

Psychometric Evaluation of the Bilingual English-Spanish Assessment Sentence Repetition Task  
for Clinical Decision-Making

Lisa Fitton<sup>1</sup>

Rachel Hoge<sup>2</sup>

Yaacov Petscher<sup>3</sup>

Carla Wood<sup>2</sup>

<sup>1</sup>Communication Sciences and Disorders, University of South Carolina, Columbia, SC

<sup>2</sup>Communication Science and Disorders, Florida State University, Tallahassee, FL

<sup>3</sup>College of Social Work & the Florida Center for Reading Research, Florida State University,  
Tallahassee, FL

Correspondence concerning this article should be addressed to Lisa Fitton, Department of Communication Sciences and Disorders, Arnold School of Public Health, University of South Carolina, 1229 Marion Street Second level, Columbia, SC 29201. Phone: (517)614-7264. Email: FITTONL@mailbox.sc.edu

**Conflict of Interest**

The authors have no relevant conflicts of interest to disclose.

**Funding**

The research reported here was supported by the Institute of Education Sciences, U. S. Department of Education through Grant R305A130460 to Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U. S. Department of Education.

## DIMENSIONALITY OF SENTENCE REPETITION

### **Abstract**

**Purpose:** The purpose of the present study was a) to examine the underlying components or factor structure of the Bilingual English-Spanish Assessment (Peña et al., 2014) sentence repetition task and b) to examine the relationship between Spanish-English speaking children's sentence repetition and vocabulary performance.

**Method:** Participants were 291 Spanish-English speaking children in kindergarten and first grade. Item analyses were used to evaluate the underlying factor structure for each language version of the sentence repetition tasks of the BESA. The tasks were then examined in relation to a measure of English receptive vocabulary.

**Results:** Bifactor models, which include a single underlying general factor and multiple specific factors, provided the best overall model fit for both the Spanish and English versions of the task. There was no relation between children's overall Spanish sentence repetition performance and their English vocabulary. However, children's pronoun, noun phrase, and verb phrase item scores in Spanish significantly predicted their English vocabulary scores. For English sentence repetition, both children's overall performance and their specific performance on the noun phrase items were predictors of their English vocabulary scores. Follow-up analyses revealed that, for the purposes of clinical assessment, the BESA sentence repetition tasks can be considered essentially unidimensional, lending support to the current scoring structure of the test.

**Conclusions:** Study findings suggest that sentence repetition tasks can provide insight into Spanish-English speaking children's vocabulary skills in addition to their morphosyntactic skills when used on a broad research scale. From a clinical assessment perspective, results indicate that the sentence repetition task has strong internal validity and support to the use of this measure in clinical practice.

# DIMENSIONALITY OF SENTENCE REPETITION

## Psychometric Evaluation of the Bilingual English-Spanish Assessment for Clinical Decision-Making

Spanish-English speaking dual language learners (DLLs) are one of the fastest growing populations in the U.S. (Kena et al., 2014). More than half of the U.S. population growth between 2000 and 2010 can be attributed to an increase in the Hispanic population (Passel, Cohn, & Lopez, 2011). DLLs are at heightened risk for poor literacy and academic achievement compared to their monolingual English-speaking peers (Hemphill & Vanneman, 2011). Consequently, this population growth has resulted in a demand for educational resources to support DLLs effectively (Kena et al., 2015). Specifically, there is need for valid and reliable assessment tools to facilitate the accurate diagnosis of DLLs with language impairment (LI).

Assessment is a foundational component of effective educational practice and can strongly impact children's educational experiences. In the present study, we focus on one diagnostic tool that was specifically designed for young Spanish-English DLLs: The *Bilingual English-Spanish Assessment* (BESA; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). To assist practitioners in maximizing the information obtained from the BESA for clinical decision-making and educational planning, we evaluated the psychometric properties of the sentence repetition task on the BESA. We examined the underlying dimensionality and predictive validity of the Spanish and English versions of the task at the item level. In the following literature review, we discuss the critical role of clinical assessment in Spanish-English DLLs' education, describe key features of the BESA and the sentence repetition task, and explain our approach to evaluating the sentence repetition items in Spanish and English.

### **Importance of Clinical Language Assessment of DLLs**

DLLs are at heightened risk for misidentification as having LI compared to their

## DIMENSIONALITY OF SENTENCE REPETITION

monolingual peers (Hamayan, Marler, Sanchez-Lopez, & Damico, 2007; MacSwan & Rolstad, 2006; Sullivan, 2011). Disproportionately-high poverty rates among bilingual families in the U.S. (Murphey, 2014), heterogeneity in dual language exposure and proficiency (Peña, Bedore, & Kester, 2015), and cultural differences (Reynolds & Suzuki, 2013) complicate the diagnostic process. There are also relatively few norm-referenced assessments designed specifically to identify bilingual children with LI. Assessment practices that do not take the unique characteristics and experiences of bilingual children into account lead to both over- and under-identification of disabilities among DLLs (Kohnert, 2010).

DLLs tend to be over-identified as having LI when practitioners use assessment tools that are not normed or vetted for identifying children from culturally and linguistically diverse backgrounds. For example, a tool designed for use with monolingual children may identify linguistic differences in DLLs as markers of impairment (Gutiérrez-Clellen & Simon-Cerejido, 2007; Montrul, Davidson, de la Fuente, & Foote, 2014). The skills that children have across both languages are often not apparent when assessment is only conducted in one language (Bedore, Peña, Gillam, & Ho, 2010; Gathercole, Thomas, Roberts, Hughes, & Hughes, 2013). Further, tests whose items are not examined carefully for cultural or linguistic bias may also disproportionately identify DLLs as having impairment (Abedi, 2006). Over-identification is problematic because it can place unneeded stress on children, their families, and service providers' time.

However, DLLs also experience heightened rates of under-identification, particularly in the lower elementary school grades. Findings from Artiles, Rueda, Salazar, and Higuera (2002) showed that children learning English as a second language were underrepresented in special education services in elementary school, but they were significantly overrepresented in the

## DIMENSIONALITY OF SENTENCE REPETITION

secondary grades with the highest rates being among students with the lowest English proficiency. This finding was replicated by Samson and Lesaux (2009). Inadequate professional training and differences in policy guidelines regarding the referral process can be contributors to systematic under-identification. For example, researchers have consistently observed the tendency of educators to adhere to a wait-and-see period allotting DLLs more time to acquire English before considering referral for a formal language evaluation despite low achievement levels (Samson & Lesaux, 2009; Sanchez, Parker, Akbayin, & McTigue, 2010). Thus, educators and service providers with more limited expertise in bilingual language development may adopt conservative approaches for identifying bilingual children with LI, leading to children being deprived of services and consequently needing more support in the later grades.

### **The BESA as a Clinical Tool**

The *Bilingual English-Spanish Assessment* (BESA) is a clinical test designed by Peña and colleagues (2014) to assist in the accurate diagnosis of Spanish-English bilingual children with speech and/or language impairment. The measure includes three subtests with separate versions in Spanish and English. The subtests target children's phonological, morphosyntactic, and semantic abilities, and were constructed to elicit skills that are reliable indicators of speech and language impairment in Spanish-English speaking DLLs. Children's performance on each of the included tasks are aggregated to create a composite Language Index, which is then used to identify children with language impairment. Unlike most available standardized language assessments, the BESA was normed on a sample of bilingual children with a broad range of proficiencies in Spanish and English ( $n = 756$ ; Peña et al., 2014). The tool was also designed to include only culturally-appropriate items for Spanish-English speaking children. This careful attention to content validity establishes the BESA as a theoretically-sound assessment for

## DIMENSIONALITY OF SENTENCE REPETITION

Spanish-English speaking DLLs (Bedore et al., 2010; Gathercole et al., 2013; Kohnert, 2010).

The morphosyntax subtest is comprised of a cloze task and the sentence repetition task examined in this paper. The subtest was designed to elicit specific grammatical features of Spanish and English. Both tasks include prompts for children to produce targeted grammatical structures in an obligatory context. The inclusion of two tasks within the subtest allows for multiple opportunities to elicit the grammatical forms in each language (Peña et al., 2014). Each task is scored individually, yielding separate scaled scores for sentence repetition and cloze. These scaled scores are then summed and converted to a standardized score for morphosyntax.

For the sentence repetition task of the BESA, the examiner instructs the child to repeat target sentences verbatim. Instead of receiving a score based on the number of errors or the number of words correctly imitated, as is typically done for sentence repetition tasks (Kapantzoglou, Thompson, Gray, & Restrepo, 2016; Klem et al., 2015; Komeili & Marshall, 2013; Pawlowska, 2014), children are scored based on their productions of the designated target items within each sentence. Target items include one to three words (e.g., “the dog”) but are dichotomously scored as either 1 (*correct*) or 0 (*incorrect*). The child receives one point for repeating an item exactly as prompted and zero points for omissions or errors in production not attributable to dialect or speech impairment. For example, the response of “is cry” or “are crying” for the target item “is crying” would receive a zero score. Similarly, in Spanish the response “tiene hambre” for the target “tenía hambre” would receive a zero score for that item. Children are required to respond in the language being assessed to receive credit for accurate repetition. The Spanish sentence repetition task consists of 37 target items embedded within 10 sentences. The English version consists of 33 target items within 9 sentences (Peña et al., 2014).

In contrast to most other language elicitation measures, sentence repetition tasks such as

## DIMENSIONALITY OF SENTENCE REPETITION

those included on the BESA are administered and scored quickly and easily, thus requiring minimal experience and training for valid implementation. Because targets are clearly specified and consistent, performance levels and error patterns can be compared across and within children (Chiat et al., 2013). For these reasons, sentence repetition tasks may be uniquely useful as progress monitoring tools for tracking the morphosyntactic development of bilingual children.

The sentence repetition task is of interest for clinicians and educators working with Spanish-English speaking children because of its ease of administration, appropriate normative sample of bilingual children, and clinical importance in identifying children with language impairment (LI). Sentence repetition performance alone has been shown to classify young Spanish-English speaking DLLs with and without LI with over 80% specificity and sensitivity (Gutiérrez-Clellen & Simon-Cereijido, 2007; Gutiérrez-Clellen, Restrepo, & Simon-Cereijido, 2006). Further, children's performance on a sentence repetition task has been shown to be minimally associated with SES (Seeff-Gabriel, Chiat, & Roy, 2008), suggesting that the tool may differentiate low performance due to socioeconomic disadvantage from LI (Chiat et al., 2013). Given the importance of accurate identification of children with and without LI, the sentence repetition tasks included on the BESA may be key tools in improving the assessment and educational experiences of Spanish-English speaking children.

### **Limitations of the Literature**

Despite the importance of the BESA as a tool for identifying Spanish-English speaking DLLs with LI, the precise clinical utility of the scores obtained from the BESA is relatively unknown. The literature that has examined the validity and reliability of scores obtained from the BESA does not include rigorous evaluation of the psychometrics of each of the specific tasks included in the measure. Although factor analysis, prediction of performance on external

## DIMENSIONALITY OF SENTENCE REPETITION

measures, and item analysis are critical to assessing the validity, reliability, and precision of individual test items (Petrillo, Cano, McLeod, & Coon, 2015; Strauss & Smith, 2009), no studies to date have utilized these techniques to satisfactorily evaluate the tasks included on the BESA.

