

Research Article

Psychometric Properties of the English–Spanish Vocabulary Inventory in Toddlers With and Without Early Language Delay

Stephanie De Anda,^a  Lauren M. Cycyk,^a  Heather Moore,^b Lidia Huerta,^a Anne L. Larson,^c 
and Marika King^d 

^aUniversity of Oregon, Eugene ^bWestern Washington University, Bellingham ^cUniversity of Minnesota, Twin Cities ^dUtah State University, Logan

ARTICLE INFO

Article History:

Received April 29, 2021

Revision received July 29, 2021

Accepted September 22, 2021

Editor-in-Chief: Stephen M. Camarata

Editor: Megan York Roberts

https://doi.org/10.1044/2021_JSLHR-21-00240

ABSTRACT

Purpose: Despite the increasing population of dual language learners (DLLs) in the United States, vocabulary measures for young DLLs have largely relied on instruments developed for monolinguals. The multistudy project reports on the psychometric properties of the English–Spanish Vocabulary Inventory (ESVI), which was designed to capture unique cross-language measures of lexical knowledge that are critical for assessing DLLs' vocabulary, including translation equivalents (whether the child knows the words for the same concept in each language), total vocabulary (the number of words known across both languages), and conceptual vocabulary (the number of words known that represent unique concepts in either language).

Method: Three studies included 87 Spanish–English DLLs ($M_{\text{age}} = 26.58$ months, $SD = 2.86$ months) with and without language delay from two geographic regions. Multiple measures (e.g., caregiver report, observation, behavioral tasks, and standardized assessments) determined content validity, construct validity, social validity, and criterion validity of the ESVI.

Results: Monolingual instruments used in bilingual contexts significantly undercounted lexical knowledge as measured on the ESVI. Scores on the ESVI were related to performance on other measures of communication, indicating acceptable content, construct, and criterion validity. Social validity ratings were similarly positive. ESVI scores were also associated with suspected language delay.

Conclusions: These studies provide initial evidence of the adequacy of the ESVI for use in research and clinical contexts with young children learning English and Spanish (with or without a language delay). Developing tools such as the ESVI promotes culturally and linguistically responsive practices that support accurate assessment of DLLs' lexical development.

Supplemental Material: <https://doi.org/10.23641/asha.17704391>

Examining early language acquisition necessitates a focus on dual language development. Such a focus is critical, given the remarkable rise of childhood bilingualism in the North America. In the United States specifically, approximately 32% of children (0–8 years of age) have at least one parent who speaks a language other than English.

These children are often referred to as dual language learners (DLLs) because they are learning languages other than English at home while also learning English (U.S. Department of Health and Human Services & U.S. Department of Education). Within DLL households, Spanish is the most common home language spoken by parents (59% of DLL parents report speaking Spanish at home; Park et al., 2017). Given the significant representation of this group of children, appropriate methods to assess early language are crucial to understand their developmental trajectories in both languages and to identify early language disability.

Correspondence to Stephanie De Anda: sdeanda@uoregon.edu. **Disclosure:** The authors have declared that no competing financial or non-financial interests existed at the time of publication.

Vocabulary is a central measurement construct in early language assessment. Typically developing children (whether learning in a mono- or multilingual context) produce their first word by the end of their first year of life, and they spend the next several months and years growing their word knowledge. Leading theories of language and literacy outcomes posit that vocabulary knowledge in toddlerhood and the preschool period sets the foundation for future development of language and literacy (e.g., Duff et al., 2015; Jin et al., 2020). In one study, vocabulary size at 2 years of age significantly predicted language and literacy achievement into fifth grade (Lee, 2011). Vocabulary size was a stronger predictor of school-age language outcomes than was lexical composition (i.e., number of verbs and closed-class words), even after controlling for child sex, birth order, ethnicity, and socioeconomic status. In addition, in clinical contexts, best practice dictates measuring vocabulary to identify early language delays and disorders (e.g., De Anda et al., 2020). Low vocabulary skills are a widely used marker of early language delay (Ellis & Thal, 2008; Weismer & Evans, 2002).

Despite the importance of measuring individual differences in early vocabulary acquisition, the available instruments that have been validated and used in the field are insufficient for children learning more than one language. Bilingualism is the norm around the world and on the rise in the United States, yet most vocabulary measures are solely available for single language learners. Indeed, available parent report measures are exclusively focused on measuring only one language, and norming populations have little representation of children who are not White (Larson, 2016). Therefore, the purpose of this article is to introduce a new bilingual parent report measure for capturing the vocabulary of toddlers with and without early language delays who are learning Spanish and English. We present a series of studies examining several measurement properties of this adapted measure in an effort to contribute a valid and reliable assessment method that clinicians and researchers can use. Although the present focus is on Spanish and English DLLs, we outline the procedure with which this bilingual tool was developed so that future researchers and practitioners can adapt the approach to support early vocabulary assessment in different multilingual contexts.

Parent Report as a Method of Vocabulary Assessment

Clinicians have developed several parent report assessment methods to examine early language development. Although there are inherent limitations to parent report, several studies show that parents from a range of educational, cultural, and linguistic backgrounds are reliable reporters of their child's development (Restrepo, 1998; Paradis et al., 2010). In the context of infants and toddlers, the use of parent

report is widespread as it provides many advantages. Parent report is more efficient than behavioral observation, which is critical because time allotted for assessment is often limited in early intervention. Similarly, assessments that require eliciting responses in young infants and toddlers can be difficult given children's emerging attentional skills and unfamiliarity with testing routines. In such cases, parent report provides a more reliable method for capturing information about the child's vocabulary knowledge than behavioral data. In addition, family involvement in the assessment process is a recommended practice for early intervention service delivery. Parent report can also help overcome challenges associated with limited linguistic diversity in the practitioner workforce as the vast majority of speech-language pathologists (SLPs) are monolingual English speakers (American Speech-Language-Hearing Association, 2021). For example, in such cases of language mismatch between the clinician and the family, parents may report on all of the child's languages. This does not negate the use of interpreters, however, who can help translate written or verbal instructions to parents and to support child language elicitation in assessment (e.g., Langdon & Saenz, 2015). Indeed, accurate assessment of all of the child's languages is imperative for differential diagnosis (Mancilla-Martinez et al., 2016). In addition, parent report is flexible enough to be collected over the phone, in person, or independently through a paper or electronic questionnaire. This flexibility and efficiency in parent report methods has contributed to a large and growing research base that documents cross-linguistic developmental trajectories for vocabulary acquisition across many of the world's languages (e.g., Frank et al., 2017).

At present, the MacArthur-Bates Communicative Development Inventories (CDIs; Fenson et al., 2007) are among the most widely used parent report tools of vocabulary in infants and toddlers. The CDI presents parents with a checklist of items, a subset of which represents the most frequent words in children's early vocabulary across several semantic categories (animal names, toys, food and drink, clothing, body parts, action words, household objects, descriptive words, pronouns, etc.). The complete measure typically takes between 20 and 40 min to administer and can be completed independently by caregivers (or with support from a professional). In terms of vocabulary, the Words and Gestures version of the CDI (normed for ages 8–18 months) captures receptive and expressive vocabulary size, whereas the Words and Sentences version (normed for ages 16–30 months) captures only expressive vocabulary. Both versions provide individual item-level information about specific words in the child's lexicon as reported by parents.

Originally developed for English monolingual speakers (Fenson et al., 2007), the CDI has been adapted to over 50 language varieties, including the *Inventario de Desarrollo de Habilidades Comunicativas* (IDHC; Jackson-Maldonado et al., 2003), an adaptation developed with Spanish speakers

in Mexico. In addition, other monolingual Spanish adaptations are available from Spain, Argentina, Chile, Colombia, Cuba, and Peru. Adaptations include variations in word lists on the CDI depending on the language being assessed, with the degree of overlap being dependent on the languages and measures being compared (e.g., Norwegian and Spanish). Jackson-Maldonado et al. (2003) and Fenson et al. (2007) provide norms for monolingual Mexican Spanish and American English speakers, respectively, allowing for comparison to other age-matched monolingual peers. When examined by individual language, the CDI has strong psychometric properties, including internal consistency, test–retest reliability, and convergent validity (Fenson et al., 2007; Jackson-Maldonado et al., 2003).

A substantial number of studies have developed and validated bilingual measures using the monolingually derived CDI and IDHC in studies of early vocabulary size in Spanish–English DLLs in the United States. Seminal studies in this literature have established the calculation of bilingual measures such as total vocabulary (TV; the sum of Spanish and English words) and total conceptual vocabulary (TCV; the number of concepts with a known word in Spanish and/or English; Pearson et al., 1993). These measures have since been validated in subsequent studies. Vocabulary assessed separately in each language (Spanish vs. English) is correlated with input characteristics in that language (i.e., Hurtado et al., 2014; Place & Hoff, 2016), and the CDI and the IDHC demonstrate concurrent validity with other measures of Spanish and English (i.e., Martínez-Sussman & Marchman, 2002). In addition, extant work provides support for the use of cross-language measures in bilingual contexts, favoring TV estimates (the sum of Spanish and English words) for identifying early language delays (Core et al., 2013).

Limitations of Current Approaches

To the authors' knowledge, only a single measure presently exists with a bilingually derived word list in Spanish and English, such as the Spanish–English Vocabulary Checklist (SEVC, Patterson, 1998, 2000), though it is not widely used in contemporary research compared to the CDI. Although the SEVC provides a combined list of Spanish and English words, it is limited in its ability to capture Spanish word knowledge given that the words were translated from the English Language Development Survey into Spanish (Rescorla, 1989). Ideally, measures should be specifically adapted and derived in their respective cultural contexts following best practice for linguistic adaptations of assessment tools (Peña, 2007). In addition, despite the importance of capturing cross-language knowledge, the SEVC is limited in its ability to describe vocabulary across languages. The SEVC was developed primarily as a screening measure for a single caregiver to serve as the informant,

even in cases where the caregiver can only report on vocabulary in one but not both languages (Patterson, 1998).

In the absence of validated bilingual word lists, current monolingual CDI adaptations are often used in dual language contexts. Despite the promise of parent report measures in bilingual contexts, it is not clear if the research findings using monolingually derived CDIs would replicate using bilingual instruments that more precisely capture cross-language knowledge. Recently, some studies have extended CDI norms specifically for DLLs exposed to British English and other languages (i.e., Bengali, Cantonese, Dutch, French, German, Greek, Hindi-Urdu, Italian, Mandarin, Polish, Portuguese, Spanish, and Welsh; Floccia et al., 2018a). However, to our knowledge, CDI extensions specifically for DLL children learning Mexican Spanish and American English in the United States are not available.

Recall that bilingual vocabulary measures have been developed and validated in DLL contexts in prior research, including TCV, TV, and translation equivalents (TEs; i.e., number of words known in both languages with the same meaning; see Table 1; e.g., Pearson et al., 1993). However, this extant research uses monolingually derived instruments that do not fully capture cross-language knowledge. For example, in the Spanish IDHC and English CDI word lists, the word *araña* appears in Spanish, but its TE, *spider*, does not appear in the English list. Even cognates, which are TEs with overlapping word forms, are sometimes not included across word lists (e.g., *sopa* appears in Spanish, but *soup* does not appear in English; *lamp* appears in English, but *lámpara* does not appear in Spanish). This is problematic for bilingual contexts because a growing body of literature shows that bilingual children who know a word in one language are also more likely to know the TE in their second language, even if that word is not common in monolingual contexts (e.g., Bilson et al., 2015; Curtin et al., 2011; De Anda & Friend, 2020; Goodrich et al., 2016; Grasso et al., 2018). It is likely that such TE facilitation is modulated by the natural linguistic distance between languages (i.e., overlap among phonology and semantics), the distance between word pairs (e.g., whether the words are cognates), and experience (e.g., Floccia et al., 2018b). Regardless of the mechanism, the use of monolingual instruments in DLL contexts leads to undercounting of such transfer of vocabulary knowledge across languages.

