*, *34*(3), 565–571. https://doi.org/10.1044/jshr.3403.565

DeAnda, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). Lexical processing and organization in bilingual first language acquisition: Guiding future research. *Psychological Bulletin*, *142*(6), 655–667. https://doi.org/10.1037/bul0000042

de Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied PsychoLinguistics*, *35*(6), 1189–1211. https://doi.org/10.1017/S0142716412000744

de Mayo, B., Kellier, D., Braginsky, M., Bergmann, C., Hendriks, C., Rowland, C. F., Frank, M. C., & Marchman, V. A. (2021). *Web-CDI: A system for online administration of the MacArthur-Bates communicative development inventories*. Language Development Research. https://doi.org/10.34758/kr8e-w591

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P., Reznick, J. S., & Bates, E. (2007). *MacArthur-bates communicative development inventories*. Paul H. Brooks.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied PsychoLinguistics*, *21*(1), 95–116. https://doi.org/10.1017/S0142716400001053

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Floccia, C., Sambrook, T., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*, 1–135. https://doi.org/10.1111/mono.12352

Frank, I., & Poulin-Dubois, D. (2002). Young monolingual and bilingual children's responses to violation of the mutual exclusivity principle. *International Journal of Bilingualism*, *6*(2), 125–146. https://doi.org/10.1177/13670069020060020201

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The Wordbank Project*. MIT Press.

Friend, M., & Keplinger, M. (2008). Reliability and validity of the computerized comprehension task (CCT): Data from American English and Mexican Spanish infants. *Journal of Child Language*, *35*(1), 77–98. https://doi.org/10.1017/s0305000907008264

Genesee, F. (2001). Bilingual first language acquisition: Exploring the limits of the language faculty. *Annual Review of Applied Linguistics*, *21*, 153–168. https://doi.org/10.1017/S0267190501000095

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Gonzalez, S. L., & Nelson, E. L. (2018). Measuring Spanish comprehension in infants from mixed hispanic communities using the IDHC: A preliminary study on 16-month-olds. *Behavioral Sciences*, *8*(12), 117–129. https://doi.org/10.3390/bs8120117

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(3), 689–705. https://doi.org/10.1017/s0305000900004414

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The Origins of grammar: Evidence from comprehension*. MIT Press.

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*, 1–27. https://doi.org/10.1017/S0305000910000759

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, *15*(2), 125–150. https://doi.org/10.1111/j.1532-7078.2009.00009.x

Jackson-Maldonaldo, D., Marchman, V. A., Fernald, A., & Lia, C. H. (2013). Short-form versions of the Spanish MacArthur-Bates communicative development inventories. *Applied PsychoLinguistics*, *34*(4), 837–868. https://doi.org/10.1017/S0142716412000045

Jackson-Maldonaldo, D., Thal, D. J., Marchman, V. A., Bates, E., & Gurierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*(3), 523–549. https://doi.org/10.1017/s0305000900008461

Kachergis, G., Marchman, V. A., Dale, P. S., Mankewitz, J., & Frank, M. C. (2022). Online computerized adaptive tests of children's vocabulary development in English and Mexican Spanish. *Journal of Speech, Language, and Hearing Research*, *65*(6), 2288–2308. https://doi.org/10.31234/osf.io/5ftsu

Kalashnikova, M., Schwarz, I.-C., & Burnham, D. (2016). Ozi: Australian English communicative development inventory. *First Language*, *36*(4), 407–427. https://doi.org/10.1177/0142723716648846

Kambanaros, M., Michaelides, M., & Grohmann, K. K. (2017). Cross-linguistic transfer effects after phonologically based cognate therapy in a case of multilingual specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *52*(3), 270–284. https://doi.org/10.1111/1460-6984.12270

Kohnert, K., Windsor, J., & Miller, R. (2004). Crossing borders: Recognition of Spanish words by English-speaking children with and without language impairment. *Applied PsychoLinguistics*, *25*(4), 543–564. https://doi.org/10.1017/S0142716404001262

Lindgren, J., & Bohnacker, U. (2020). Vocabulary development in closely related languages: Age word type and cognate facilitation effects in bilingual Swedish-German preschool children. *Linguistic Approaches to Bilingualism*, *10*(5), 587–622. https://doi.org/10.1075/lab.18041.lin

MacWhinney, B. (2019). Understanding spoken language through TalkBank. *Behavior Research Methods*, *51*(4), 1919–1927. https://doi.org/10.3758/s13428-018-1174-9

Mancilla-Martinez, J., Pan, B. A., & Vagh, S. B. (2011). Assessing the productive vocabulary of Spanish–English bilingual toddlers from low-income families. *Applied PsychoLinguistics*, *32*(2), 333–357. https://doi.org/10.1017/S0142716410000433

Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, *20*(1), e12425. https://doi.org/10.1111/desc.12425

Marchman, V. A., & Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research*, *45*, 938–997. https://doi.org/10.1044/1092-4388(2002/080)

Mayor, J., & Mani, N. (2019). A short version of the MacArthur–Bates Communicative Development Inventories with high validity. *Behavior Research Methods*, *51*(5), 2248–2255. https://doi.org/10.3758/s13428-018-1146-0

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769–785. https://doi.org/10.1111/j.1467-7687.2010.01024.x

McCartney, K., Burchinal, M. R., & Bub, K. L. (2006). Best practices in quantitative methods for developmentalists. *Monographs of the Society for Research in Child Development*, *71*, 1–145. https://doi.org/10.1111/j.1540-5834.2006.07103001.x

Miller, J. F., Sedey, A. L., & Miolo, G. (1995). Validity of parent report measures of vocabulary development of children with down syndrome. *Journal of Speech, Language, and Hearing Research*, *38*(5), 1037–1044. https://doi.org/10.1044/jshr.3805.1037

Millsap, R. E. (2010). Testing measurement invariance using item response theory in longitudinal data: An introduction. *Child Development Perspectives*, *4*(1), 5–9. https://doi.org/10.1111/j.1750-8606.2009.00109.x

Naigles, L. (2021). It takes all kinds (of information) to learn a language: Investigating the language comprehension of typical children and children with autism. *Current Directions in Psychological Science*, *30*(1), 11–18. https://doi.org/10.1177/0963721420969404

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

Pace, A., Luo, R., Levine, D., Iglesias, A., de Villiers, J., Golinkoff, R. M., Wilson, M., & Hirsh-Pasek, K. (2021). Competencies within and across languages in dual language learners: The role of exposure and process. *Child Development*, *92*(1), 35–53. https://doi.org/10.1111/cdev.13418

Pae, S. Y. (2003). Validity and reliability of the Korean adaptation of MCDI. *Communication Sciences and Disorders*, *8*, 1–14. https://doi.org/10.1016/j.ijom.2015.11.006

Pearson, B. Z., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, *44*(4), 617–653. https://doi.org/10.1111/j.1467-1770.1994.tb00633.x

Potter, C. E., Fourakis, E., Morin-Lessard, E., Byers-Heinlein, K., & Lew-Williams, C. (2019). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, *22*(4), e12794. https://doi.org/10.1111/desc.12794

Ring, E. D., & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language*, *20*(59), 141–159. https://doi.org/10.1177/014272370002005902

Sheng, L., Lam, B. P. W., Cruz, D., & Fulton, A. (2016). A robust demonstration of the cognate facilitation effect in first-language and second-language naming. *Journal of Experimental Child Psychology*, *141*, 229–238. https://doi.org/10.1016/j.jecp.2015.09.007

Simpson Baird, A., Palacios, N., & Kibler, A. (2016). The cognate and false cognate knowledge of young emergent bilinguals. *Language Learning*, *66*(2), 448–470. https://doi.org/10.1111/lang.12160

Song, L., Tamis-LeMonda, C. S., Yoshikawa, H., Kahana-Kalman, R., & Wu, I. (2012). Language experiences and vocabulary development in Dominican and Mexican infants across the first 2 years. *Developmental Psychology*, *48*(4), 1106–1123. https://doi.org/10.1037/a0026401

Thal, D. J., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research*, *42*(2), 482–496. https://doi.org/10.1044/jslhr.4202.482

Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, *4*, 1–21. https://doi.org/10.1080/14769670500215647

Tsao, F.-M., Liu, H.-M., & Kuhl, P. M. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*(4), 1067–1084. https://doi.org/10.1111/j.1467-8624.2004.00726.x

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Developmental Psychology*, *50*(2), 336–348. https://doi.org/10.1037/a0033238

West, K. L., Leezebaum, N. B., Northrup, J. B., & Iverson, J. M. (2019). The relation between walking and language in infant siblings of children with autism spectrum disorder. *Child Development*, *90*(3), 356–372. https://doi.org/10.1111/cdev.12980

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tamis-LeMonda, C. S., Kachergis, G., Masek, L. R., Gonzalez, S. L., Soska, K. C., Herzberg, O., Xu, M., Adolph, K. E., Gilmore, R. O., Bornstein, M. H., Casasola, M., Fausey, C. M., Frank, M. C., Goldin-Meadow, S., Gros-Louis, J., Hirsh-Pasek, K., Iverson, J., Lew-Williams, C., MacWhinney, B., … Yurovsky, D. (2024). Comparing apples to manzanas and oranges to naranjas: A new measure of English-Spanish vocabulary for dual language learners. *Infancy*, *29*(3), 302–326. https://doi.org/10.1111/infa.12571