Preliminary investigations of the BESA's internal functioning have been limited. The measure's overall factor structure has been examined using principal components analysis (PCA) (Peña et al., 2014). PCA is a data reduction technique that relies on data-driven components extraction. Unlike confirmatory factor analysis (CFA), which is the preferred technique for specifying and comparing theoretically-plausible latent factor structures, PCA identifies components based only on the observed data, treating each data point as if they were measured without error (Costello & Osborne, 2005). Although useful to simplify interpretation of an individual dataset, PCA does not include adjustments for communalities among data loaded onto the same component or factor. Any component structure identified through PCA requires follow-up testing with a separate participant sample and theoretically-driven confirmatory factor analyses to allow for generalization to a larger population.

Another key limitation of the literature supporting the BESA is that factor analyses have only been conducted at the level of the full measure. No within-domain analyses have been conducted to evaluate the dimensionality of any of the individual subtests or tasks. This is problematic because there is potential for multidimensionality within each task that has gone untested. Treating a multidimensional measure as unidimensional can result in scaling errors, imprecise estimation of item discrimination, and inaccurate estimates of item fit on the general factor (Demars, 2012). All these issues are relevant in considering how to interpret the scaled scores obtained from each individual task and their contribution to the overall standardized scores derived for each language (AERA, APA, NCME, 2014 Standards 1.13, 2.3). This

## DIMENSIONALITY OF SENTENCE REPETITION

information is essential, given that score interpretation directly influences children's eligibility for educational services.

The sentence repetition tasks are particularly open to the influence of factors outside of children's morphosyntactic ability. Given the design of the tasks, which include words and word phrases from multiple word classes (i.e., parts of speech), children may perform similarly on clusters of items that have similar characteristics. There is evidence that DLLs learn words belonging to some word classes earlier than those from other classes. Specifically, children tend to produce concrete nouns before they begin producing verbs and more complex grammatical function words (Bornstein et al., 2004; Jackson-Maldonado, Thal, Marchman, Bates, & Gutiérrez-Clellen, 1993). Therefore, it is feasible that children's knowledge and general use of words from the different classes may influence their performance on the sentence repetition items. Children may demonstrate higher levels of accuracy on nouns and verbs than prepositions or adjectives, independent of their underlying morphosyntactic ability.

Because no studies to date have assessed the underlying factor structure of the individual tasks included on the BESA, the validity of the included items and the scoring system is called into question. One of the assumptions of classical test theory (CTT), which was used to assess the items included on the BESA (Peña et al., 2014), is that the scale being evaluated is unidimensional. For this assumption to be satisfied for the sentence repetition task, children's ability to respond correctly to any given item would need to be attributed only to his or her morphosyntactic knowledge in the language being assessed. As noted previously, this assumption has not been directly tested for the BESA. Consequently, the use of CTT is insufficient to facilitate understanding of how performance on this tool informs clinical practice.

The CTT-based item analysis also included Cronbach's alpha as a measure of internal

## DIMENSIONALITY OF SENTENCE REPETITION

consistency reliability, which relies upon rigid assumptions that have not been tested. For example, Cronbach's alpha requires that each item on a subtest or scale contribute equally to the total score (McNeish, 2017). This assumption requires every item on each task to load equally onto the general factor underlying that task. Cronbach's alpha also requires unidimensionality (McNeish, 2017), which similarly has not been tested at the task level of the BESA.

Consequently, Cronbach's alpha is not currently a trustworthy index of reliability for the BESA.

### **Investigating the Internal Structure of the BESA**

To determine the precise clinical utility of the BESA as a diagnostic measure for Spanish-English DLLs, deeper exploration of the tasks and subtests is needed. In the present paper, we focus specifically on the sentence repetition task of the BESA. Although this task is one of two included in the morphosyntax, it yields an independent norm-referenced scaled score that can be interpreted independently. Given that educators and practitioners are often under time constraints and therefore may elect to only administer portions of the BESA test battery, this paper focuses on the internal and predictive validity of administering the sentence repetition task in isolation. Results are intended to guide service providers in understanding how the sentence repetition task yields information about different components of children's underlying dual language skills.

Factor analysis at the item level can reveal the underlying structure of skills that influence children's performance on any given item. For the sentence repetition task of the BESA, we used a confirmatory factor analysis approach to compare three types of structural models for the task in Spanish and English. These included a unidimensional (one-factor) model, multidimensional correlated factors models, and bifactor models that include both a single underlying factor and specific factors.

The unidimensional model was included as a theoretically-plausible model for the task

## DIMENSIONALITY OF SENTENCE REPETITION

because the BESA creators intended it to be a measure of a single underlying construct: morphosyntax (Peña et al., 2014). The scoring schema currently used for the BESA suggests that variation in children's responses to the individual items can be attributed to a single underlying factor. We therefore compared the competing models against this one-factor structure.

The models including multiple correlated factors were constructed based on the targets' word classes to assess the influence of word class on children's performance. Word classes were highlighted as potentially-important factors because of the differing rates at which children begin producing words from separate word classes (Bornstein et al., 2004). Further, if the sentence repetition tasks yielded underlying factors based on word classes, information obtained from word class subscales may help service providers select treatment targets for children assessed using the BESA (Montrul, 2010).

Finally, the bifactor models incorporated both the single underlying factor and specific factors based on the word classes. Bifactor models can accommodate the overall similarities of all the test items while simultaneously modeling further similarities within clusters of items inside the overall structure (Reise, 2012; DeMars, 2013). Bifactor models include one general factor, which represents the primary underlying construct of interest, and specific factors, which represent the item clusters within the task. Given the hypothesized importance of word classes to DLLs' performance on the targets for the sentence repetition tasks (Bornstein et al., 2004; Jackson-Maldonado et al., 1993), the specific factors were specified by target word class.

### **Validity**

One advantage of bifactor modeling is that specific factors can be examined independently of the general factor because they are uncorrelated (Chen, West, & Sousa, 2006). For example, with sentence repetition task it is possible to test the relations between

## DIMENSIONALITY OF SENTENCE REPETITION

hypothesized word class factors and children's performance on an external measure above and beyond the general factor. Restated, both children's overall performance on all the items and their performance on specific clusters of items can be examined as predictors of an external measure. This feature of bifactor modeling allows for tests of discriminant and convergent validity, which in turn can help determine what latent abilities are being tapped by each factor included in the model.

For the present study, the final structural models were evaluated in relation to a measure of receptive English vocabulary. This external measure was selected because there is an established positive relation between English vocabulary and English sentence repetition performance among bilingual children (Klem et al., 2015; Komeili & Marshall, 2013). Additionally, using an external measure of vocabulary allowed for examination of the convergent validity of the word classes factors tested. Because children tend to begin producing nouns and verbs before words from more complex grammatical word classes (Bornstein et al., 2004; Jackson-Maldonado et al., 1993), a positive association between children's performance on the English noun and/or verb sentence repetition items and their English vocabulary scores may be anticipated (Polišenská, Chiat, & Roy, 2015).

The relation between children's English vocabulary scores and Spanish sentence repetition performance was examined from a discriminant validity perspective. The literature focused on the cross-linguistic associations between vocabulary and morphosyntax is limited (Simon-Cerejido & Gutiérrez-Clellen, 2009), but there is some evidence that bilingual children's grammatical skills in one language do not tend to be strongly associated with their vocabulary knowledge in the other language. Several papers examining the cross-linguistic relations between morphosyntax and vocabulary have found no relation between the constructs

## DIMENSIONALITY OF SENTENCE REPETITION

when they are examined across languages (Bedore et al., 2010; Conboy & Thal, 2006; Marchman, Martínez-Sussman, & Dale, 2004; Taliencich-Klinger, Bedore, & Peña, 2018). Consequently, no association between participants' English vocabulary scores and their overall performance on the Spanish sentence repetition task was expected. However, associations between children's performance on clusters of items on the Spanish sentence repetition task and their English vocabulary were possible. Children's English vocabulary knowledge could be related to their ability to repeat Spanish words belonging to certain word classes. Because of the limitations of the current literature, these relations were considered exploratory.

### **The Present Study**

Precision and reliability in standardized assessment are essential, particularly for DLLs who are misidentified as having impairment at disproportionate rates. The present study was conducted to examine the internal structure and predictive validity of the sentence repetition tasks of the BESA, a promising standardized measure designed to assist in the diagnosis of Spanish-English speaking children with speech and/or language impairment. We focused on the sentence repetition task because of its potential as an efficient and informative measure appropriate for DLLs. Employing an item-based approach, we examined the dimensionality of the sentence repetition tasks in Spanish and English. We then evaluated the relation between children's performance on the sentence repetition tasks and their English receptive vocabulary skills to inform the continued development of the tool and its interpretation.

## **Method**

### **Participants**

Participants included 291 children enrolled in elementary schools in Florida ( $n = 196$ ) and in Kansas ( $n = 95$ ). Teachers recruited the children based on parent report that they spoke

## DIMENSIONALITY OF SENTENCE REPETITION

Spanish at home. Approximately half of the participants were male ( $n = 149$ , 51.20%). Schools reported that 98% of the participants were eligible for free lunch and 2% were eligible for reduced lunch. Of families who completed phone demographics interviews with the research team ( $n = 220$ ), 67% of the primary caregivers reported completing less than a high school diploma, 23% had a high school diploma, and 7.7% attended some college. At the time of testing, 138 of the children were in kindergarten (47.42%) and 153 were in first grade (52.58%). Reported language use and linguistic background information for the children and families is provided in Table 1. All participants were receiving English-only classroom instruction. None of the children were identified with speech-language disorders or receiving special education services. All the children fell within the intended age range for the *Bilingual English-Spanish Assessment* (Peña et al., 2014), with an average age of 66 months ( $SD = 9.8$ ).

### **Procedures**

Tests of morphosyntax and English vocabulary were collected as part of a battery of baseline measures administered during a larger intervention study funded by the Institute of Education Sciences, U. S. Department of Education (Wood, Fitton, Petscher, Rodriguez, Sunderman, & Lim, 2018). The present study used extant data collected in elementary schools in Florida and Kansas prior to delivery of the intervention program. The study procedures were approved by the Florida State University human subjects institutional review board.

### **Measures**

Investigators and trained research assistants administered tests of sentence repetition, vocabulary, emergent literacy skills, and non-verbal intelligence at the beginning of the 2015-2016 school year. The sentence repetition and vocabulary measures included Spanish and English versions. The emergent literacy assessment assessed English only. Two trained research

## DIMENSIONALITY OF SENTENCE REPETITION

assistants scored all the tests and then an independent coder cross-checked all scores.