Several studies suggest that accurate assessment of cross-language skills, such as knowledge of TEs, will best account for the lexical skills of DLLs by more precisely characterizing combined word knowledge. In particular, TEs are crucial for calculating TV and TCV, both of which are lexical measures that are frequently used to describe word knowledge across languages in DLLs. Research has shown that measures that consider children's *both* languages are, in general, better than single language measures

Table 1. Key vocabulary measures for dual language learners.

Measure	Construct	Definition	Calculation
Within-language measures	Spanish vocabulary	Words understood and/or said in Spanish	Sum of Spanish words
	English vocabulary	Words understood and/or said in English	Sum of English words
Cross-language measures	Total vocabulary	All words understood and/or said across languages	Sum of Spanish and English total vocabulary
	Total conceptual vocabulary	Concepts with one or two words understood and/or said (lexicalized concepts)	Total vocabulary minus translation equivalents
	Translation equivalents	Concepts with two words understood/said	Sum of concepts with words that are understood and/or said in <i>both</i> Spanish and English

for achieving diagnostic accuracy in clinical contexts (Mancilla-Martinez & Vaugh, 2013; Oh & Mancilla-Martinez, 2021). Furthermore, researchers have demonstrated that TV is better than TCV in identifying children potentially at risk for language impairment (Core et al., 2013) when using the English CDI and the Spanish IDHC. However, at present, the use of monolingual instruments in bilingual contexts has greatly limited the validity of cross-language measures.

In clinical contexts, valid cross-language measures form part of intervention planning and progress monitoring when targeting early vocabulary development. Tools that measure the words and concepts children have in both languages are therefore needed. From a practical perspective, providing two monolingual word lists to bilingual parents means the burden to caregivers is doubled as they are required to complete two separate questionnaires. As such, a procedure that allows for valid measurement of cross-language measures while also maximizing efficiency by providing a single bilingual list would support effective assessment practices that are responsive to the unique learning contexts of DLLs.

This Study

This study introduces the English–Spanish Vocabulary Inventory (ESVI) as a new tool that captures the breadth of cross-language knowledge in young Spanish and English learners while reducing parent burden of completion. Specifically, the ESVI presents the word lists from both the monolingual English CDIs and the Spanish IDHCs, with the addition of TEs. Recall that the CDI/IDHC Words and Gestures and Words and Sentences forms are widely used to capture lexical knowledge during the first, second, and third years of life. Given the critical importance of language assessment for prevention and intervention in the birth-to-3 period, two versions of the ESVI were created: The ESVI Expressive–Receptive (ESVI-ER) is based on the Words and Gestures version of

the CDI/IDHC and captures both comprehension and production in children with limited expressive skills, whereas the ESVI Expressive (ESVI-E) is based on the Words and Sentences version of the CDI/IDHC and captures only production for children with emerging expressive skills. Critically, additional lexical items were included in both of the ESVI questionnaires compared to the CDI/IDHC so that every word appears along with its TE. In this way, caregivers can report if a child understands or says a word in English, Spanish, or both languages.

The overarching aim of this study is to provide a methodological approach for combining questionnaires while also examining several important psychometric variables relevant to measure development, including validity and reliability. The measurement properties of the ESVI were evaluated across three different studies to provide preliminary evidence regarding the psychometric properties of the ESVI when employed with Spanish and English learners with and without suspected early language delays. Study 1 examines the content, criterion, and construct validity of expressive vocabulary estimates on the ESVI-E, whereas Study 2 replicates the criterion validity findings and examines social validity and reliability using both expressive and receptive vocabulary on the ESVI-ER. Study 3 introduces criterion validity and further extends the social validity findings in a geographically distinct sample of children compared to Studies 1 and 2 using the ESVI-E.

For an assessment tool to be of any value, evidence of strong psychometric properties when the tool is used with the intended population is essential. It is important that preliminary psychometrics be established for the ESVI because this approach uses a different format and scoring method than the CDI/IDHC (described in detail in the Method section). Such an examination in the initial stages of piloting the measure can inform whether the ESVI shows promise for further research into its diagnostic accuracy, for example, and if revisions to enhance the appropriateness of the tool are warranted. Specifically, language assessments must demonstrate acceptable

psychometric properties to support their use across contexts (e.g., Andersson, 2005; Mokkink et al., 2010). The inclusion of children with typical and atypical language learning also supports the measure's clinical utility, given that vocabulary is a common area of assessment in this population. In terms of psychometrics, content validity is important to establish because it captures the degree to which the measure is truly representative of the domain of interest (in this case, within and cross-language vocabulary knowledge). Conversely, social validity captures the measure's acceptability to caregivers, whereas criterion validity describes the degree to which a measure is related to an outcome. In addition, we ask: To what extent does the ESVI explain variance in vocabulary outcomes as compared to the business-as-usual approach of presenting families with two single-language lists? Furthermore, how do children with and without language delays compare in terms of vocabulary as measured on the ESVI? Together, the results of these measures of validity and reliability will help establish whether the ESVI has acceptable metrics in its present form. Given that the present measure extends and adapts best practice methods for bilingual assessment in the context of vocabulary and aligns closely with extant approaches, we expected that content, construct, social, and criterion validity would be acceptable or strong. Nevertheless, we expected that the findings would support further refinement of the measure to maximize diagnostic accuracy and efficiency and provide recommendations for clinical practice. These studies were approved by the institutional review boards of the respective institutions where the research was conducted.

Study 1

Method

Participants

Children and their mothers were originally recruited for a longitudinal study of child Spanish–English language development, maternal language input, and maternal well-being. To participate, mothers identified as Latina or Hispanic and reported that Spanish comprised at least 20% of their children's language exposure at home. A total of 50 mothers ($M_{\text{age}} = 33.86$ years, $SD = 6.01$) and their 52 children, including two sets of twins ($M_{\text{age}} = 26.57$ months, $SD = 2.86$), participated. Twenty-nine children were girls (56%) and 23 were boys (44%). Fifty-one children were born in the United States. According to mothers, 21 children (39.6%) had concerns for their behavioral, neurological, language, learning, or hearing development

and/or had received special education services for communication. On average, children's language exposure since birth was 77.8% Spanish ($SD = 21.3\%$) and 22.2% English ($SD = 21.3\%$), as measured on the Language Exposure Assessment Tool (De Anda et al., 2016). Four children also had 2%–4% exposure to a third language (Japanese, Zapoteco, Mixteco, or Mayan). The majority of mothers (67%) had a high school education or higher, were born outside the United States (80%), and identified their ethnicity as Mexican (76%). Participants were recruited in a metropolitan city in the Pacific Northwest of the United States. See Table 2 for additional detail on child and family characteristics.

Measures

ESVI-ER

The ESVI is a parent-reported receptive and expressive vocabulary measure that provides words common to the early vocabularies of English- and Spanish-learning children by adapting existing monolingual methods. The ESVI provides Spanish and English TEs side by side (e.g., *mesa/table*) to aid in capturing cross-language knowledge in bilingual children while also facilitating efficient completion for the caregiver. With permission from the CDI Advisory Board, all words from the monolingual norming samples of the CDI Words and Sentences and its Spanish adaptation, the Inventarios de Habilidades Comunicativas (IDHC), were listed on the ESVI-E version of the questionnaire. Because the IDHC and the CDI were adapted to Spanish and English contexts, respectively, many words on the CDI and the IDHC do not overlap. Two authors who are Latinx and native Mexican Spanish speakers (S. D. A. and L. H.) and have extensive clinical experience with bilingual children reviewed the word lists and added TEs where they were available. Any disagreements in translation were discussed until a consensus was reached among three of the study authors who all speak and use Spanish, including in clinical practice (S. D. A., L. H., and L. M. C.). Where TEs were not available, the original word was nevertheless kept in the ESVI to ensure that comparison to the CDI and the IDHC was possible. This process yielded an additional 206 Spanish words and 214 English words added to the original lists from the CDI and the IDHC to create the English–Spanish ESVI-E inventory. Spanish and English instructions are provided to parents, mirroring those provided by the original CDI and IDHC, with additional specification of how to respond when the child understands and/or says the word in English only, Spanish only, or both languages.

Five key expressive vocabulary variables are extracted from the measure. Table 1 summarizes the key definitions of each vocabulary measure of interest. First, to capture within-language vocabulary in (a) Spanish and (b) English, respectively, the total numbers of words

Table 2. Demographics of participating children and their families.

Variable	n	Study 1			Study 2				Study 3				
		n	M	SD	%	n	M	SD	%	n	M	SD	%
Child characteristics													
Age in months	87	52	26.57	2.86		13	32.35	8.24		22	26.1	6.56	
Sex													
Boy	50	23			44.2	11			84.6	16			72.7
Girl	37	29			55.8	2			15.4	6			27.3
Born in the United States		51			98.1		not asked			21			95.4
Developmental concerns ^a	56	21			42.0	13			100	22			100
Child language exposure													
% Exposure to Spanish ^b	87	52	77.8	22.2		13	77.54	28.03		22			
% Exposure to English		43	26.9	20.5		11	26.09	29.43					
% Exposure other language		4	0.03	0.01		1	8	N/A					
Caregiver characteristics													
Relationship to child													
Mother	85	50			100	13			86.7	22			100
Other	2					2			15.3				
Age in years			33.86	6.01			40.8	11.21			31.4	5.27	
Ethnicity													
Mexican	72	38			76.0	13			86.7	21			95.4
Guatemalan	4	3			6.0	1			6.7				
Colombian	1	1			2.0								
Dominican	1	1			2.0								
Salvadoran	1	1			2.0								
Nicaraguan	1	1			2.0								
Chilean	1	1			2.0								
Argentine	1	1			2.0								
Other	2	0				1			6.7	1			4.5
Multiethnic	3	3			6.0								
Education (n = 86)													
Less than high school	20	17			34.0	2			13.3	1			4.5
High school diploma or GED	47	22			44.0	11			73.3	14			63.6
Associate's degree	3	1			2.0	1			6.7	1			4.5
Bachelor's degree	13	7			14.0					6			27
Master's degree or higher	3	3			6.0								
Born in the United States	16	10			19.2	1			6.7	5			22.7
Annual family income													
< \$10,000	8	6			12.0					2			9.1
\$10,001–20,000	8	5			10.0	1			6.7	2			9.1
\$20,001–30,000	22	11			22.0	5			33.3	6			27
\$30,001–40,000	7	6			12.0	1			6.7				
\$40,001–50,000	4	2			4.0					2			9.1
\$50,001–60,000	8	6			12.0					2			9.1
\$60,001–70,000	3	1			2.0					2			9.1
> \$70,001	8	6			12.0	1			6.7	1			4.5
Unknown	17	7			14.0	5			33.3	5			23

^aRefers to parent-reported concerns and/or participation in early intervention or early childhood special education. ^bCumulative exposure as measured by the Language Exposure Assessment Tool was only used in Studies 1 and 2. N/A = not applicable; GED = General Educational Development.

reported to be understood and said in each language are summed. To capture overlapping cross-language knowledge, a count of the total reported number of (c) TEs is calculated. To capture (d) TV, the numbers of words in Spanish and English are summed. To capture (e) TCV, the number of TEs is subtracted from TV to provide a count of the number of concepts with a word in either Spanish or English. In order to compare the utility of the ESVI against the monolingual CDI and IDHC word lists, the same variables (a–e) were calculated using the same formulas, but this time using only the words that appear

on the original CDI and IDHC, removing those words that were added to create the ESVI-E's fully translated word lists in English and Spanish. In this way, we were able to compare these key vocabulary estimates across the ESVI-E and the CDI/IDHC.