**Sentence repetition.** As described previously, the sentence repetition tasks are part of the morphosyntax subtests of the *Bilingual English-Spanish Assessment* (BESA; Peña et al., 2014). The BESA was developed using a normative sample of 756 Spanish-English bilingual children ages 4-6;11, with 17 dialects of Spanish and seven dialects of English represented in the sample. Sentence repetition is one of two tasks included within each of the BESA's morphosyntax subtests. The second is a cloze task designed to measure children's use of specific grammatical morphemes in obligatory contexts. The morphosyntax subtest is reported to have high internal consistency reliability in Spanish ( $\alpha = 0.96$ ) and in English ( $\alpha = 0.95$ ), and inter-rater reliability of 96% for Spanish and English (Peña et al., 2014). The sentence repetition task was administered in English and in Spanish to all participants in the study.

**English receptive vocabulary.** The *Peabody Picture Vocabulary Test, 4<sup>th</sup> Ed.* (PPVT-4; Dunn & Dunn, 2007) Form A was used to measure English receptive vocabulary skills. The test is a norm-referenced measure designed for English-speaking individuals (from 2:6 to 90 years old) in the United States. The assessment requires 10-15 minutes to administer and the child is asked to point to an auditorily labeled picture given a choice of four. The measure was normed on 3,540 individuals in the United States reflecting the national population distribution for sex, race/ethnicity, geographic region, socioeconomic status (SES), and clinical diagnosis. The coefficient alphas reported for Form A range from 0.95 to 0.97 for children aged 5 years to 6 years, 11 months. Split-half reliability ranges from 0.93 to 0.97 for the same age group.

**Descriptive measures.** Participants' scores on three additional measures are reported to provide more description of the sample. They were the *Test de Vocabulario en Imágenes Peabody* (TVIP; Dunn, Lugo, Padilla, & Dunn, 1986); the *Woodcock Reading Mastery Tests*,

## DIMENSIONALITY OF SENTENCE REPETITION

*Third Edition* (WRMT-III; Woodcock, 2011) letter identification (LI), phonological awareness (PA), and rapid automatic naming (RAN) subtests; and the *Primary Test of Nonverbal Intelligence* (PTONI; Ehrler & McGhee, 2008). The TVIP is a measure of Spanish receptive vocabulary constructed for monolingual Spanish-speaking children ages 2;6-17;11 years. The WRMT-III is intended to quantify language and literacy skills in English-speaking monolinguals and was normed for individuals 4-79 years old. The PTONI was designed to measure nonverbal intelligence and uses pictures to assess reasoning in children without requiring a verbal response. All standardized scores obtained from these descriptive should be interpreted with caution, given that these tests rely on normative samples that do not directly match the characteristics of the present Spanish-English speaking sample.

### **Analyses of the Sentence Repetition Task**

First, descriptive statistics for the individual items included in the BESA sentence repetition tasks were obtained. Average percent correct, standard deviations, and item-total correlations were computed for each item. This information was used to provide a general overview of item functioning within the Spanish and English versions of the task.

Next, to evaluate the dimensionality of the BESA sentence repetition task, item-level confirmatory factor analyses were conducted in Mplus 7.31 (Muthén & Muthén, 1998-2012). Three types of structural models were specified for the task in each language using weighted least squares means and variances (WLSMV) estimation. These included: (a) a unidimensional model with a single general factor, (b) multidimensional correlated factor models including specific factors specified based on the word classes represented by the items, and (c) bifactor models including both a single general factor and multiple specific factors. The Spanish versions of these models are shown in Figure 1 and the English versions are provided in Figure 2. The

## DIMENSIONALITY OF SENTENCE REPETITION

English items were divided into six word classes (pronouns, noun phrases, prepositions, subordinating conjunctions, verb phrases, and copula/auxiliary), but the Spanish items were only divided into five word classes. None of the Spanish targets were copula or auxiliary.

[insert Figure 1]

[insert Figure 2]

### **Model Comparisons**

Each model represents a plausible underlying structure for the BESA sentence repetition task and has distinct implications for interpreting children's scores on the task. As such, the models were compared systematically through consideration of both statistical and theoretical fit. The unidimensional models were specified and examined for each language first. These models served as the basis for comparison for all other models because the unidimensional models represent the current scoring scheme employed for the BESA sentence repetition task. If none of the comparison models had exhibited better statistical fit than the unidimensional models, then no further investigation would be required.

The word classes correlated factors models were examined next. The most complex models were specified based on all the hypothesized word classes represented by the items (model B in Figure 1 and model E in Figure 2). Simpler models were then specified to compare against the most complex model and against the unidimensional model. These more parsimonious models were constructed by combining word classes based on their function within the sentences. For example, in English the most complex correlated factors model included a separate factor for prepositions or prepositional phrases (e.g., "for school") and for noun phrases (e.g., "the doctor"). Simpler model structures combined these word classes into a single factor because both generally included function words followed by a lexical word. Similarly, in

## DIMENSIONALITY OF SENTENCE REPETITION

Spanish, the most complex correlated factors model included separate factors for prepositions (e.g., “con”) and subordinating conjunctions (e.g., “cuando”). These word classes were combined for the simpler models because both serve as connectors for dependent clauses. Model fit and item loadings were also taken into consideration during model specification, as later described. This iterative process facilitated the identification of the correlated factors model that provided the simplest, most accurate representation of item functioning for the sentence repetition task.

The bifactor model structures were specified similarly. First, the most complex models were specified based on all the hypothesized word classes represented within the items (model C in Figure 1 and model F in Figure 2). Simpler models were then specified to compare against the unidimensional model and against the most complex bifactor model. These simpler models were initially constructed by combining any specific factors that were theoretically similar, as with the correlated factors models. After combining the specific factors that could theoretically load onto a single specific factor, additional specific factors were removed based on the item loadings onto the specific factors. Specific factors with low factor loadings were eliminated so that the items only loaded onto the general factor. This process continued until the model with the best balance of model fit and parsimony was identified.

Each model was examined individually for evidence of misspecification prior to comparison against other models. Overall fit indices including the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis index (TLI) were inspected. As recommended by Kline (2015), an RMSEA below .10 and a CFI and TLI above .90 were considered indicators of good model fit. The item loadings, factor covariances, and residuals were also examined. As noted by Kline (2015), standardized loadings or covariances above 1.0 and residuals below zero are suggestive of model misfit. Any model

## DIMENSIONALITY OF SENTENCE REPETITION

exhibiting evidence of misfit was removed from consideration.

Models that exhibited good model fit indicators were then compared to identify the structure that best represented the functioning of the sentence repetition items in each language. Nested models were compared using chi-square difference testing in Mplus (Muthén & Muthén, 1998-2012). Because the models were initially estimated using WLSMV, comparisons were conducted using the DIFFTEST option (Muthén & Muthén, 1998-2012). A significant result from these comparisons indicated that the more parsimonious (i.e., simpler) model was a significantly worse fit to the data compared to the more complex model. Additionally, Akaike information criterion (AIC) and Sample-size adjusted Bayesian information criterion (BIC) values were obtained through re-estimating the models with good fit using maximum likelihood (ML). Lower AIC and BIC values were generally preferred (Kline, 2015). However, it should be noted that fit indices are sample-dependent and often overlap in comparisons of correlated factors and bifactor models (Morgan, Hodge, Wells, & Watkins, 2015). As such, the substantive interpretation of the models was also taken into consideration in identifying the best structure.

Upon identification of the Spanish and English factor structures with the best balance of fit, parsimony, and consistency with theory, reliability coefficients were computed. Coefficient omega was used to accommodate the violation of tau equivalence (McNeish, 2017). Coefficients omega hierarchical, indices of explained common variance, and the average relative bias in parameter estimates were also computed for the bifactor models to address issues related to bifactor model overfitting (Bonifay, Lane, & Reise, 2017; Rodriguez, Reise, & Haviland, 2016a, 2016b).

Finally, two structural regression models were specified to examine the relation between children's performance on the sentence repetition items and their English vocabulary scores. For

## DIMENSIONALITY OF SENTENCE REPETITION

the first structural regression, children's vocabulary scores in English, as measured by the PPVT-4, were predicted by the factors identified in the best-fitting Spanish sentence repetition model. In the second structural regression, PPVT-4 scores were predicted by the factors in the best-fitting English sentence repetition model.

### Results

#### Descriptive Results Performance

Table 2 provides descriptive information about children's performance on all the measures included in the test battery. Preliminary analysis of the sentence repetition subtest items revealed that the average percent correct was .56 ( $SD = .17$ ) for the Spanish items. The lowest value obtained for percent correct was .20 and the greatest value was .84. The items categorized as nouns exhibited the highest percent accuracy (.67,  $SD = .14$ ) compared to the other word classes. Item-total correlations ranged from .39 to .68, and Cronbach's alpha within the Spanish subtest was .95. For English, the children generally responded correctly more often, with an average percent correct of .72 ( $SD = .11$ ) for the English items. The lowest percent correct value obtained was .52 and the greatest value was .94. The items categorized as noun phrases exhibited the highest average percent correct at .80 ( $SD = .10$ ). Item-total correlations for the English measure ranged from .37 to .65. Cronbach's alpha for the English subtest was .93.

#### Dimensionality in Spanish

Model fit statistics for the measurement models in Spanish are shown in Table 3. The best-fitting Spanish multidimensional correlated factors model included 3 factors. Both the 5-factor and 4-factor correlated factors models revealed evidence of misspecification; correlations among the latent variables were observed to be above 1.0. Collapsing the prepositional phrases, subordinating conjunctions, and noun phrases to load onto a single factor did not result in a

## DIMENSIONALITY OF SENTENCE REPETITION

significantly worse fit to the data compared to the more complex models:  $\Delta\chi^2 = 7.51$ ,  $\Delta df = 3$ ,  $p = .0574$ . However, simplifying the factor structure further through loading the pronoun items onto the noun phrase factor resulted in significantly worse model fit:  $\Delta\chi^2 = 19.55$ ,  $\Delta df = 2$ ,  $p = .0001$ . Consequently, the 3-factor model was identified as the best multidimensional correlated factors model in Spanish.

[insert Table 3]

The first Spanish bifactor model evaluated reflected the structure of the 5-factor correlated factors model, with five specific factors and one general factor (model C in Figure 1). However, this model exhibited evidence of misspecification related to the prepositional phrase and subordinating conjunction specific factors. Removal of these two specific factors from the model did not result in a significantly worse fit to the data:  $\Delta\chi^2 = 12.21$ ,  $\Delta df = 8$ ,  $p = .1422$ . Consequently, the prepositional phrase items and subordinating conjunction items were constrained to load onto only the general factor (see model G in Figure 3). Theoretically, this result indicated that the general factor accounted for most of the similarities in the children's performance between the prepositional phrase items and the subordinating conjunction items. Neither the specific factors for prepositional phrases nor for subordinating conjunctions accounted for additional significant variance in the children's performance on the Spanish sentence repetition task above and beyond the general factor.

[insert Figure 3]

The bifactor model including three specific factors (pronouns, noun phrases, and verbs) was identified as model with the best balance of model fit, parsimony, and theory. The model exhibited significantly better fit compared to the unidimensional model:  $\Delta\chi^2 = 105.53$ ,  $\Delta df = 28$ ,  $p < .0001$ . This bifactor model structure yielded the lowest RMSEA and AIC value, as well as

## DIMENSIONALITY OF SENTENCE REPETITION

the highest CFI and TLI. Further, the bifactor model with only two specific factors (pronouns and verbs) was a significantly worse fit to the data compared to the model with three specific factors:  $\Delta\chi^2 = 21.84$ ,  $\Delta df = 9$ ,  $p = .0094$ . Finally, this model fit the theoretical expectation for the sentence repetition task. The bifactor model suggests that children's performance on each sentence repetition item informs a single underlying ability. However, the model also allows for associations among the residuals for each item. The residuals cluster by three of the word classes included in the task. Given that the task was designed to assess a single underlying ability, this model structure best fits that original intent of the test authors.

For this bifactor model with three specific factors, a coefficient omega hierarchical of 0.998 was obtained for the total Spanish sentence repetition score. Coefficients omega hierarchical subscale of 0.542, 0.095, and 0.392 were obtained for the pronoun, noun phrase, and verb item total scores, respectively. The explained common variance obtained based on this model was .92, indicating that the general factor accounted for 92% of the common variance among items. The remaining 8% of shared item variance was attributable to the word class groupings. The parameter estimates for the item loadings on the general factor within the bifactor structure were then compared to those obtained from the unidimensional model (Rodriguez et al., 2016b). The average item bias was revealed to be 0.72%, indicating little difference in parameter estimates obtained from the two models (Muthén, Kaplan, & Hollis, 1987).

All 37 of the Spanish sentence repetition items loaded positively onto the general factor. These loadings ranged from .50 to .87. For the pronouns specific factor, 60.00% ( $n = 3$ ) of the items loaded positively onto the factor and the rest ( $n = 2$ ) loaded negatively. Similarly, 71.43% ( $n = 10$ ) of the item loadings onto the verb phrases specific factor were positive. However, only 44.44% ( $n = 4$ ) of the item loadings onto the noun phrases specific factor were positive. As such,

## DIMENSIONALITY OF SENTENCE REPETITION

when interpreting findings relative to each of these factors, the pronouns and verb phrases factors can be interpreted simply. Higher accuracy on the pronouns and verb phrases items is associated with higher factor scores for pronouns and verb phrases. For noun phrases, however, the opposite interpretation is needed. Higher accuracy on the noun phrase items is associated with lower factor scores for noun phrase items.

### **Dimensionality in English**

Model fit statistics for the measurement models in English are provided in Table 3. The best-fitting English multidimensional correlated factors model included 2 factors. The 6-factor, 5-factor and 4-factor correlated factors models exhibited evidence of misspecification, with correlations among the latent variables observed to be above 1.0. Noun phrases and prepositional phrases were subsequently loaded onto the same factor; pronouns, copula/auxiliary, and subordinating conjunctions were loaded onto a second factor; and verb phrases were loaded onto a third factor. Although the fit of this model to the data was no worse than the 4-factor correlated factors model ( $\Delta\chi^2 = 3.76$ ,  $\Delta df = 3$ ,  $p = .2880$ ), it did not provide better fit to the data than a more parsimonious 2-factor model  $\Delta\chi^2 = 2.81$ ,  $\Delta df = 2$ ,  $p = .2453$ . For the 2-factor model, the second and third factors were collapsed (pronouns, copula/auxiliary, subordinating conjunctions, and verb phrases). This model was a significantly better fit to the data than the unidimensional model:  $\Delta\chi^2 = 17.10$ ,  $\Delta df = 1$ ,  $p < .0001$ . Consequently, the 2-factor model was identified as the best multidimensional correlated factors model in English.

The first English bifactor model evaluated included one general factor and six specific factors, mirroring the structure of the 6-factor correlated factors model (see F in Figure 2). This model revealed evidence of misspecification through negative residual variances. Similar to observations made while specifying the correlated factors models, collapsing prepositional

## DIMENSIONALITY OF SENTENCE REPETITION

phrases and noun phrases onto the same factor and copula/auxiliary and verb phrases onto the same factor resolved all indications of model misspecification. Subsequent model fit comparisons were made against this bifactor model with four specific factors. More parsimonious models, specified by constraining pronouns, subordinating conjunctions, and the combined copula/auxiliary and verb phrases factor to load only onto the general factor, did not result in a significantly worse fit to the data:  $\Delta\chi^2 = 24.29$ ,  $\Delta df = 19$ ,  $p = .1852$ . However, the model including one general factor and one specific factor, constructed from the preposition phrase and noun phrase items, was a significantly better fit to the data compared to the unidimensional model:  $\Delta\chi^2 = 77.67$ ,  $\Delta df = 12$ ,  $p < .0001$ .

The bifactor model including one specific factor was identified as the model with the best balance of model fit, parsimony, and theory for English sentence repetition. The model provided the best statistical fit to the data, evidenced by its AIC and sample-size adjusted BIC values. It was no worse a fit to the data than any of the more complex models examined and was a significantly better fit than the unidimensional model. Finally, as with the Spanish bifactor model, this English bifactor model maps onto the theoretical rationale behind the creation of the sentence repetition task. Each item provides information about a single underlying ability. However, some of the residual variance in children's performance on the items can be explained by the word class of the item. Children tended to perform similarly on items that were noun phrases or prepositional phrases, above and beyond their general performance on the sentence repetition task in English.

For this bifactor model with one specific factor, a coefficient omega hierarchical of 0.997 was obtained for the total English sentence repetition score. A coefficient omega hierarchical subscale of 0.166 was obtained for the preposition and noun phrase item total score. This model

## DIMENSIONALITY OF SENTENCE REPETITION

yielded an explained common variance of .93, indicating that the general factor accounted for 93% of the common variance among items. The remaining 7% of shared item variance was attributable to the specific factor. The parameter estimates for the item loadings on the general factor within the bifactor structure were also compared to those obtained from the unidimensional model (Rodriguez et al., 2016b), revealing an average item bias of 0.67%. This value indicates little difference in parameter estimates obtained from the two models (Muthén et al., 1987).

Item loadings onto the general factor were all positive ( $n = 33$ ), ranging from .52 to .84. For the specific factor, 75% ( $n = 9$ ) of the items loaded negatively onto the factor and the rest ( $n = 3$ ) loaded positively. These findings indicate that children with higher overall performance across all the sentence repetition items were more likely to respond incorrectly to the noun and prepositional phrase items. Children who tended to have more incorrect responses overall were more likely to respond correctly to the noun and prepositional phrase items specifically. Rephrased, the noun and prepositional phrase items were easier for children with lower overall sentence repetition ability in English.

### **Predictive Validity**

The structural equation models conducted to predict children's concurrent receptive vocabulary scores in English also were a good fit to the data. For Spanish,  $\chi^2(635) = 847.74, p < .001$ ; RMSEA = .034 (90% CI = .028 - .040); CFI = 0.98, TLI = .98. This model, which included children's Spanish sentence repetition performance predicting PPVT-4 scores (see Figure 4), accounted for 25.8% of the variance in children's PPVT-4 performance. Although the general factor in Spanish did not significantly predict English vocabulary (est < 0.01,  $p = .943$ ), all three specific factors significantly predicted PPVT-4 scores. Children with higher performance on the

## DIMENSIONALITY OF SENTENCE REPETITION

pronoun, noun phrase, and verb phrase items in Spanish tended to have lower PPVT-4 scores. Consideration of the factor loadings is needed to arrive at this interpretation; the pronoun and verb phrase items generally loaded positively onto their specific factors, but the noun phrase items were negatively loaded overall.

[insert Figure 4]

The model constructed to predict PPVT-4 performance with English sentence repetition (see Figure 5) was a good fit to the data:  $\chi^2(514) = 676.35, p < .001$ ; RMSEA = .033 (90% CI = .026 - .040); CFI = 0.97, TLI = .97. The model revealed that the general English factor was a significant, positive predictor. The English nouns and prepositional phrases specific factor also uniquely contributed to predicting English vocabulary scores. Together the two factors accounted for 41.9% of the variance in children's PPVT-4 scores. Children with higher overall performance on the English sentence repetition task tended to have higher PPVT-4 scores. Above and beyond their overall performance, children who responded correctly to the noun and prepositional phrases items exhibited even higher PPVT-4 scores. Similar to the Spanish model, consideration of the factor loadings is needed to accept this interpretation. All the English sentence repetition items loaded positively on the general factor, but the noun and prepositional phrase items generally loaded negatively onto the specific factor.

[insert Figure 5]

## Discussion

### Dimensionality

The present study was conducted to assess the dimensionality of the Spanish and English versions of the BESA sentence repetition task, which was designed as a measure of the morphosyntactic skills of young Spanish-English speaking children. Item-level factor analyses

## DIMENSIONALITY OF SENTENCE REPETITION

revealed that the Spanish and English versions of the task are most precisely described as multidimensional, with children's performance on each item being influenced by multiple underlying constructs. Bifactor models yielding the best global fit statistics indicate that not only do children tend to exhibit similar performance across all the items included, but they also perform similarly on specific subsets of items above and beyond their overall performance. However, further analyses revealed that, although the bifactor models provided the best overall fit to the data, the sentence repetition tasks can be treated as essentially unidimensional. For the purposes of scoring, the specific factors identified within the bifactor model structures did not account for a significant amount of variance in children's performance on the task. Further, there was little difference in the parameter estimates obtained for the sentence repetition items following a unidimensional framework compared to the bifactor frameworks. These findings provide support for the current approach to scoring this task.

The single general factor found to fit the Spanish and English versions of the task is most likely representative of children's underlying morphosyntactic knowledge in each language, given previous findings that children's general ability to repeat sentences is linked to grammaticality and morphological awareness (Komeili & Marshall, 2013; Polišenská et al., 2015). In both languages, all the items loaded onto this factor positively with excellent reliability, suggesting that a child who performs well on all the items is likely to have strong morphosyntactic skills. Conversely, a child who performs poorly on all items is likely to have weak morphosyntactic skills.

The dimensionality findings have practical implications for the use of this sentence repetition task as a measure of morphosyntax. Importantly, these findings support the use of a unidimensional scoring system. It is worth noting at this point that there is a close relationship

## DIMENSIONALITY OF SENTENCE REPETITION

between the estimation of categorical item-level confirmatory factor analyses (CFA) and item response theory (IRT) models. CFA models are designed to model covariance between test items, while IRT models directly connect and model individual test takers' responses. When the underlying scale is found to be unidimensional (or essentially unidimensional), results from a categorical CFA and 2-parameter IRT model provide the same information. IRT has been described as a special case of CFA, where unidimensionality and local independence are assumed (de Ayala, 2013). This is relevant in the present work because IRT models guide the specific selection and refinement of approaches to equating, scaling, and adaptive testing. The present findings support the application of unidimensional IRT approaches with the BESA sentence repetition task. These approaches can extend and further specify the use of this task in diverse and potentially broader contexts.