Computerized Comprehension Task

The Computerized Comprehension Task (CCT; Friend & Keplinger, 2003; Friend et al., 2012) is a behavioral assessment intended to capture children's receptive vocabulary knowledge in Spanish and English. The CCT

has significant test–retest reliability across English and Spanish adaptations in children as young as 16 months of age. The CCT also shows convergent validity with parent reports of vocabulary (Friend & Keplinger, 2008; Friend & Zesiger, 2011) and predicts significant variance in vocabulary production outcomes (Friend et al., 2012). Recent adaptations of the English and Spanish CCT have extended administration up to 36 months of age, following the same procedures for item selection. These extensions also show convergent validity, construct validity, and predictive validity such that performance on the CCT at 24 months of age is positively associated with performance at 30 months of age in a preliminary study (De Anda et al., 2020).

The CCT presents children with pairs of images (a target and distractor object) presented on a touch-sensitive screen by a trained experimenter. Children are prompted to touch the target after a standard elicitation prompt (e.g., “¿Dónde está el zapato? Toca zapato”, or “Where is the shoe? Touch shoe.”). In this study, children received both the Spanish and English versions of the CCT (Friend & Keplinger, 2003; Friend et al., 2012). To capture children’s within-language receptive vocabulary knowledge, an accuracy score was computed by counting the number of trials in which the child correctly identified the target object on the Spanish and English versions, respectively. Consistent with published procedures (De Anda et al., 2018), coders used the Eudico Linguistics Annotator (<http://tla.mpi.nl/tools/tla-tools/elan/>, Max Planck Institute for Psycholinguistics, The Language Archive; Lausberg & Sloetjes, 2009) to denote the onset of the first target word presentation and the onset of the child’s touch response for each trial. A trial was marked as correct if the child’s first touch was to the target object and if it occurred after the first presentation of the target word. Coders were trained to allow pointing responses for children who preferred not to touch the screen. Trained coders analyzed video recordings of the CCT administration and achieved acceptable interrater reliability (95% agreement) on data from 25% of participants.

Spontaneous Language Sample

Mothers and their children completed a 10-min interaction from which the child’s expressive language was analyzed. Mothers and children had access to a standard set of toys appropriate for toddlers to facilitate opportunities for concrete and symbolic play: (a) toy cookware and plastic food items, (b) farm animals and farmhouse or building blocks, and (c) the same book in Spanish and English. Similar to the Three Bags Task commonly used to study early parent–child interactions (e.g., Tamis-LeMonda et al., 2004), mothers were encouraged to use all three sets of toys in ways that felt natural to them using their language(s) of preference. No instruction was given as to whether mothers should use English or Spanish, but rather they were allowed to use one or both languages in whatever degree they felt was typical.

To code the language samples, the language sample recordings were transcribed and coded in their original languages by trained bilingual Spanish–English research assistants, using the conventions from the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2018). To ensure accuracy of transcription and coding, the transcript was reviewed alongside the video of the recording by a second research assistant. Then, the SALT company staff reviewed and updated each transcript to ensure accuracy to coding conventions.

For the purposes of describing child productive communication, several key variables were generated using SALT software: number of different words (NDW) and number of total words (NTW). These estimates included only complete and intelligible utterances. Given that mothers and children used Spanish and English during the language sample to varying degrees, language mixing (i.e., code switching) was expected. Thus, NDW and NTW capture vocabulary across languages, so as to better compare with the critical cross-language measures from the ESVI for the purposes of examining content validity.

In addition, to account for vocalizations that were unintelligible but still communicative, all nonword intentional child vocalizations were coded. This did not include reflexive vocalizations or vocalizations that were not made for the purposes of communicating. Whereas NDW and NTW describe intelligible productions of words, vocalizations include child speech productions that are intentional for the purpose of communication but unintelligible as true words. For example, children with emerging expressive skills might point to a toy and produce speech, in an attempt to draw their mother’s attention to the object, modeling early object naming. Such vocalizations are not true intelligible words but are included because children at 24 months are still producing intentional communicative vocalizations alongside words, thereby contributing to the child’s quantity and productivity of communication. Language samples had, on average, 55.08 total utterances ($SD = 35.27$), which included vocalizations and verbalizations, and 22.96 complete and intelligible verbal utterances ($SD = 25.41$) coming from the child.

To achieve reliability, a random selection of 20% of the observations were transcribed and coded in full by research assistants who were not involved in their original transcription. The reliability transcripts were also reviewed by SALT staff for coding accuracy. Interrater reliability was determined by calculation of intraclass correlation coefficients based on a one-way random effects models for NDW, NTW, and child vocalization counts. The intraclass correlation coefficient was .94 for NDW (95% CI [.82, .98]), .93 for NTW (95% CI [.80, .98]), and .97 for child vocalizations (95% CI [.92, .92]), indicative of excellent interrater reliability.

Procedure

Following consent and a warm-up period, children and their mothers completed the behavioral tasks (the CCT and spontaneous language sample), followed by administration of the ESVI. All mothers were offered the choice of filling out the questionnaire independently or with assistance (e.g., Jackson-Maldonado et al., 2005). At this time, caregivers also reported whether they ever had concerns about their child's language development and whether their child had at any time received early intervention or special education services due to language or communication concerns.

Results

Table 3 summarizes children's ESVI-E scores across the key variables of interest at the group level. The table also includes the same variables estimated using only those words that appear on the CDI and the IDHC for comparison. Given that the ESVI-E uses the words from the CDI and the IDHC, the correlation among the two measures was high for all key variables (all $r_s > .9$, $p_s < .01$). Nevertheless, we expected that the ESVI-E's approach of providing all words across languages would reveal additional cross-language lexical knowledge that would not otherwise be captured using existing monolingual methods (i.e., using the CDI and the IDHC separately for each language). Indeed, t tests revealed that the ESVI-E captured significantly more cross-language knowledge than the CDI/IDHC, as measured by the proportion of the child's vocabulary composed of TEs, $t(50) = 3.929$, $p < .001$, $d = 0.550$. The effect size shows that the difference in vocabulary knowledge captured by the ESVI-E compared to the CDI/IDHC was moderate.

Content Validity

Content validity describes the degree to which the ESVI-E is truly representative of children's vocabulary knowledge. To show evidence of acceptable content validity, we expect that parent report on the ESVI-E will be associated with concurrent behavioral measures of receptive (the CCT) and emerging expressive vocabulary skills (language sample: NDW, NTW, and number of vocalizations) that directly assess the child's communication skills.

We first examined whether parent report on the ESVI-E predicted performance on the CCT. Since ESVI-E and CDI/IDHC vocabulary scores did not meet the normality assumptions for linear regression, log transformations of vocabulary scores were used for all such analyses. Two regression models examined vocabulary size in Spanish and English, respectively, and included children's CCT accuracy score as the dependent variable and the log of ESVI-E vocabulary scores as the independent variable. Parent report of Spanish vocabulary on the ESVI-E predicted children's accuracy scores on the CCT in Spanish, $F(1, 43) = 4.482$, $p = .040$, $R^2 = .094$. In English, the ESVI-E was also a significant predictor of scores on the English CCT, $F(1, 42) = 12.79$; $p < .001$, $R^2 = .234$. See Supplemental Material S1 for correlations and Supplemental Material S2 regression results across studies.

To compare the ESVI-E results to the CDI and the IDHC, separate linear regression models included only CDI and IDHC scores in English and Spanish, respectively: Spanish IDHC vocabulary, $F(1, 43) = 4.483$, $p = .040$, $R^2 = .094$; English CDI vocabulary, $F(1, 42) = 12.02$, $p = .001$, $R^2 = .223$. Results showed that the ESVI-E captured similar variance in children's receptive vocabulary scores on the CCT, suggesting that we did not lose information by using the ESVI-E's bilingual approach compared to the CDI and the IDHC validated in monolinguals.

Table 3. Descriptive statistics for key vocabulary measures across studies.

Measure	Study 1		Study 2		Study 3
	Expressive		Receptive	Expressive	Expressive
	<i>M (SD)</i>		<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
ESVI					
Spanish total vocabulary	106.61 (121.847)	162.077 (115.758)	63.462 (79.667)	84.818 (128.588)	
English total vocabulary	76.373 (105.137)	68.846 (65.127)	19.154 (24.252)	51.091 (107.642)	
Total vocabulary	182.98 (215.616)	230.923 (149.41)	83.385 (118.417)	135.91 (203.891)	
Total conceptual vocabulary	139.078 (143.672)	187.615 (104.376)	73 (100.071)	114 (162.412)	
Translation equivalents	43.902 (82.09)	43.308 (52.680)	10.385 (22.198)	21.91 (53.225)	
CDI/IDHC					
Spanish total vocabulary	94.45 (103.95)	128.539 (79.279)	50.615 (77.821)	75.273 (116.414)	
English total vocabulary	67.69 (90.60)	52.461 (46.596)	14.077 (14.529)	45.681 (92.596)	
Total vocabulary	162.14 (184.56)	181.0 (88.649)	64.692 (82.922)	120.955 (159.28)	
Total conceptual vocabulary	129.29 (132.95)	156.923 (71.620)	59.769 (76.937)	101.0 (133.712)	
Translation equivalents	32.84 (58.17)	24.077 (24.692)	4.923 (8.967)	19.954 (47.570)	

Note. ESVI = English-Spanish Vocabulary Inventory; CDI = MacArthur-Bates Communicative Development Inventories; IDHC = Inventario de Desarrollo de Habilidades Comunicativas.

Next, we examined whether the ESVI-E predicted children's productions during a language sample. Unlike the CCT analyses, which evaluated Spanish and English word knowledge separately, we favored cross-language variables (e.g., ESVI-E TCV and TV) in order to facilitate and simplify comparison of the ESVI-E across various language sample variables (NDW, NTW, and number of vocalizations). Children's NDWs produced during the language sample were entered as the dependent variable, and the log of ESVI-E TCV was entered as the independent variable, resulting in a significant model, $F(1, 49) = 48.73$, $p < .001$, $R^2 = .50$. As with the CCT analyses, a separate model including only CDI/IDHC TCV as the dependent variable showed similar variance explained compared to the ESVI-E TCV, $F(1, 49) = 46.06$, $p < .001$, $R^2 = .484$. The same overall pattern of results held when using TV as the dependent variable across the ESVI-E and CDI/IDHC models, respectively, such that the ESVI-E vocabulary scores were significant predictors and explained similar variance than CDI/IDHC scores: ESVI-E TV, $F(1, 49) = 43.41$, $p < .001$, $R^2 = .470$; CDI/IDHC TV, $F(1, 49) = 42.65$, $p < .001$, $R^2 = .465$. The same pattern of results was shown for NTW, such that both ESVI-E estimates of TCV and TV explained similar variance as estimates from the CDI and the IDHC: ESVI-E TCV, $F(1, 49) = 46.51$, $p < .001$, $R^2 = .487$; CDI/IDHC TCV, $F(1, 49) = 44.26$, $p < .001$, $R^2 = .474$; ESVI-E TV, $F(1, 49) = 40.03$, $p < .001$, $R^2 = .450$; CDI/IDHC TV, $F(1, 49) = 39.73$, $p < .001$, $R^2 = .448$.