The specific factors identified within the bifactor frameworks are worth discussing, however, due to their relations to children's English vocabulary scores. Although essentially ignorable from a scoring standpoint in clinical practice, there may be value in accounting for these factors when examining children's BESA sentence repetition performance on a large scale in research or when considering treatment targets for individual children. Specifically, children's ability to repeat pronoun, noun phrase, and verb phrase items in Spanish appears to be associated with their underlying knowledge of these respective word classes. Children with strong pronoun skills in Spanish tend to repeat pronoun items within the Spanish sentence repetition task with higher accuracy than children with weaker pronoun skills, after accounting for overall morphosyntactic skills. The same was observed for the noun phrase and verb phrase items in Spanish. In English, children's ability to repeat nouns and prepositional phrases was linked to their knowledge of these word classes. Children with strong noun and prepositional phrase skills

## DIMENSIONALITY OF SENTENCE REPETITION

tended to repeat items belonging to those word classes with greater accuracy than children with weaker skills, again after accounting for overall performance.

Interestingly, both language versions of the sentence repetition task revealed that children performed similarly on the noun phrase targets above and beyond their overall ability to repeat the sentences. This finding may be attributable in part to the developmental nature of nouns. Prior work suggests that nouns are learned at an earlier age than other word classes in both Spanish (Jackson-Maldonado et al., 1993) and English (Bornstein et al., 2004). Verbs are also a relatively early-acquired word class in English, evidenced by their presence in young children's vocabularies (Tomasello & Merriman, 1995). As such, it is possible that the specific factors identified for each language represent children's vocabulary knowledge within the word classes.

### **Predictive Validity**

The findings from the predictive models provide intriguing insight into children's performance on the sentence repetition task. English vocabulary has been shown to have a positive association with English morphosyntactic skills (Marchman et al., 2004) and no significant association with Spanish morphosyntactic skills (Simon-Cereijido & Gutiérrez-Clellen, 2009) among bilingual children. Results revealed that, within the bifactor framework, both the general factor and the specific factor representing noun and prepositional phrases in English were unique predictors of English receptive vocabulary. In Spanish, the general factor was not associated with English receptive vocabulary, but all three specific factors were unique predictors. Overall, the English morphosyntax model predicted 41.9% of the variation in children's scores on the vocabulary measure, while the Spanish morphosyntax model predicted only 25.8% of the variance in English vocabulary.

For English, the interpretation of these predictive results is straightforward. The general

## DIMENSIONALITY OF SENTENCE REPETITION

factor, which can be interpreted as overall English morphosyntactic knowledge, accounted for a significant, large portion of the variance in English vocabulary. This is consistent with prior work suggesting a positive relation between English vocabulary and English morphosyntax among bilingual children (e.g., Conboy & Thal, 2006; Simon-Cereijido & Gutiérrez-Clellen, 2009). This finding also aligns with previous research suggesting morphological awareness and vocabulary tasks overlap as they are both measures of underlying language ability (Spencer et al., 2015), with morphosyntactic skills being an integral part of vocabulary knowledge. The specific factor for noun phrases, which can be interpreted as children's vocabulary knowledge specific to noun and prepositional phrases, also contributed uniquely to predicting English vocabulary performance. This finding is similarly unsurprising given the nature of the receptive vocabulary measure, the PPVT-4 (Dunn & Dunn, 2007). Over 65% of target words included on the PPVT-4 can be categorized as nouns. This percentage is even higher for items included in the earliest sets of the test. Consequently, it is reasonable for children's noun phrase vocabulary to explain their performance on a measure that includes primarily nouns.

The observed relation between the Spanish specific factors and English vocabulary is less straightforward. The general factor, representing children's Spanish morphosyntactic knowledge, did not significantly contribute to predicting their English vocabulary scores. This finding is consistent with prior work exploring the cross-linguistic relations between morphosyntax and vocabulary (Simon-Cereijido & Gutiérrez-Clellen, 2009). However, the pronouns, nouns, and verb phrase specific factors all uniquely contributed to predicting children's PPVT-4 scores. As noted previously, consideration of the item loadings onto each of the factors is essential to interpreting the associations between these specific factors and children's vocabulary scores. The general and specific factors specified in the English sentence repetition model had primarily

## DIMENSIONALITY OF SENTENCE REPETITION

positive item loadings. In Spanish, however, the noun phrases specific factor exhibited primarily negative item loadings, contrasting the mostly-positive loadings observed for the Spanish general factor and the pronouns and verb phrase specific factors. This consequently affects the interpretation of the parameter estimate for Spanish noun phrases predicting English vocabulary. The estimate is positive in the overall predictive model, but the negative loadings indicate that the estimate should be interpreted inversely. Children with lower performance on the Spanish noun phrase items tended to have higher English vocabulary scores. This interpretation is consistent with the remaining findings from the predictive model. Children with higher scores on the Spanish pronoun, noun phrase, and verb phrase items (above and beyond their overall performance on the measure) tended to have lower English vocabulary scores. If these specific factors are in fact indicators of children's vocabulary knowledge, then this negative association is consistent with evidence that young Spanish-English speaking children with strong skills in one language tend to have weaker skills in the opposite language (Hoff & Core, 2013; Kohnert, 2010; Scheffner Hammer et al., 2012). Notably, this negative association was obtained above and beyond children's general morphosyntactic skills in Spanish. As such, only children with very strong (or very weak) Spanish skills tended to exhibit the opposite pattern in English.

### **Limitations**

Given the relatively small sample size, caution is recommended in extending the findings beyond the current sample. These results offer a framework for future evaluation of the BESA and other assessments designed for DLLs, but it is possible that the findings are dependent on the characteristics observed in the sample. For example, the participants for this study were all from relatively low SES backgrounds. All the children were identified as eligible for free or reduced price lunch. It is possible that the resulting factor structure and relations obtained would not

## DIMENSIONALITY OF SENTENCE REPETITION

generalize to a higher-SES sample. Additionally, results may not generalize to Spanish-English speaking DLLs outside of the United States nor to those who are outside the BESA's normative age range of 4-6;11 years old. The dimensionality of the task may differ for children being educated in different language learning environments, and response patterns may vary as a function of age. These factors were outside the scope of the present work. Replication is needed with larger, more diverse samples that more accurately represent the young Spanish-English DLL population in the U.S. to generalize the findings beyond the current sample of children.

Further, caution is recommended in generalizing findings from this work to other sentence repetition tasks. Sentence repetition tasks like those used on the BESA are often included within language test batteries intended to distinguish children with language impairment. Differences in test development procedures, administration, and scoring of these tasks, however, may influence the underlying factor structure of these tasks. For example, sentence repetition tasks that yield scores corresponding to each sentence (e.g., child is given a score ranging from 0-3 for each sentence based on the number of errors or omissions instead of given a point for each correctly repeated target in a sentence) may provide information about underlying language constructs that are different from those identified in the present work. Further research examining the factor structure of multiple sentence repetition tasks of differing formats may provide insight into how scoring and administration procedures influence children's responses to these types of tasks.

Finally, due to the complex nature of factor structure analysis and restrictions related to sample size, quantity of parameters, and model identification, this study did not include sentence-level covariates (e.g., length, complexity) nor all possible item-level covariates (e.g., relative position in sentence). Because the present findings suggest that specific item

## DIMENSIONALITY OF SENTENCE REPETITION

characteristics influence children's performance on this sentence task, it is recommended that future work explore additional sentence-level and item-level factors that may explain performance above and beyond children's underlying morphosyntactic skills. Accounting for the influence of these characteristics may help to reduce residual error of measurement. Further research is needed on the identified specific word classes' factors to examine stability, predictive relationships to other language and literacy skills, and utility for progress monitoring children's broader morphosyntactic skills.

### **Future Directions**

The results from the present paper add to the evidence based regarding the precise clinical utility of the BESA, its tasks, and its subtests. Similar item-level analyses are needed to vet each of the portions of the tool. Additionally, after determining the individual tasks' dimensionality, internal consistency, and predictive functioning, it would be valuable to examine the entire battery at the item level. This can lead to more efficient interpretation of test results and provide insight into future development of subsets of items that may be used in screening.

### **Conclusions**

The results from this paper provide empirical support for the current scoring system of the BESA sentence repetition task in both Spanish and English. Findings also add to the evidence base for the construct and internal validity of the task as a measure of morphosyntax in Spanish and English. For clinicians, this provides support for the use of the sentence repetition task scaled score in clinical reporting as evidence for children's morphosyntactic skills. For researchers, these results support the treatment of the BESA sentence repetition tasks as unidimensional, opening opportunities for more further examination of item functioning from an IRT approach. However, results also suggest that children's performance on specific items

## DIMENSIONALITY OF SENTENCE REPETITION

within the BESA sentence repetition task can provide further insight into additional skills. There may be value for clinicians in examining children's item-level errors. Consistently-low performance on a specific word class, such as noun phrases, may indicate weaknesses in a child's knowledge of that particular word class and consequently guide treatment target identification. Researchers may consider the potential clustering of residual variance in follow-up studies, where meaningful information may be obtained from examining children's item-level performance. Overall, findings support the construct validity of the task as a measure of morphosyntax in Spanish and in English, with the possibility that additional meaningful information can be gleaned from item analysis of children's responses.

### **Acknowledgements**

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130460 to Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

**References**

- Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record*, 108(11), 2282-2303. <http://dx.doi.org/10.1111/j.1467-9620.2006.00782.x>
- Artiles, A. J., Rueda, R., Salazar, I., & Higareda, J. (2002). Of rocks and soft places: English-language learner representation in special education in California urban school districts. In D. J. Losen & G. Orfield (Eds.), *Racial inequality in special education* (pp. 117–136). Cambridge, MA: Harvard Education Press.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bedore, L. M., Peña, E. D., Gillam, R. B., & Ho, T. (2010). Language sample measures and language ability in Spanish English bilingual kindergarteners. *Journal of Communication Disorders*, 43(6), 498-510. <https://doi.org/10.1016/j.jcomdis.2010.05.002>
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184-186. <https://doi.org/10.1177/2167702616657069>
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., . . . Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75(4), 1115-1139. <http://dx.doi.org/10.1111/j.1467-8624.2004.00729.x>

## DIMENSIONALITY OF SENTENCE REPETITION

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225.

[https://doi.org/10.1207/s15327906mbr4102\\_5](https://doi.org/10.1207/s15327906mbr4102_5)

Chiat, S., Armon-Lotem, S., Marinis, T., Polišenská, K., Roy, P., & Seeff-Gabriel, B.

(2013). Assessment of language abilities in sequential bilingual children: The potential of sentence imitation tasks. In V. C. M. Gathercole (Ed.), *Issues in the Assessment of Bilinguals* (pp. 56-89). Bristol: Multilingual Matters.

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

<https://doi.org/10.1111/j.1467-8624.2006.00899.x>

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practice Assessment, Research & Evaluation, 10*(7), 1-9. Retrieved from <http://pareonline.net/pdf/v10n7.pdf>

de Ayala, R. J. (2013). Factor analyses with categorical indicators. In Y. Petscher, C.

Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analyses in the educational and social sciences* (pp. 208–242). New York, NY: Routledge.

Demars, C. E. (2012). Confirming testlet effects. *Applied Psychological Measurement, 36*(2),

104–121. <https://doi.org/10.1177/0146621612437403>

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of*

*Testing, 13*, 354-378. <https://doi.org/10.1080/15305058.2013.799067>

Dunn, L. M., & Dunn, D. M. (2007). *The Peabody Picture Vocabulary Test* (4th ed.). Circle Pines, MN: American Guidance Service.

Dunn, L., Lugo, S., Padilla, R., & Dunn, L. (1986). *Test de Vocabulario en Imagenes Peabody*.

## DIMENSIONALITY OF SENTENCE REPETITION

Circle Pines, NM: American Guidance Service.

Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary Test of Nonverbal Intelligence*. Pro-Ed.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, 78(2), 350-365. <http://dx.doi.org/10.1037/0022-3514.78.2.350>

Gathercole, V. C. M., Thomas, E. M., Roberts, E. J., Hughes, C. O., & Hughes, E. K. (2013).

Why assessment needs to take exposure into account: Vocabulary and grammatical abilities in bilingual children. In V. C. M. Gathercole (Ed.), *Issues in the Assessment of Bilinguals* (pp. 20-55). Bristol: Multilingual Matters.

Gutiérrez-Clellen, V. F., Restrepo, M.A., & Simón-Cerejido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research*, 49, 1209-1223.

[https://dx.doi.org/10.1044/1092-4388\(2006/087\)](https://dx.doi.org/10.1044/1092-4388(2006/087))

Gutiérrez-Clellen, V. F., & Simon-Cerejido, G. (2007). The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech, Language, and Hearing Research*, 50(4), 968-981. [https://dx.doi.org/10.1044/1092-4388\(2007/068\)](https://dx.doi.org/10.1044/1092-4388(2007/068))

Hamayan, E., Marler, B., Sanchez-Lopez, C., & Damico, J. (2007). Reasons for the misidentification of special needs among ELLs. In *Special education considerations for English language learners: Delivering a continuum of services* (pp. 2-7). Philadelphia, PA: Caslon.

Hemphill, F. C., & Vanneman, A. (2011). *Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of*

## DIMENSIONALITY OF SENTENCE REPETITION

*Educational Progress* (NCES 2011-459). NCES, Institute of Education Sciences, U.S. Department of Education: Washington, DC.

Hoff, E., & Core, C. (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, 34(03), 215-226. <https://doi.org/10.1055/s-0033-1353448>

Jackson-Madonado, D., Thal, D., Marchman, V., Bates, E., & Gutiérrez -Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, 20, 523-549. <https://doi.org/10.1017/S0305000900008461>

Kapantzoglou, M., Thompson, M. S., Gray, S., & Restrepo, M. A. (2016). Assessing measurement invariance for Spanish sentence repetition and morphology elicitation tasks. *Journal of Speech, Language, and Hearing Research*, 59, 254-266. [https://doi.org/10.1044/2015\\_jslhr-l-14-0319](https://doi.org/10.1044/2015_jslhr-l-14-0319)

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., . . . Kristapovich, P. (2014). *The Condition of Education 2014* (NCES 2014-083). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., . . . Velez, E. D. V. (2015). *The Condition of Education 2015* (NCES 2015-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental science*, 18(1), 146-154.

## DIMENSIONALITY OF SENTENCE REPETITION

<https://doi.org/10.1111/desc.12202>

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.) New York: Guilford Press.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence, and implications for clinical actions. *Journal of Communication Disorders*, 43(6), 456-473.

<https://doi.org/10.1016/j.jcomdis.2010.02.002>

Komeili, M., & Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clinical linguistics & phonetics*, 27(2), 152-162.

<https://doi.org/10.3109/02699206.2012.751625>

MacSwan, J., & Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teacher's*

*College Record*, 108, 2304–2328. <http://dx.doi.org/10.1111/j.1467-9620.2006.00783.x>

Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental*

*Science*, 7(2), 212-224. <https://doi.org/10.1111/j.1467-7687.2004.00340.x>

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*.

Advance online publication. Retrieved from <https://doi.org/10.1037/met0000144>

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30, 3-23. <https://doi.org/10.1017/s0267190510000103>

Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism:*

*Language and Cognition*, 17(1), 118-138. <https://doi.org/10.1017/s1366728913000114>

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in

## DIMENSIONALITY OF SENTENCE REPETITION

- favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3, 2-20. <https://doi.org/10.3390/jintelligence3010002>
- Murphey, D. (2014). *The academic achievement of English language learners: Data of the U.S. and each of the states* (Research Brief). Bethesda, MD: Child Trends.
- Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52, 431-462.  
<http://dx.doi.org/10.1007/BF02294365>
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Passel, J. S., Cohn, C., & Lopez, M. H. (2011). *Hispanics account for more than half of nation's growth in past decade. Census 2010: 50 million Latinos*. Washington, DC: Pew Hispanic Center.
- Pawlowska, M. (2014). Evaluation of three proposed markers for language impairment in English: A meta-analysis of diagnostic accuracy studies. *Journal of Speech, Language, and Hearing Research*, 57, 2261-2273. [https://doi.org/10.1044/2014\\_jslhr-1-13-0189](https://doi.org/10.1044/2014_jslhr-1-13-0189)
- Peña, E. D., Bedore, L. M., & Kester, E. S. (2015). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders*, 51, 192-202.  
<https://doi.org/10.1111/1460-6984.12199>
- Peña, E. D., Gutiérrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *Bilingual English-Spanish Assessment Manual*. San Rafael, CA: AR-Clinical Publications.
- Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item

## DIMENSIONALITY OF SENTENCE REPETITION

- response theory, and Rasch measurement theory to evaluate patient-report outcome measures: A comparison of worked examples. *Value in Health*, 18, 25-34. doi: <http://dx.doi.org/10.1016/j.jval.2014.10.005>
- Polišenská, K., Chiat, S., & Roy, P. (2015). Sentence repetition: What does the task measure?. *International Journal of Language & Communication Disorders*, 50(1), 106-118. <https://doi.org/10.1111/1460-6984.12126>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667-696. <https://doi.org/10.1080/00273171.2012.715555>
- Reynolds, C. R., & Suzuki, L. (2013). Bias in psychological assessment: An empirical review and recommendations. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of Psychology Vol 10: Assessment Psychology* (2<sup>nd</sup> ed., pp. 82-113). Hoboken, NJ: Wiley.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, 98(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Samson, J. F., & Lesaux, N. K. (2009). Language minority learners in special education: Rate and predictors of identification for services. *Journal of Learning Disabilities*, 42, 148-162. <https://doi.org/10.1177/0022219408326221>
- Sanchez, M. T., Parker, C., Akbayin, B., & McTigue, A. (2010). Processes and challenges in identifying learning disabilities among students who are English language learners in three New York state districts. *Institute for Educational Sciences National Center for*

## DIMENSIONALITY OF SENTENCE REPETITION

*Education Evaluation and Regional Assistance*, 85.

Scheffner Hammer, C., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpina, S. E., & Goldstein, B. (2012). Predicting Spanish-English bilingual children's language abilities. *Journal of Speech, Language, and Hearing Research*, 55(5), 1251-1264.

[https://doi.org/10.1044/1092-4388\(2012/11-0016\)](https://doi.org/10.1044/1092-4388(2012/11-0016))

Seeff-Gabriel, B., Chiat, S., & Roy, P. (2008). *Early Repetition Battery*. London: Pearson Assessment.

Simon-Cereijido, G., & Gutiérrez-Clellen, V. A (2009). Cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, 30(2), 315-337. <https://doi.org/10.1017/s0142716409090134>

Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., . . . Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and writing*, 28(7), 959-988. <https://doi.org/10.1007/s11145-015-9557-0>

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1-25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639>

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, 77, 317-334. <https://doi.org/10.1177/001440291107700304>

Taliancich-Klinger, C. L., Bedore, L. M., & Peña, E. D. (2018). Preposition accuracy on a sentence repetition task in school age Spanish-English bilinguals. *Journal of Child Language*, 45, 97-119. <https://doi.org/10.1017/s0305000917000125>

## DIMENSIONALITY OF SENTENCE REPETITION

Tomasello, M., & Merriman, W. E. (1995). *Beyond the name for things: Young children's acquisition of verbs*. Hillsdale: Lawrence Erlbaum Associates.

Wood, C., Fitton, L., Petscher, Y., Rodriguez, E., Sunderman, G., & Lim, T. (2018). The effect of e-book vocabulary instruction on Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research*. [https://doi.org/10.1044/2018\\_jslhr-1-17-0368](https://doi.org/10.1044/2018_jslhr-1-17-0368)

Woodcock, R. W. (2011). *Woodcock Reading Mastery Tests, Third Edition (WRMT)*. San Antonio, TX: Pearson.

## DIMENSIONALITY OF SENTENCE REPETITION

Table 1.

*Reported Language Use and Linguistic Background of Participants*

	% of Respondents
<hr/> <hr/> <b>Primary Language of Child</b> <hr/> <hr/>	
English	5%
Primarily English, some Spanish	4%
English and Spanish Equally	34%
Primarily Spanish, some English	10%
Spanish	45%
<hr/> <hr/> <b>Reported Country of Origin of Family</b> <hr/> <hr/>	
Mexico	56%
Guatemala	14%
El Salvador	10%
Cuba	6%
Honduras	5%
Other Country	9%

## DIMENSIONALITY OF SENTENCE REPETITION

Table 2

*Participants' Standard Scores on Language and Literacy Assessments*

Assessment	Kindergarten ( <i>n</i> = 138)			First Grade ( <i>n</i> = 153)			Full Sample ( <i>n</i> = 291)		
	M	<i>SD</i>	<i>n</i>	M	<i>SD</i>	<i>n</i>	M	<i>SD</i>	<i>n</i>
PTONI	95.10	19.42	128	90.28	18.46	148	95.52	19.03	276
WRMT-III Letter ID	87.89	18.40	57	96.66	11.73	73	92.82	15.58	130
WRMT-III PA	96.50	16.05	26	93.71	19.37	63	94.53	18.41	89
WRMT-III RAN	95.22	11.00	73	97.38	13.65	130	96.61	12.77	203
WRMT-III Readiness	96.30	13.92	20	95.70	15.06	54	95.86	14.67	74
PPVT-4	75.84	17.27	128	84.95	12.17	152	80.79	15.38	280
TVIP	81.05	19.34	80	86.52	16.27	90	83.95	17.94	170
Spanish SR <sup>1</sup>	8.46	3.64	127	7.94	3.45	112	8.21	3.55	239
English SR <sup>1</sup>	7.62	3.43	113	9.94	2.66	104	8.73	3.29	217

<sup>1</sup>Spanish and English sentence repetition scores are reported as scaled scores, where the normative mean = 10 and *SD* = 1.5.