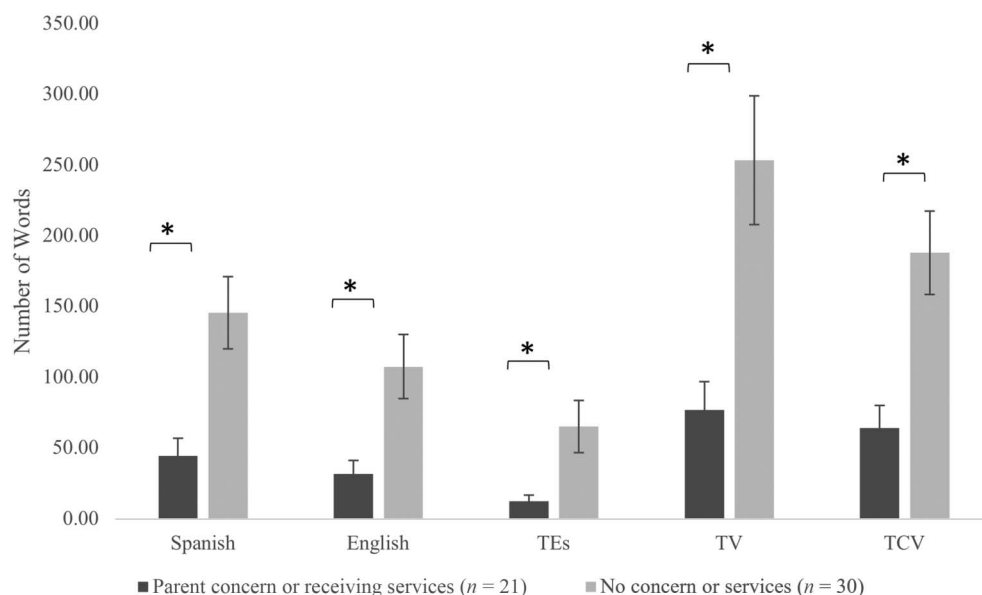
Criterion Validity

In addition to examining NDW and NTW from children's language samples, we analyzed whether the ESVI-E predicted children's nonword vocalizations. Once again, the logs of TCV and TV scores on the ESVI-E and the CDI/IDHC were entered as predictors across four separate models. Results were consistent with the findings from the CCT, NDW, and NTW such that no explanatory power was lost with ESVI-E estimates compared to those derived from the CDI/IDHC: ESVI-E TCV, $F(1, 49) = 15.41$, $p < .001$, $R^2 = .239$; CDI/IDHC TCV, $F(1, 49) = 14.72$, $p < .001$, $R^2 = .231$; ESVI-E TV, $F(1, 49) = 13.57$, $p < .001$, $R^2 = .217$; CDI/IDHC TV, $F(1, 49) = 13.33$, $p < .001$, $R^2 = .214$.

Construct Validity

Given that children with a range of language abilities were included in this study, we described construct validity by examining ESVI-E vocabulary sizes as a function of (a) whether parents identified language concerns and (b) whether children had at any time received early intervention or special education services due to language or communication concerns (see Figure 1). The t tests compared ESVI-E scores across all vocabulary variables (Spanish vocabulary, English vocabulary, TV, TCV, and TEs). Results showed that children for whom parents had language concerns or that were currently receiving special education services to support language and communication had significantly smaller TV (language delay: $M = 77$, $SD = 92.44$; no language delay: $M = 253.70$, $SD = 249.23$), $t(39.93) = 3.52$,

Figure 1. English–Spanish Vocabulary Inventory vocabulary scores as a function of parental concerns for language and communication. Spanish = vocabulary size in Spanish; English = vocabulary size in English; TEs = translation equivalents; TV = total vocabulary; TCV = total conceptual vocabulary. * $p < .05$.



$p = .001$, $d = 0.88$; TCV (language delay: $M = 64.37$, $SD = 73.22$; no language delay: $M = 188.27$, $SD = 161.39$), $t(43.51) = 3.653$, $p < .001$, $d = 0.93$; Spanish vocabulary (language delay: $M = 44.90$, $SD = 56.24$; no language delay: $M = 145.87$, $SD = 139.77$), $t(41.37) = 3.531$, $p < .001$, $d = 0.89$; English vocabulary, (language delay: $M = 32.11$, $SD = 42.99$; no language delay: $M = 107.83$, $SD = 124.177$), $t(38.778) = 3.063$, $p = .004$, $d = 0.76$; and TEs (language delay: $M = 12.63$, $SD = 19.824$; no language delay: $M = 65.43$, $SD = 101.033$), $t(32.44) = 2.779$, $p = .009$, $d = 0.67$, than their peers without language concerns or services at a Bonferroni corrected alpha level of .01.

Interim Summary

Study 1 demonstrated content and criterion validity in that parent report of expressive vocabulary on the ESVI-E significantly predicted receptive vocabulary on the CCT and expressive language and vocalizations. Furthermore, the ESVI-E showed acceptable construct validity in that children with parent-identified language concerns had significantly lower vocabulary scores than children for whom no concerns or prior receipt of speech-language services were reported. These differences were most apparent in Spanish compared to English, likely due to the fact that Spanish was the dominant language of exposure for this group of children.

Next, Study 2 sought to replicate the criterion validity findings against a more naturalistic measure of child communication outside the lab setting (e.g., home language sampling) and examine social validity in Spanish-learning children with language delays. Given that children with communication concerns often have limited expressive skills, we additionally collected receptive vocabulary using the ESVI-ER to examine the extent to which results were replicated across comprehension and production. We expected that parent-reported vocabulary size on the ESVI-ER would be associated with children's rate of communication at home. Furthermore, although the sample size in Study 2 is somewhat small, the findings were included because they contribute to the extremely limited research base on vocabulary assessment in Spanish–English DLL toddlers with identified language delays and provide preliminary evidence of the utility of the ESVI-ER. We return to interpreting the findings in light of such limitations in the Discussion section.

Study 2

Method

Participants

Children and their caregivers were recruited in the U.S. Pacific Northwest for a larger study examining language

intervention in children exposed to Spanish with early language delays (Cycyk et al., 2020). Caregivers were eligible for the study if they spoke Spanish at home and had a child 5 years of age or younger who they believed spoke less than 200 total words across Spanish and English. A total of 11 families participated in the intervention, including 15 caregivers ($M_{\text{age}} = 40.8$ years, $SD = 11.21$) and 13 children (two sets of twins; $M_{\text{age}} = 32.25$ years, $SD = 8.24$; two girls and 11 boys). Caregivers participating in the study were predominantly mothers ($n = 13$); one father and one grandmother also participated. Most caregivers ($n = 13$) identified as Mexican. All but one caregiver were born outside the United States, and most (73%) completed high school. All heard Spanish and English at home, and on average, exposure estimates indicated that children as a group heard Spanish ($M = 78\%$, $SD = 28\%$) more often than English ($M = 22\%$, $SD = 28\%$) as measured on the Language Exposure Assessment Tool (De Anda et al., 2016). Eleven children were receiving special education services at the time of the study and had been identified by the local early intervention agency as having an expressive communication disorder. Two children were referred due to parental concern about language development (Paradis et al., 2010; Restrepo, 1998), which was confirmed by a certified Spanish–English bilingual SLP (see Table 2).

Measures

ESVI-ER

The ESVI-ER used in Study 2 followed the same procedure for development used in the ESVI-E described in Study 1. Unlike in Study 1 (but again with the permission of the CDI Advisory Board), Study 2 used the Words and Gestures versions of the CDI and the IDHC as the initial word lists from which the ESVI-ER was developed. The Words and Gestures version captures caregiver report of words the child understands (receptive) and the words the child can say (expressive). A comparison of word lists across the English CDI and the Spanish IDHC for TEs yielded an additional 217 Spanish words and 228 English words to create the final receptive and expressive English–Spanish inventory on the ESVI-ER. Recall that all words on the CDI and the IDHC were included in the ESVI-ER to ensure comparison to the CDI and the IDHC was possible. The ESVI-ER word list was developed and reviewed by the same team in Study 1 following identical procedures.

Five variables were calculated for participants in Study 2 for receptive and expressive knowledge separately, yielding a total of 10 vocabulary estimates from the ESVI-ER: (a) Spanish vocabulary size: receptive and expressive; (b) English vocabulary size: receptive and expressive; (c) TEs: receptive and expressive; (d) TV: receptive and expressive; and (e) TCV: receptive and expressive. Similarly, in order to compare the utility of the ESVI-ER against the monolingual CDI and IDHC word lists, the same 10

variables were calculated using only the words that appear on the original Words and Gestures versions of the CDI and the IDHC, removing those words that were added to create the ESVI-ER's fully translated word lists in English and Spanish.

Language ENvironment Analysis

For a subset of participants ($n = 9$), home audio recordings of the child during typical waking hours were also collected. The Language ENvironment Analysis (LENA) Pro System (2012) captures the child's rate of vocalization as a measure of expressive language. Vocalizations as captured by the LENA are only speech vocalizations (i.e., reflexive and vegetative sounds are not included). The LENA Digital Language Processor (DLP) device sits on the inside pocket of a vest worn by the child and records all audio in the environment. The accompanying software uses speech recognition technology to calculate the total number of child vocalizations, with approximately 75% accuracy in English contexts (Xu et al., 2009). The use of LENA has also been expanded successfully to multilingual contexts that include Spanish speakers (e.g., Jackson & Callender, 2014; Marchman et al., 2017; Weisleder & Fernald, 2013; Wood et al., 2016).

Caregivers were provided visual and verbal instructions for using the DLP and asked to record a full day of their child's life (up to 16 hr in the home). As a group, families provided an average of 692.0 min ($SD = 238.56$) recorded prior to the intervention and an average of 547.72 min recorded ($SD = 275.46$) following the intervention. To calculate rate of communication, we divided the LENA-estimated total number of vocalizations by the number of minutes recorded by the LENA. The estimate of vocalizations from the LENA's full-day recording is comparable to the coding of vocalizations from the language samples in Study 1.

Parent Report of Social Validity

Given that the ESVI relies entirely on caregiver report, four questions were posed to all caregivers regarding their experience with the measure. Caregivers were asked whether they had difficulty filling out the questionnaire or answering questions and whether they recruited others to help fill out the questionnaire. In cases where the caregiver was unable to report on the child's skills in a language (because they only observed the child in one and not both language contexts for example), they were allowed to recruit the aid of the additional caregiver for the language in question. Caregivers were also asked to report on their time to completion in minutes and on their confidence that the answers on the ESVI correctly reflected their child's communication skills.