DIMENSIONALITY OF SENTENCE REPETITION

Table 3

*Model Fit Indices*

Model	$\chi^2$	df	$\Delta\chi^2$	RMSEA	LB	UB	CFI	TLI	AIC	Adj. BIC
A. Spanish 1-Factor	879.29	629	---	0.038	0.032	0.044	0.976	0.975	9707.61	9740.35
Spanish 2-Factor	878.33	628	1.21	0.038	0.032	0.044	0.976	0.975	9706.73	9739.90
Spanish 3-Factor	862.63	626	22.43*	0.037	0.031	0.043	0.978	0.976	9699.24	9733.30
Spanish 4-Factor	857.73	623	32.32*	0.037	0.031	0.043	0.978	0.976	--	--
B. Spanish 5-Factor	848.27	619	46.67*	0.037	0.030	0.043	0.978	0.977	--	--
C. Spanish BiFactor (5)	802.32	593	106.78*	0.036	0.029	0.042	0.980	0.978	--	--
Spanish BiFactor (3)	812.66	602	91.05*	0.036	0.029	0.042	0.980	0.978	9677.55	9832.70
Spanish BiFactor (2)	828.48	611	70.42*	0.036	0.030	0.042	0.979	0.978	9680.27	9784.41
D. English 1-Factor	722.59	495	---	0.043	0.036	0.050	0.955	0.952	7272.75	7295.94
English 2-Factor	710.23	494	17.10*	0.042	0.035	0.049	0.957	0.954	7266.17	7289.72
English 3-Factor	708.21	492	22.80*	0.042	0.035	0.049	0.957	0.954	7265.32	7289.56
English 4-Factor	705.07	489	28.05*	0.042	0.035	0.049	0.957	0.954	--	--
English 5-Factor	700.50	485	35.54*	0.042	0.035	0.049	0.957	0.954	--	--
E. English 6-Factor	695.29	480	42.91*	0.042	0.035	0.049	0.957	0.953	--	--
F. English BiFactor (6)	678.28	463	67.01*	0.043	0.036	0.050	0.957	0.951	--	--
English BiFactor (4)	632.44	464	118.37*	0.038	0.030	0.045	0.967	0.962	7233.20	7401.18
English BiFactor (3)	637.27	470	107.81*	0.038	0.030	0.045	0.967	0.963	7225.39	7356.85
English BiFactor (2)	640.20	473	101.16*	0.038	0.030	0.045	0.967	0.963	7223.25	7314.82
English BiFactor (1)	651.25	483	77.67*	0.037	0.030	0.044	0.967	0.964	7224.06	7251.47

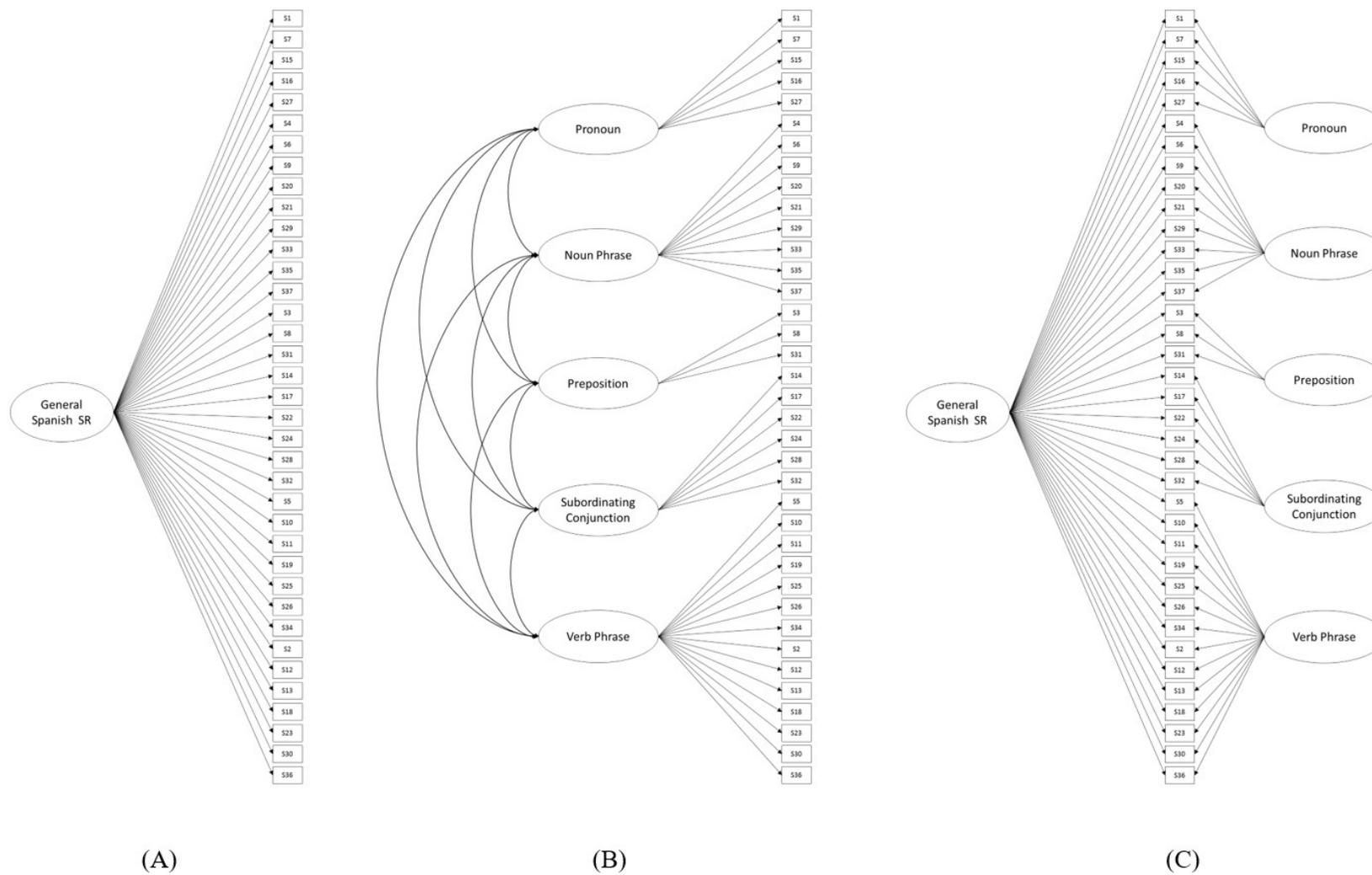
*Note.*  $\Delta\chi^2$  is reported for the model comparisons against the 1-factor model. \* $p < .001$

The bifactor models are identified by the number of specific factors included, which is reported in parentheses.

# DIMENSIONALITY OF SENTENCE REPETITION

Figure 1

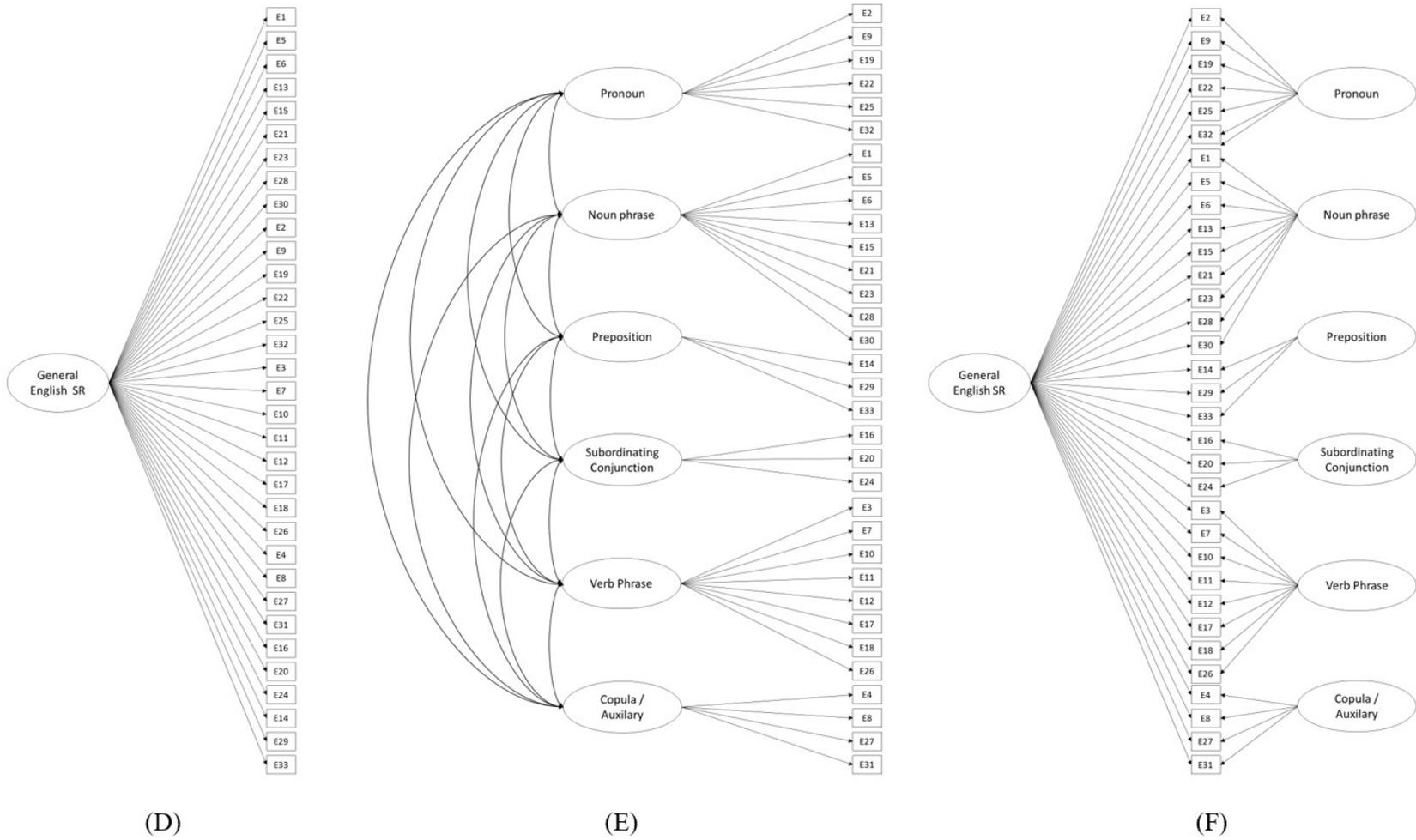
*Spanish Sentence Repetition Models Tested*