Procedure

Caregivers were provided the ESVI-ER questionnaire in a one-on-one context with graduate student

clinicians prior to receiving the intervention delivered in the larger study from which these data were drawn. Clinicians provided verbal instructions for the questionnaire and provided examples before checking for comprehension. Families were encouraged to ask questions at this stage and instructed to take the ESVI-ER home and return it approximately 1 week later. During this same meeting, families were also shown how to operate the LENA DLP through visual and verbal instructions. The clinician discussed the requirements for audio recording (a full typical day at home) and developed a plan with the family for the best dates and times for recording. As with the ESVI-ER, the audio recording was returned approximately 1 week later. Once the ESVI-ER was returned, the clinician verified completion of the questionnaire and completed the social validity questionnaire verbally with the caregiver. Following the collection of the ESVI-ER, LENA, and social validity questionnaire, families received approximately 7 weeks of instruction for a culturally and linguistically adapted parent-mediated naturalistic communication intervention (Cycyk et al., 2020). At the end of the intervention, families were once more asked to complete the ESVI-ER questionnaire by updating their first version to capture changes in their child's vocabulary knowledge and to report on the duration of time it took to complete.

Results

Criterion Validity

Table 3 presents descriptive results for key vocabulary estimates. As in Study 1, log transformations were completed for vocabulary estimates on the ESVI-ER and CDI/IDHC for inclusion in regression models given that they violated normality assumptions. Results replicated findings from Study 1, such that expressive ESVI-ER TV scores significantly predicted children's rate of vocalizations on the LENA home language recordings, $F(1, 10) = 5.313$, $p = .043$, $R^2 = .347$, but the results were marginal for the ESVI-ER TCV model: TCV, $F(1, 10) = 4.505$, $p = .05$, $R^2 = .311$. However, neither expressive CDI/IDHC TCV nor expressive CDI/IDHC TV scores were significant predictors of children's rate of vocalization.

Unlike expressive estimates of TCV and TV, receptive estimates on the ESVI-ER did not predict children's rate of vocalization. Similarly, receptive TCV and TV estimates from the CDI/IDHC did not yield significant models in predicting vocalization.

Vocabulary Change

We expect that the ESVI-ER should track vocabulary change consistent with developmental expectations.

Indeed, vocabulary growth is linear in Spanish-learning toddlers (Jackson-Maldonado et al., 2003), and thus, performance at Time 1 should correlate with Time 2. To assess the reliability of the ESVI-ER for capturing developmental vocabulary change over time, correlations between Time 1 and Time 2 were examined across each of the four key measures for receptive and expressive vocabulary estimates: (a) Spanish vocabulary: receptive, $r(12) = .81, p = .001$; expressive, $r(12) = .92, p < .001$; (b) English vocabulary: receptive, $r(12) = .98, p = .001$; expressive, $r(12) = .89, p < .001$; (c) TEs: receptive, $r(12) = .58, p = .047$; expressive, $r(12) = .95, p < .001$; and (d) TCV: receptive, $r(12) = .86, p = .001$; expressive, $r(12) = .98, p < .001$. Across receptive and expressive vocabulary estimates, correlations were consistently strong and in the positive direction between Time 1 and Time 2 administrations.

In addition, we conducted t tests to examine whether children demonstrated significant change in ESVI vocabulary over the 2-month time window. Paired t tests revealed that vocabulary size was significantly larger at Time 2 compared to Time 1 for total receptive and expressive Spanish vocabulary, receptive: $t(11) = 4.29, p = .001$; expressive: $t(11) = 2.95, p = .013$, and receptive and expressive TCV, receptive: $t(11) = 4.81, p < .001$; expressive: $t(11) = 3.5, p = .005$. Total receptive and expressive English vocabulary and TEs showed less variability in the group of Spanish-dominant toddlers and did not evince a significant difference between Time 1 and Time 2 (all *n.s.* $ps > .06$).

Social Validity

Most caregivers ($n = 9$ out of 13 surveyed; 69%) felt “very confident” that their ESVI-ER responses reflected their child’s communication skills, whereas the remainder said they were “somewhat confident.” Notably, four caregivers reported difficulty completing the questionnaire and recruited family support for reporting vocabulary across languages. The recruited family members included the child’s sibling ($n = 1$), grandmother ($n = 2$), and father ($n = 1$). Lastly, most caregivers completed the questionnaire below the maximum expected 80 min at Time 1 ($Mdn = 60$ min, $M = 75.83$ min, $SD = 47.19$ min; recall that administration of each English and Spanish CDI/IDHC questionnaire can take up to 40 min to complete; Fenson et al., 2007). Administration time was almost cut in half at Time 2 ($M = 41.89$ min, $SD = 36.049$) when caregivers were asked to update their responses from Time 1 (see Table 4 for responses).

Interim Summary

Results from Study 2 replicated and extended the findings from Study 1 to a group of Spanish-learning

Table 4. Summary of responses for the social validity questionnaire in Study 2.

Social validity questionnaire items	<i>n</i> (%)
¿Tuvo alguna dificultad en llenar este documento o en contestar alguna de las preguntas? (Did you have any difficulty filling this document out or answering any of the questions?)	
Si (Yes)	4 (31)
No (No)	9 (69)
¿Alguien le ayudo a llenar este documento? (Did someone help you fill out this document?)	
Si (Yes)	4 (31)
No (No)	9 (69)
¿Cuanto tiempo aproximadamente le tomo en llenar este documento? (How long did it take you to fill out this document?)	
30 min or less	4 (31)
Between 31 and 60 min	4 (31)
Between 60 and 90 min	2 (15)
Greater than 90 min	3 (23)
¿Cuan seguro/a está de que sus respuestas en este documento reflejan correctamente las destrezas de comunicación de su niño/a? (How confident are you that your answers in this document correctly reflect your child’s communication skills?)	
Muy seguro/a (very confident)	9 (69)
Algo seguro/a (somewhat confident)	4 (31)
No muy seguro/a (not confident)	0

toddlers and preschoolers with concerns about early language delay. Specifically, we showed that criterion validity was replicated when examining the correlation between the ESVI-ER and children’s rate of vocalization as captured through a naturalistic home language sample from a full-day audio recording on the LENA DLP. Furthermore, parent report on the ESVI-ER demonstrated strong reliability for tracking vocabulary change over a 2-month period and acceptable social validity in children at risk of language disorders. Study 3 next sought to replicate and extend the social validity findings and examine the criterion validity of the ESVI against a standardized language assessment used widely in practice and research.

Study 3

Method

Participants

Families who participated in Study 3 were recruited from an early intervention program and the broader community in a Western state in the United States and formed part of a larger study examining language-screening tools in Spanish-English-speaking children from Latinx backgrounds (King et al., 2021). Child participants were included in the study if they were between 12 and 36 months,

had at least one parent who identified as having Hispanic or Latino heritage, and lived in a home where their caregiver(s) spoke Spanish or a mix of Spanish and English. A total of 22 mothers ($M_{\text{age}} = 31.27$ years, $SD = 5.34$) and their 22 children ($M_{\text{age}} = 26.14$ months, $SD = 6.56$; 16 boys and six girls) participated. Demographic and language exposure information was collected using the Center for Early Care and Education Research Dual Language Learner Child and Family Questionnaire (Hammer et al., 2020). All but one child were identified by their caregiver as Hispanic or Latino. Three children were also identified as White, and the remaining child was identified as White and “1/8 Mexican.” Twenty-one children (95.5%) were born in the United States. The parents of 17 children reported concern regarding their child’s expressive and receptive language development, and 13 children were participating in early intervention services at the time of the study. Most mothers ($n = 21$) identified as of Mexican descent; the remaining mother responded “North American” for race/ethnicity and explicitly referenced her Mexican heritage. Five were born in the United States. The vast majority ($n = 21$) completed high school or greater. Although there was a range of Spanish and English exposure, children, on average, heard more Spanish from their mothers than English ($M = 77\%$ of maternal input to children was in Spanish, $SD = 22.66$, range: 30%–100%). Overall, children’s language output was relatively balanced as a group: per maternal report, approximately half of children tended to speak to their caregiver primarily in Spanish, either entirely or to a greater degree than English ($n = 11$), whereas the remaining children spoke English and Spanish to a similar degree ($n = 2$) or predominantly in English ($n = 7$). Table 2 provides demographic characteristics of children and their caregivers.

Measures

ESVI-E

The same ESVI-E measure described in Study 1 (capturing expressive vocabulary only) was used in Study 3.

Preschool Language Scales–Fifth Edition Spanish

Children were administered the Preschool Language Scales–Fifth Edition Spanish (PLS-5 Spanish; Zimmerman et al., 2011), which is a standardized, comprehensive dual-language assessment of receptive and expressive English and Spanish language skills for young children birth to 7;11 (years;months). The PLS-5 was normed on 1,150 Spanish–English DLLs living in the United States and has adequate internal consistency score reliability ($r = .87$ – $.97$) and test–retest score reliability ($r = .91$ – $.92$; Zimmerman et al., 2012). The measure includes two subtests (Auditory Comprehension and Expressive Communication) that each yield separate raw scores that were used for analyses. All

items on the PLS-5 are presented first in Spanish. Once a ceiling is achieved in Spanish, the items are presented a second time in English. The final score is based on the correct responses across languages. Items are scored with caregiver report, observation, or through direct elicitation by the examiner.

Centers for Disease Control and Prevention Developmental Milestones Form

Parents completed the Spanish translation of the Centers for Disease Control and Prevention (CDC) Developmental Milestones Checklist (CDC, n.d.), which lists activities typically associated with development between 2 months and 5 years of age. Respondents indicate whether their child has met social–emotional, language, and motor skills typically achieved by same-age children. Because the number of milestones listed on the form differs based on age, we calculated the proportion of items checked within the child’s chronological age to allow for a comparable scale across participants 12–24 months.

Parent Report of Social Validity

The social validity questionnaire for Study 3 was different than the one employed in Study 2, though it also queried parents on their perceptions. For Study 3, the social validity questionnaire presented nine statements (read aloud in Spanish by the experimenter), which parents were to rate on a scale of 1–4 (see Table 5). The questions were similar to those in Study 2 but further expanded on the nature of caregiver perceptions.

Procedure

Data collectors included bilingual (Spanish–English) certified SLPs and an SLP master’s student. Data collectors were trained on the administration and scoring of the PLS-5 Spanish by reading the examiner’s manual, observing at least one administration of the PLS-5 Spanish (live or recorded), and role-playing administration items. All measures were introduced to families using standardized directions in Spanish. Testing was conducted in the home environment, and most assessments occurred on the same day. No assessments occurred more than a month apart.

Results

Concurrent Criterion Validity: PLS-5 and CDC Developmental Milestones

Concurrent criterion validity examines the degree to which a measure is related to a concurrent and theoretically related outcome. Unlike content validity, which captures the degree to which the ESVI-E measures the key aspects of interest (i.e., early vocabulary), criterion validity describes whether ESVI-E vocabulary scores are associated

Table 5. Summary of responses for the social validity questionnaire in Study 3.

Social validity questionnaire items	M (SD)
1. <i>Estoy interesado(a) en saber las palabras que mi hijo(a) entiende y dice al desarrollar su lenguaje.</i> (I am interested in knowing the words my child understands and says as they develop their language.)	3.773 (0.429)
2. <i>Fue fácil para mí saber recordar e indicar las palabras que mi hijo(a) entiende/dice en la lista de vocabulario.</i> (It was easy for me to know how to remember and mark the words my child understands/says on the vocabulary list.)	3.545 (0.510)
3. <i>Usar la lista de vocabulario más frecuentemente (por ejemplo una vez cada mes) sería fácil para mí hacer.</i> (Using the vocabulary list more frequently [for example once a month] would be easy for me to do.)	3.364 (0.492)
4. <i>Me gustaría usar una aplicación para llevar un registro de las palabras específicas que dice y entiende mi hijo(a) en inglés y español.</i> (I would like to use an application to keep a register of the specific words my child knows and says in English and Spanish.)	3.500 (0.512)
5. <i>Confío en que mis respuestas en la lista de vocabulario representan el vocabulario de mi hijo(a) en inglés.</i> (I am confident that my responses on the vocabulary list represent my child's vocabulary in English.)	3.500 (0.598)
6. <i>Confío en que mis respuestas en la lista de vocabulario representan el vocabulario de mi hijo(a) en español.</i> (I am confident that my responses on the vocabulary list represent my child's vocabulary in Spanish.)	3.591 (0.590)
7. <i>Después de cumplir la lista de vocabulario, recordé palabras adicionales que mi hijo(a) entiende o dice que no noté en la lista de vocabulario.</i> (After completing the vocabulary list, I remembered additional words that my child understands and says that I did not mark on the vocabulary list.) ^a	2.833 (0.983)
8. <i>Todas las palabras mi hijo(a) dice están incluidas en la lista de vocabulario.</i> (All of the words my child says are included in the vocabulary list.)	3.227 (0.612)

Note. Respondents were asked to select from the following answer options: 1 = *totalmente en desacuerdo* (totally disagree); 2 = *en desacuerdo* (disagree); 3 = *de acuerdo* (agree); 4 = *totalmente de acuerdo* (totally agree).

^aOnly six respondents provided a response for this question.

with other related cognitive skills. In this case, we expected that caregiver report of lexical knowledge on the ESVI-E will predict children's concurrent broader language abilities as measured on a standardized assessment (the PLS-5) and also reflect parental perceptions of children's overall development as reported on the CDC Developmental Milestones Checklist. To answer these questions, a series of regression models first examined whether TCV and TV on the ESVI-E would predict raw scores for the comprehension and production subtests on the PLS-5. Theoretically, we expected a positive association such that children who have lexicalized

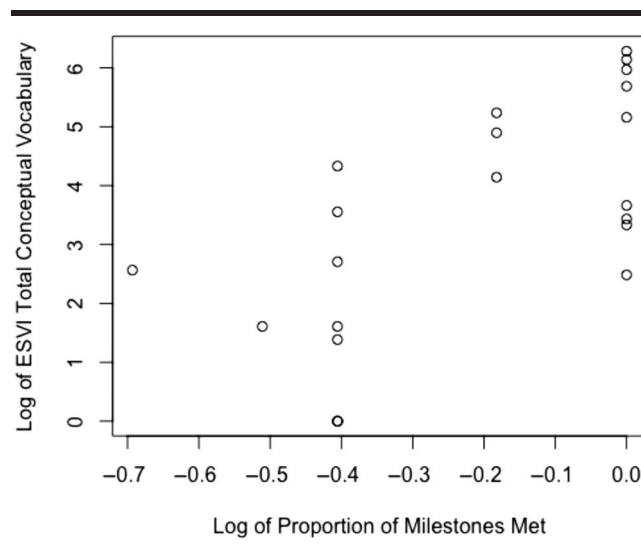
many concepts across Spanish and English will also demonstrate strong overall language skills across comprehension and production. Children's raw scores on the PLS-5 did not meet the normality assumptions for linear regression, and therefore, the log score was calculated as with all vocabulary scores on the ESVI-E. Given the use of raw scores and the relatively wide age range of participants in Study 3, age was included as a covariate in the model. Results indicated a significant association between ESVI-E TCV and TV estimates and raw scores on the PLS-5 Expressive subtest: TCV, $F(2, 19) = 18.74, p < .001, R^2 = .664$; TV, $F(2, 19) = 18.38, p < .001, R^2 = .659$. Similarly, ESVI-E TCV and TV predicted raw scores on the Receptive subtest of the PLS-5: TCV, $F(2, 19) = 17.86, p < .001, R^2 = .653$; TV, $F(2, 19) = 17.38, p < .001, R^2 = .647$.

To further assess criterion validity, we examined whether the ESVI-E predicted caregiver report of age-appropriate developmental milestones. The distribution of developmental milestones did not meet the assumptions of normality, and log transformations, once more, were conducted. In addition, age was included as a covariate to parallel the PLS-5 analyses. Regression results showed that TCV on the ESVI-E was a significant predictor of the amount of age-appropriate developmental milestones that have been met, $F(1, 20) = 8.174, p = .002, R^2 = .463$ (see Figure 2). Results using TV on the ESVI-E also yielded a significant model with similar variance explained, $F(1, 20) = 8.497, p = .002, R^2 = .472$.

Social Validity

Table 5 provides a summary of responses from caregivers. All caregivers (100%) agreed that they were interested in capturing their child's vocabulary over time, felt it

Figure 2. Association between total conceptual vocabulary and proportion of developmental milestones met. ESVI = English-Spanish Vocabulary Inventory.



was “easy” to record their responses and track changes, and overall felt confident that their responses on the measure represented their child’s vocabulary in both English and Spanish. All but two parents (91%) agreed that the words their child knew or said were listed on the questionnaire, though six parents (27%) agreed that they remembered additional words their child knew or said after completing the questionnaire. When asked to list the additional words, eight caregivers (36%) listed a combined total of 34 items that they reported were not included in the ESVI-E. Approximately 17 were indeed listed on the ESVI-E, eight items were lexical items that were not included, and nine were phrases or sentences rather than discrete lexical items.

Interim Summary

Study 3 contributed unique findings regarding the concurrent criterion validity of the ESVI-E. Results showed that parent report of total expressive conceptual vocabulary size across Spanish and English on the ESVI-E significantly predicted children’s performance on the PLS-5 and the CDC Developmental Milestones Checklist. The results from Study 3 also further extended the social validity findings from Study 2 to a second geographically distinct group of families. Similar to the sample of parents in Study 2, parents in Study 3 reported feeling confident in the ESVI-E measuring the words their child could say. Additional feedback from some parents suggests opportunities to further revise the ESVI-E.

Discussion

There remains a dearth of research examining vocabulary instruments that are specifically adapted for DLL infants and toddlers with and without early language delays. As we reviewed in the introduction, many extant approaches for vocabulary assessment of DLLs using parent report follow best practice guidelines in that both languages are measured and cross-language measures are derived (TV and TCV). However, each language is typically measured separately using methods that were developed and intended for monolingual learners, thereby limiting the precision and validity of current instruments in assessing vocabulary in the DLL context. This study therefore introduces an adaptation of extant approaches by combining word lists across languages and including TEs: the ESVI, Expressive and Expressive–Receptive versions. The primary aim of this study was to establish preliminary evidence regarding the psychometric properties of the ESVI when employed with young Spanish and English learners with and without early language delays. Across three distinct samples of Spanish–English DLLs, including those with language disabilities, we showed promising results that indicated the ESVI-E and the ESVI-ER stand to contribute to the

measurement of early child language in research and clinical practice. Below, we review key findings for each construct of interest and provide recommendations.

Content Validity and Criterion Validity

Results across Studies 1 and 2 showed that parent report of expressive and receptive vocabulary on the ESVI is positively associated with direct observations of child language skills, including in children with identified language concerns. In Study 1, the number of words caregivers reported their child could say on the ESVI significantly predicted children’s accuracy score on a Spanish and English receptive vocabulary behavioral task (the CCT). Similarly, expressive TV and TCV as reported on the ESVI predicted the number of different and total words children produced when interacting with their caregiver during a short 10-min free-play observation. The association held such that expressive TV and TCV on the ESVI predicted children’s nonword vocalizations, as well. Indeed, early vocalizations predict vocabulary outcomes (e.g., Donnellan, et al., 2020). Similarly, in Study 2, home language recordings across multiple hours revealed that parent report of expressive vocabulary (i.e., ESVI TV scores) predicts rate of vocalizations on LENA home language recordings in a sample of children with language delays. Together, these results suggest that Spanish and English expressive vocabulary as measured on the ESVI predicts accuracy on a word identification task and rate of vocalization across lab and home settings in children with and without early language delays. The findings from receptive vocabulary were similar but not statistically significant likely because the association between receptive vocabulary and expressive verbalizations is somewhat attenuated compared to expressive vocabulary. Findings from Study 3 further showed that vocabulary size indeed predicts other related outcomes as expected. Parent reports of expressive vocabulary on the ESVI predicted total and expressive standard scores on a language assessment (the PLS-5 Spanish) and also whether children met general developmental milestones.

Taken together, the content and criterion validity findings followed our expectations. We know from a large body of work that parent report is a widely used and valid method for capturing vocabulary (e.g., Marchman & Martínez-Sussman, 2002). The present results further strengthen the validity of parent-reported vocabulary for DLL children. We also know that vocabulary is representative of early communication skills more broadly, and our findings show that the ESVI expressive scores do indeed closely relate to other indicators of child expressive communication. Furthermore, across all analyses, words from the CDI and the IDHC alone did not predict variance above and beyond that captured by the ESVI, but rather the ESVI performed comparably. However, results from Study 1 showed that the

CDI and IDHC monolingual approach to dual language assessment undercount key cross-language measures such as TEs and effect sizes showed the difference was moderate. Together, the results show that the additional words included in the ESVI to capture cross-language knowledge and that our approach of presenting words side by side does not compromise the content validity or criterion validity of parent-reported vocabulary. This is an important finding as the addition of TEs provides greater specificity in the assessment of DLLs' vocabulary skills.

Construct Validity and Tracking Change

Children with parent-reported or confirmed language and communication concerns also reported significantly smaller expressive TVs, TCVs, TEs, Spanish vocabulary, and English vocabulary on the ESVI in Study 1—an effect that was most robustly observed in cross-language measures (TV, TCV, and TEs). This finding is consistent with assessment recommendations that underscore the importance of measuring both languages when assessing children at risk for language impairment (e.g., Mancilla-Martinez & Vaughn, 2013). Furthermore, the fact that the difference between children with and without caregiver language and communication concerns was largest in TV also replicates the finding that this variable is better than TCV in identifying children potentially at risk for language impairment (Core et al., 2013). Similarly, the ESVI can capture meaningful changes in receptive and expressive TCV and vocabulary size in the dominant language (in this case Spanish) over a 2-month period in children with early language delays as shown in Study 2. Future studies must examine children with more balanced exposure and with larger vocabulary sizes to replicate findings across all vocabulary measures and to establish test–retest reliability. This suggests that the ESVI can be used to track vocabulary in a subset of Spanish–English DLLs who are experiencing atypical development and with known language delays. As with the content validity findings, construct validity results show that the ESVI conforms to the expected pattern of results and provide preliminary support for the use of the ESVI in clinical contexts.

Social Validity

To our knowledge, this is the first study to report on the social validity of parent-reported vocabulary in a DLL context for children with and without language delay. Given that the ESVI relies entirely on caregiver report, it is important to examine whether the questionnaire is perceived as a valid and practical measure of Spanish and English vocabulary by caregivers themselves as a preliminary step. The social validity findings suggest that families find the ESVI useful but that there are additional

revisions needed to improve the tool. Specifically, findings from Studies 2 and 3 showed that parents with and without concerns over their child's language development felt confident in reporting words their child could comprehend and those they could say in English and Spanish. In Study 3, caregivers of Spanish–English DLLs further reported their interest in capturing their child's language skills. This is consistent with extant research, which reports that Latino Mexican families in the United States value their children's early communication development and employ several strategies toward this end (Cycyk & Hammer, 2020). Together, these findings lend further support to the appropriateness of parent report on this aspect of child language development. At the same time, it is important to note that the social validity questionnaires used in this study have not been previously published and their lack of broad validation may limit the rigor of the results. Social desirability also could have played a role in the social validity findings, and we return to these limitations in discussing future directions below. However, the pattern of results across the three studies and across the several parent report questionnaires suggests that caregivers of Spanish–English bilingual children are generally reliable in their responses.

In terms of practicality of the tool, parents believed it was easy to record their responses on the questionnaire, which presented TEs in Spanish and English side by side. Notably, the average time to completion reported by parents in Study 1 was slightly shorter than the estimated time required to complete two separate measures, and for some parents, the time to completion was substantially reduced. The addition of TEs did not appear to impact participant burden despite the increased number of words on the ESVI as compared to the separate versions of the CDI and the IDHC. Therefore, this method of combining existing monolingual CDI adaptations into a single dual language questionnaire may be useful to parents of other populations and for other language combinations as well, though their validity and reliability should be examined independently.

The social validity results from Studies 2 and 3 also point toward revisions to the ESVI that may strengthen its usefulness. Caregivers reported recruiting other family members to help complete the form as they filled out the questionnaire at home. The inclusion of additional family members may indicate that DLLs were often learning different languages from different people with different proficiencies in Spanish and English and that multiple caregivers played a significant role in the child's language experience. This observation is consistent with the experiences of many Spanish–English DLLs in the United States and with some extant practices (e.g., Marchman & Martínez-Sussman, 2002). For example, in one study of Latinx Mexican families in the United States, caregivers reported a focus on tight-knit families that collectively supported the child's growth (Cycyk & Hammer, 2020).

In light of these findings, we recommend that users of the ESVI avoid assuming that any one caregiver can report on the child's words. In general, best practice for language assessment in bilingual contexts involves engaging multiple reporters for data collection (i.e., for a full discussion on considerations in the use of the CDI in bilingual contexts, see De Houwer, 2019). Instead, users of the ESVI should first discuss who might best report on each of the child's languages based on caregiving role and Spanish and English child language input in order to systematically ensure an accurate description of all of the child's languages. In this study, caregivers were provided with the combined English and Spanish word list from the ESVI, which was completed collaboratively with other informants in some cases. We favor and recommend such a bilingual approach over providing the English-only or Spanish-only words to different informants because many of the informants themselves are bilingual. In addition, although it is not clear how collaborative versus independent administrations across raters can influence the reliability of vocabulary estimates, evidence suggests that raters are generally reliable with each other, including in neurodiverse child populations (i.e., Nordahl-Hansen et al., 2013). In addition, several parents in Study 3 provided a list of words that they believed were not included on the ESVI. Given that some of the examples parents provided indeed had equivalents on the ESVI, these findings suggest that the ESVI instructions could be improved for clarity while also still allowing room for parents to report additional words known by their child. In addition, this suggests further adaptations should prioritize reducing the length of the questionnaire.

Summary and Future Directions

This study demonstrates preliminary support for the use of the ESVI in research and clinical contexts with Spanish–English DLLs where precise measures of within- and cross-language knowledge are especially needed. The ESVI departs from extant approaches to parent-reported vocabulary by presenting TEs for all words in Spanish and English side by side. The ESVI showed acceptable levels of content validity, criterion validity, construct validity, and social validity across three distinct groups of Spanish–English learners in the Pacific Northwest and Western regions of the United States. Across studies, more than half of participants had parent-reported or confirmed language delays, thereby demonstrating promising utility of the ESVI in clinical contexts. Furthermore, across analyses, the measures derived from the monolingual instruments did not explain additional variance beyond that captured by the ESVI. Taken together, the findings suggest that the ESVI provides a valid and reliable approach for assessing vocabulary in Spanish–English learners with and without early language delays.

In clinical contexts, we remind practitioners that although parent report confers several advantages and contributes meaningfully to evaluation procedures, it must be used in conjunction with other measures to support effective triangulation of assessment findings and accurate diagnosis of DLLs (Castilla-Earls et al., 2020; De Anda et al., 2020). Scores resulting from the ESVI should therefore be used primarily for descriptive purposes to support early language assessment (e.g., to describe the child's relative vocabulary knowledge within and across languages) and to monitor progress in vocabulary growth over time. Continued research with larger samples of diverse Spanish–English DLLs is needed to replicate findings, to further evaluate the utility of the measure, and to determine its validity with DLL populations. Examining the role of exposure and vocabulary in each language will also help understand the role of input and lexical development in DLLs. The findings of this study provide preliminary evidence of promise of the ESVI for measuring development, suggesting that this effort would be worthwhile.

In addition, this study provides potential avenues for further refinement of assessment approaches for DLLs learning languages other than Spanish and English. First, the ESVI approach of combining the monolingual CDI questionnaires and adding TEs may potentially be adapted to other languages and to multilingual contexts. Second, multistudy designs such as the one presented here may afford a short-term solution to advancing assessment research for the multilingual children with language delays. Across all three studies presented here, more than half of the children had parent-reported and/or confirmed developmental concerns. This represents a robust sample of underrepresented children with atypical early language in which we were able to detect differences by language status (Study 1) while also demonstrating that the tool aligned with other key outcome measures among the children with language concerns (Studies 2 and 3). Part of the difficulty in developing assessment measures for DLLs stems from challenges to recruiting large and representative samples of neurodiverse children, including typically developing learners and children with suspected and/or known developmental disabilities in underserved and vulnerable communities. Collaborative partnerships and utilizing participants across geographical locations can minimize these challenges. Thus, although the three studies presented here contribute significantly to the extremely limited research base of Spanish–English DLL toddlers with early language delays, future research must conduct more rigorous studies with larger sample sizes. Third, this study included many families facing economic hardship, with the results suggesting the ESVI can be used with such communities. Nevertheless, future studies should examine the role of sociodemographic factors (maternal education, income, etc.) on vocabulary outcomes by recruiting a heterogeneous sample of young Spanish learners.

Although the ESVI is a significant departure from the status quo that uses measures intended for monolinguals in DLL contexts, there remain some challenges that limit its applicability. Further research on the social validity of the ESVI is warranted given that social desirability could have played a role in responses. Including a measure of overall positive bias may be useful in future research that aims to improve the social validity of the ESVI for caregivers and the use of anonymous responses. Inclusion of key stakeholders early in the development process would also be beneficial, perhaps through cognitive interviewing to confirm that the social validity surveys indeed measure what they are intended to measure (Willis, 2005). In addition, the ESVI included words from Mexican (derived from the IDHC) and presumably Standardized American English (derived from the CDI) dialects. Future extensions should consider further refinements based on language varieties within both English and Spanish. Our participants were primarily families of Mexican descent living in the United States and were not representative of the entire U.S. Spanish-speaking population. Furthermore, due to the differences between the ESVI and the CDI instruments, the available CDI norms should not be interpreted as applying to the ESVI at the level of individual children. Thus, a future large-scale norming effort of the ESVI is needed to replicate the validity and reliability of the measure and to provide relevant descriptive criteria for the purposes of assessment. Moreover, caregivers in Study 2 were provided with their prior questionnaire and asked to mark changes. This could artificially inflate correlations in this analysis while also potentially discouraging parents from reporting words their child no longer knew or said in the case of regression. Future research should evaluate test-retest reliability with independent administrations. Similarly, examining the diagnostic accuracy of the ESVI (i.e., sensitivity and specificity) will be useful as measures for identifying DLL children at risk for later language impairment are relatively scarce. Lastly, administration time could be significantly reduced to improve efficiency. As such, efforts to reduce the number of words on the questionnaire may prove especially useful for clinical contexts in which time allotted for assessments of multilingual children is already limited. Previous research examining short forms of the IDHC have shown promise for use with toddlers from Spanish-speaking homes (Guiberson et al., 2011). Future directions include development of a brief version of the ESVI.

Materials

This article described the use of two versions of the ESVI: one that examines only expressive vocabulary (ESVI-E; preferable for children with some expressive language skills) and the other that captures both expressive

and receptive vocabulary (ESVI-ER; preferable for children with limited or absent expressive skills). The questionnaires and their respective scoring tools can be downloaded at <https://edldlab.uoregon.edu/> or by contacting the first author.

Acknowledgments

This work was supported in part by National Institutes of Health Grant K23DC018033 awarded to Stephanie De Anda, a grant from the ASHFoundation awarded to Lauren M. Cyczk, and funding from Utah State University's Office of Research awarded to Anne L. Larson. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors are grateful to the families who participated and the students at the University of Oregon's Early Dual Language Development Lab and the Communication Across Languages and Modalities Lab at Utah State University who facilitated data collection and processing. They also gratefully acknowledge the permission of the MacArthur-Bates Communicative Development Inventories (CDI) Advisory Board to use the vocabulary lists for the CDI and the *Inventario de Desarrollo de Habilidades Comunicativas* as the foundation of the English-Spanish Vocabulary Inventory.

References

- American Speech-Language-Hearing Association. (2021). *Demographic profile of ASHA members providing bilingual services, year-end, 2020*. <https://www.asha.org/siteassets/surveys/demographic-profile-bilingual-spanish-service-members.pdf>
- Andersson, L. (2005). Determining the adequacy of tests of children's language. *Communication Disorders Quarterly*, 26(4), 207–225. <https://doi.org/10.1177/15257401050260040301>
- 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>
- Castilla-Earls, A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Peña, E. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American Journal of Speech-Language Pathology*, 29(3), 1116–1132. https://doi.org/10.1044/2020_AJSLP-19-00179
- Centers for Disease Control and Prevention. (n.d.). *CDC's developmental milestones*. <https://www.cdc.gov/ncbddd/actearly/milestones/index.html>
- 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](https://doi.org/10.1044/1092-4388(2013)11-0044)
- Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in

- focus. *Journal of Phonetics*, 39(4), 492–504. <https://doi.org/10.1016/j.wocn.2010.12.002>
- Cycyk, L. M., & Hammer, C. S. (2020). Beliefs, values, and practices of Mexican immigrant families towards language and learning in toddlerhood: Setting the foundation for early childhood education. *Early Childhood Research Quarterly*, 52(A), 25–37. <https://doi.org/10.1016/j.ecresq.2018.09.009>
- Cycyk, L. M., Moore, H. W., De Anda, S., Huerta, L., Mendez, S., Patton, C., & Bourret, C. (2020). Adaptation of a caregiver-implemented naturalistic communication intervention for Spanish-speaking families of Mexican immigrant descent: A promising start. *American Journal of Speech-Language Pathology*, 29(3), 1260–1282. https://doi.org/10.1044/2020_AJSLP-19-00142
- De Anda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The Language Exposure Assessment Tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research*, 59(6), 1346–1356. https://doi.org/10.1044/2016_JSLHR-L-15-0234
- De Anda, S., & Friend, M. (2020). Lexical-semantic development in bilingual toddlers at 18 and 24 months. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.508363>
- De Anda, S., Hendrickson, K., Zesiger, P., Poulin-Dubois, D., & Friend, M. (2018). Lexical access in the second year: A study of monolingual and bilingual vocabulary development. *Bilingualism: Language and Cognition*, 21(2), 314–327. <https://doi.org/10.1017/s1366728917000220>
- De Anda, S., Larson, A., & Cycyk, L. M. (2020). Considerations in the evaluation and assessment process for Latinx infants and toddlers. Division for Early Childhood (DEC) recommended practices. *Monograph Series (No. 7 Assessment)*, 7, 25–40. <https://www.dec-sped.org/product-page/dec-recommended-practices-monograph-series-no-7-assessment>
- De Houwer, A. (2019). Equitable evaluation of bilingual children's language knowledge using the CDI: It really matters who you ask. *Journal of Monolingual and Bilingual Speech*, 1(1), 32–54. <https://doi.org/10.1558/jmbs.11184>
- Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science*, 23(1), e12843. <https://doi.org/10.1111/desc.12843>
- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, 56(8), 848–856. <https://doi.org/10.1111/jcpp.12378>
- Ellis, E. M., & Thal, D. J. (2008). Early language delay and risk for language impairment. *SIG 1 Perspectives on Language Learning and Education*, 15(3), 93–100. <https://doi.org/10.1044/lle15.3.93>
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, S., & Bates, E. (2007). *The MacArthur Communicative Development Inventories—Second Edition (MCIDI-2): User's guide and technical manual*. Brookes.
- Floccia, C., Sambrook, T. D., 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. (2018a). II: Methods. *Monographs of the Society for Research in Child Development*, 83(1), 30–42. <https://doi.org/10.1111/mono.12349>
- Floccia, C., Sambrook, T. D., 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. (2018b). III: Analyses and results for Study I: Estimating the effect of linguistic distance on vocabulary development. *Monographs of the Society for Research in Child Development*, 83(1), 43–60. <https://doi.org/10.1111/mono.12350>
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. <https://doi.org/10.1017/S0305000916000209>
- Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, & Computers*, 35, 302–309. <https://doi.org/10.3758/BF03202556>
- 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>
- Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the Computerized Comprehension Task: Comprehension predicts production. *Developmental Psychology*, 48(1), 136–148. <https://doi.org/10.1037/a0025511>
- Friend, M., & Zesiger, P. (2011). Une réplification systématique des propriétés psychométriques du Computerized Comprehension Task dans trois langues [A systematic replication of the psychometric properties of the Computerized Comprehension Task in three languages]. *Enfance*, 2011, 329–344. <http://doi.org/10.4074/S0013754511003053>
- Goodrich, J. M., Lonigan, C. J., Kleuver, C. G., & Farver, J. M. (2016). Development and transfer of vocabulary knowledge in Spanish-speaking language minority preschool children. *Journal of Child Language*, 43(5), 969–992. <http://doi.org/10.1017/S030500091500032X>
- Grasso, S. M., Peña, E. D., Bedore, L. M., Hixon, J. G., & Griffin, Z. M. (2018). Cross-linguistic cognate production in Spanish–English bilingual children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 61(3), 619–633. https://doi.org/10.1044/2017_JSLHR-L-16-0421
- Guiberson, M., Rodriguez, B. L., & Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Language, Speech, and Hearing Services in Schools*, 42(4), 536–549. [https://doi.org/10.1044/0161-1461\(2011\)10-0076](https://doi.org/10.1044/0161-1461(2011)10-0076)
- Hammer, C. S., Cycyk, L. M., Scarpino, S. E., Jury, K. A., & Sawyer, B. E. (2020). Development of the CECER-DLL child and family questionnaire: A new tool for documenting the language and literacy experiences of Latino/a dual language learners. *International Journal of Bilingual Education and Bilingualism*, 1–23. <https://doi.org/10.1080/13670050.2020.1840507>
- Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A. (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, 17(1), 189–202. <http://doi.org/10.1017/S136672891300014X>
- Jackson, C. W., & Callender, M. F. (2014). Environmental considerations: Home and school comparison of Spanish–English speakers' vocalizations. *Topics in Early Childhood Special Education*, 34(3), 165–174. <http://doi.org/10.1177/0271121414536623>
- Jackson-Maldonado, D., Thal, D., Marchman, V., Newton, T., Fenson, L., & Conboy, B. (2003). *Communicative Development Inventory: User's guide and technical manual*. Brookes.
- Jackson-Maldonado, D., Thal, D., Marchman, V., Newton, T., Fenson, L., & Conboy, B. (2005). *Inventarios MacArthur-Bates del desarrollo de habilidades comunicativas: User's guide and technical manual* [MacArthur-Bates Communicative Development Inventories: User's guide and technical manual]. Brookes.
- Jin, F., Schjolberg, S., Wang, M. V., Eadie, P., Nes, R. B., Roysamb, E., & Tambs, K. (2020). Predicting literacy skills at 8 years from preschool language trajectories: A population-based cohort

- study. *Journal of Speech, Language, and Hearing Research*, 63(8), 2752–2762. https://doi.org/10.1044/2020_JSLHR-19-00286
- King, M., Larson, A., & Buzhardt, J.** (2021). Exploring the classification accuracy of the early communication indicator (ECI) with dual-language learners from Latinx backgrounds. *Assessment for Effective Intervention*. Advance online publication. <https://doi.org/10.1177%2F15345084211027138>
- Langdon, H., & Saenz, T.** (2015). *Working with interpreters and translators: A guide for speech-language pathologists and audiologists*. Plural.
- Larson, A. L.** (2016). Language screening for infants and toddlers: A literature review of four commercially available tools. *Communication Disorders Quarterly*, 38(1), 3–12. <http://doi.org/10.1177/1525740115627420>
- Lausberg, H., & Sloetjes, H.** (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, 41(3), 841–849. <http://doi.org/10.3758/BRM.41.3.841>
- Lee, J.** (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied PsychoLinguistics*, 32(1), 69–92. <https://doi.org/10.1017/S0142716410000299>
- Mancilla-Martínez, J., Gámez, P. B., Vagh, S. B., & Lesaux, N. K.** (2016). Parent reports of young Spanish–English bilingual children’s productive vocabulary: A development and validation study. *Language, Speech, and Hearing Services in Schools*, 47(1), 1–15. https://doi.org/10.1044/2015_LSHSS-15-0013
- Mancilla-Martínez, J., & Vagh, S. B.** (2013). Growth in toddlers’ Spanish, English, and conceptual vocabulary knowledge. *Early Childhood Research Quarterly*, 28(3), 555–567. <https://doi.org/10.1016/j.ecresq.2013.03.004>
- 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(5), 983–997. [https://doi.org/10.1044/1092-4388\(2002\)080](https://doi.org/10.1044/1092-4388(2002)080)
- Miller, J., & Iglesias, A.** (2018). *Systematic Analysis of Language Transcripts (SALT) (Research Version 2010)* [Computer software]. SALT Software.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W.** (2010). The COSMIN study reached international consensus on taxonomy, terminology and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Nordahl-Hansen, A., Kaale, A., & Ulvund, S. E.** (2013). Inter-rater reliability of parent and preschool teacher ratings of language in children with autism. *Research in Autism Spectrum Disorders*, 7(11), 1391–1396. <https://doi.org/10.1016/j.rasd.2013.08.006>
- Oh, M. H., & Mancilla-Martínez, J.** (2021). Comparing vocabulary knowledge conceptualizations among Spanish–English dual language learners in a new destination state. *Language, Speech, and Hearing Services in Schools*, 52(1), 369–382. https://doi.org/10.1044/2020_LSHSS-20-00031
- Paradis, J., Emmerzael, K., & Duncan, T. S.** (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43(6), 474–497. <https://doi.org/10.1016/j.jcomdis.2010.01.002>
- Park, M., O’Toole, A., & Katsificas, C.** (2017). *Dual language learners: A national demographic and policy profile*. Fact Sheet. Migration Policy Institute (ED589153). ERIC. <https://eric.ed.gov/?id=ED589153>
- Patterson, J. L.** (1998). Expressive vocabulary development and word combinations of Spanish–English bilingual toddlers. *American Journal of Speech-Language Pathology*, 7(4), 46–56. <https://doi.org/10.1044/1058-0360.0704.46>
- Patterson, J. L.** (2000). Observed and reported expressive vocabulary and word combinations in bilingual toddlers. *Journal of Speech, Language, and Hearing Research*, 43(1), 121–128. <https://doi.org/10.1044/jslhr.4301.121>
- Pearson, B. Z., Fernández, S. C., & Oller, D. K.** (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43(1), 93–120. <https://doi.org/10.1111/j.1467-1770.1993.tb00174.x>
- Peña, E. D.** (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, 78(4), 1255–1264. <https://doi.org/10.1111/j.1467-8624.2007.01064.x>
- Place, S., & Hoff, E.** (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, 19(5), 1023–1041. <https://doi.org/10.1017/S1366728915000322>
- Rescorla, L.** (1989). The language development survey. *Journal of Speech and Hearing Disorders*, 54(4), 587–599. <https://doi.org/10.1044/jshd.5404.587>
- Restrepo, M. A.** (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, 41(6), 1398–1411. <https://doi.org/10.1044/jslhr.4106.1398>
- Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., & Lamb, M. E.** (2004). Fathers and mothers at play with their 2- and 3-year-olds: Contributions to language and cognitive development. *Child Development*, 75(6), 1806–1820. <https://doi.org/10.1111/j.1467-8624.2004.00818.x>
- U.S. Department of Health and Human Services, & U.S. Department of Education.** (2017). *Policy statement on supporting the development of children who are dual language learners in early childhood programs (Log No. ODAS, ECD-ACF-PS- 2017-02)*. Office of the Deputy Assistant Secretary for Early Childhood Development.
- Weisleder, A., & Fernald, A.** (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177%2F0956797613488145>
- Weismer, S. E., & Evans, J. L.** (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders*, 22(3), 15–29. <https://doi.org/10.1097/00011363-200205000-00004>
- Willis, G. B.** (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage. <https://doi.org/10.4135/9781412983655>
- Wood, C., Diehm, E. A., & Callender, M. F.** (2016). An investigation of language environment analysis measures for Spanish–English bilingual preschoolers from migrant low-socioeconomic-status backgrounds. *Language, Speech, and Hearing Services in Schools*, 47(2), 123–134. https://doi.org/10.1044/2015_LSHSS-14-0115
- Xu, D., Yapanel, U., & Gray, S.** (2009). *Reliability of the LENA language environment analysis system in young children’s natural home environment*. LENA Foundation.
- Zimmerman, I. L., Steiner, V. G., & Pond, E.** (2011). *Preschool Language Scales–Fifth Edition (PLS-5)*. Pearson.
- Zimmerman, I. L., Steiner, V. B., & Pond, R. E.** (2012). *Preschool Language Scales–Fifth Edition Spanish (PLS-5 Spanish): Administration and scoring manual*. Pearson.