# DIMENSIONALITY OF SENTENCE REPETITION

Figure 2

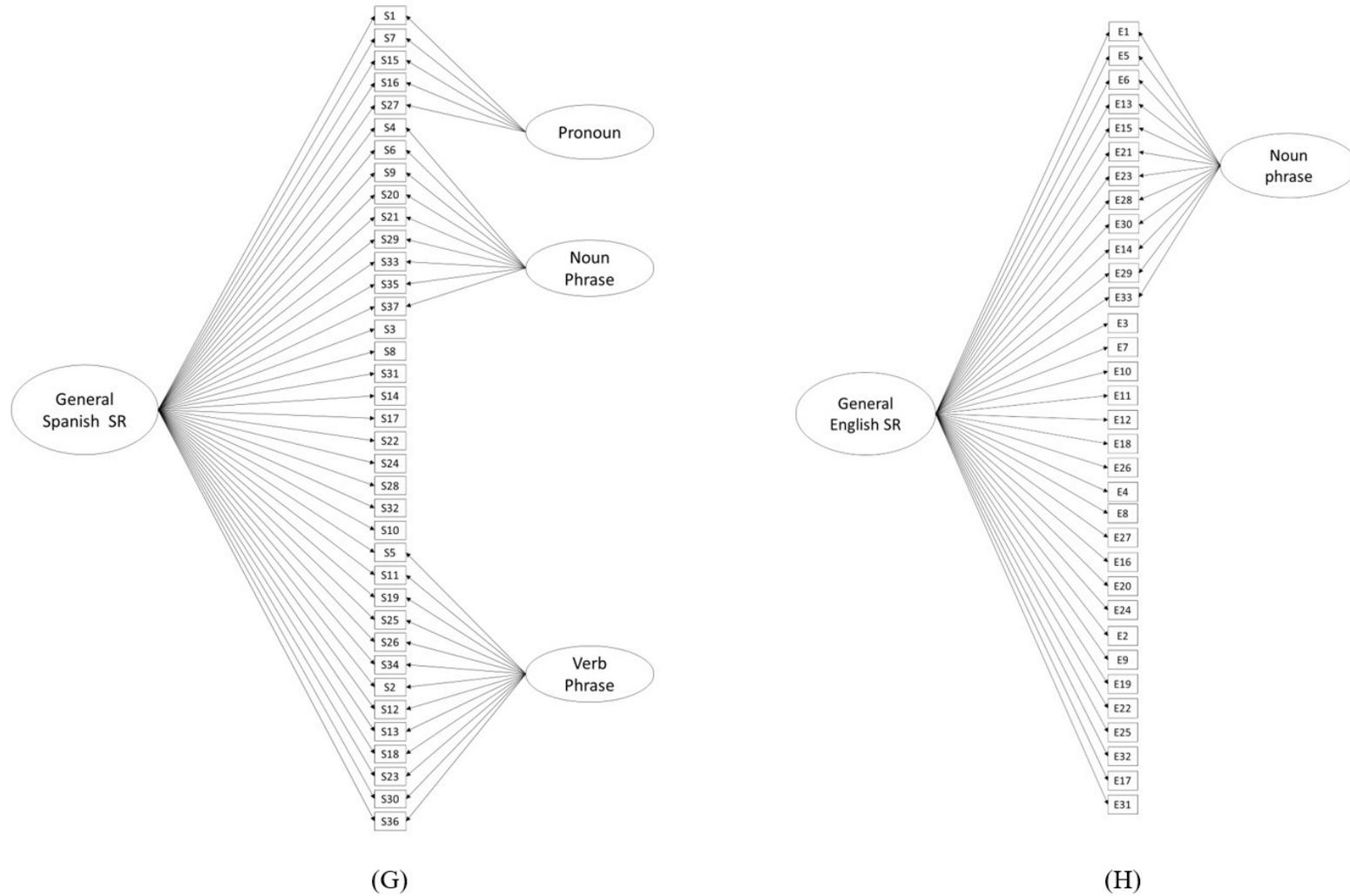
*English Sentence Repetition Models Tested*



# DIMENSIONALITY OF SENTENCE REPETITION

Figure 3

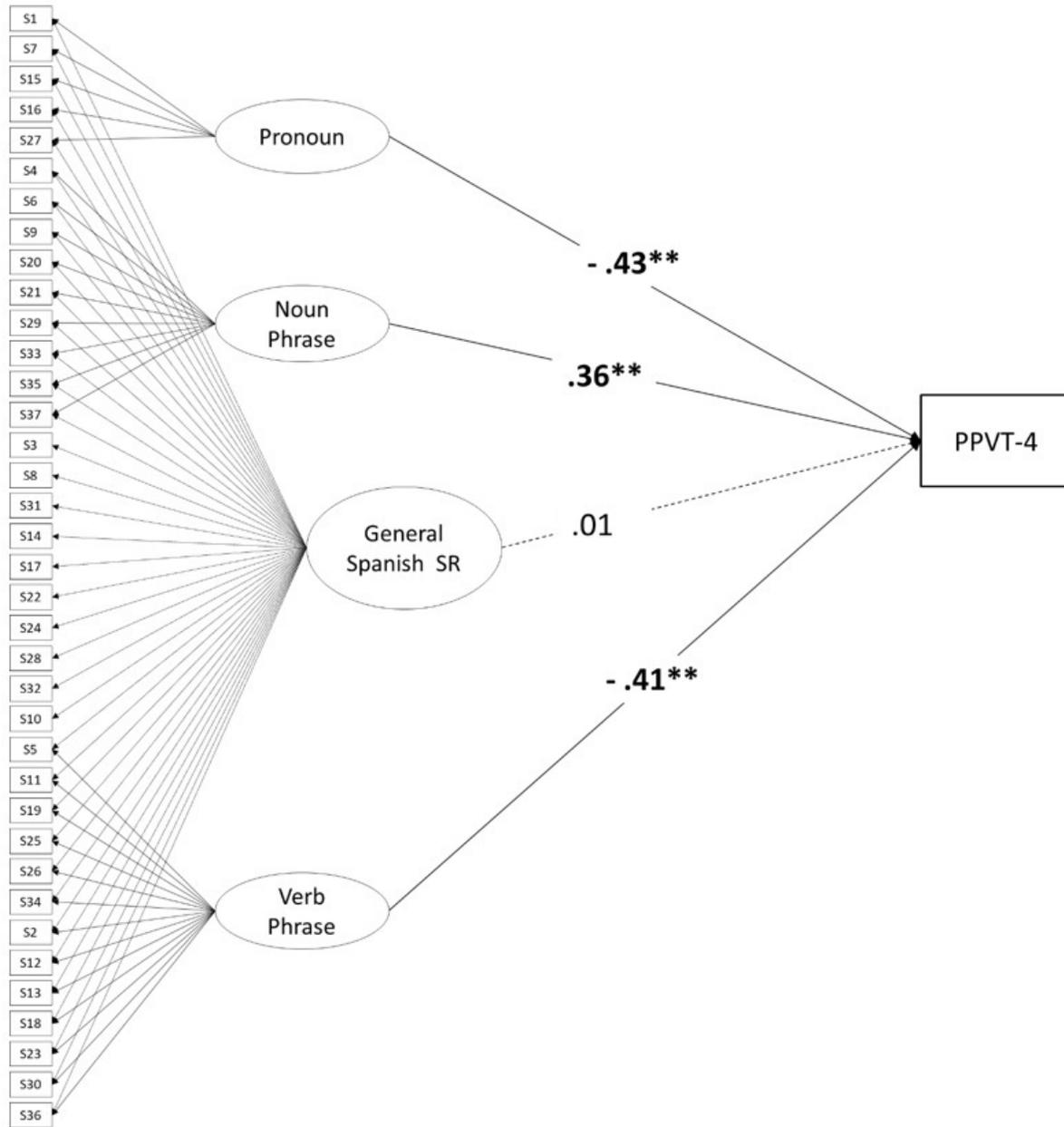
*Final Spanish and English Sentence Repetition Models*



# DIMENSIONALITY OF SENTENCE REPETITION

Figure 4

*Spanish Sentence Repetition Predicting PPVT-4 Performance*



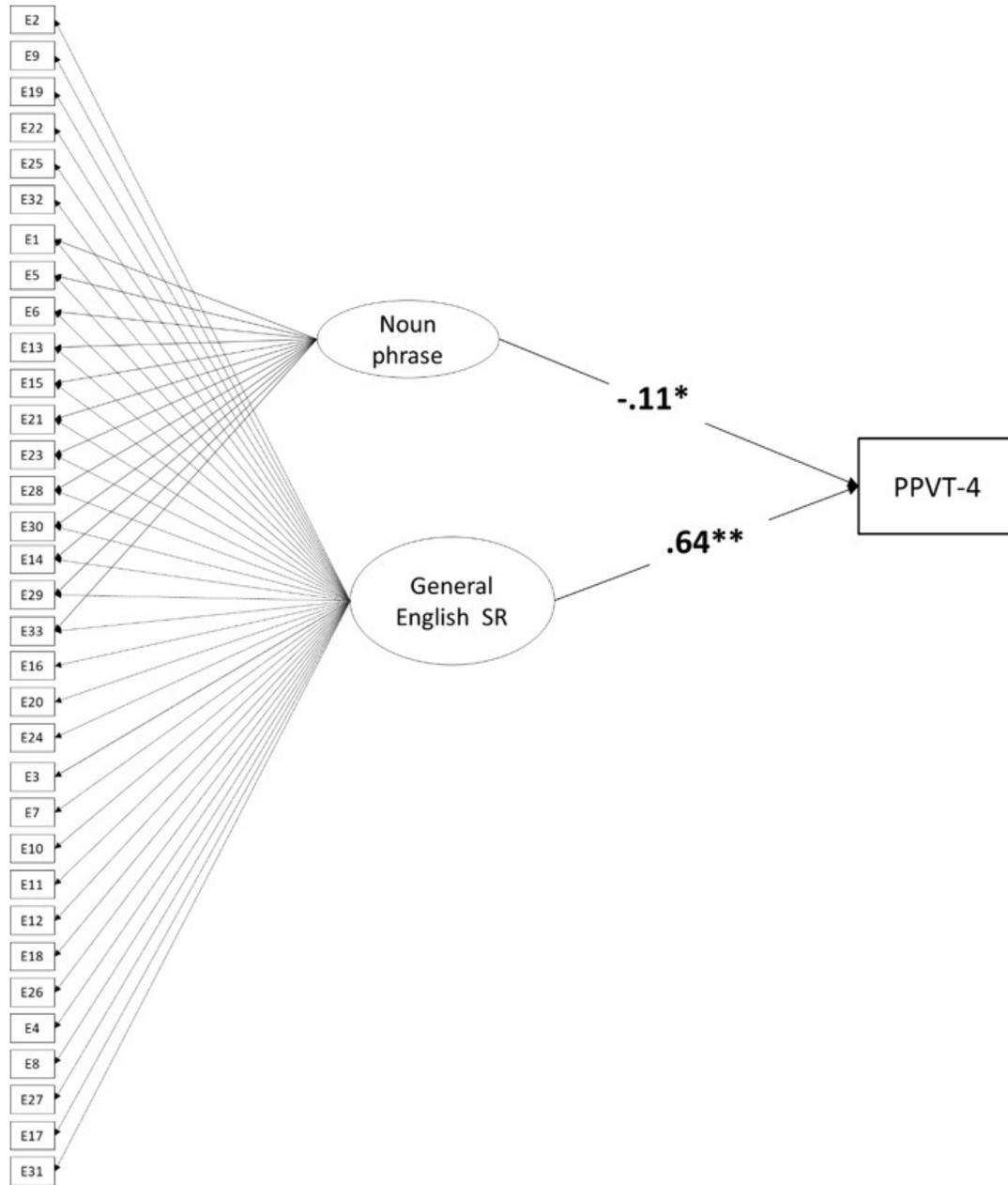
*Note.* Path coefficients reported are the standardized values.

\*\* $p < .001$

# DIMENSIONALITY OF SENTENCE REPETITION

Figure 5

*English Sentence Repetition Predicting PPVT-4 Performance*



*Note.* Path coefficients reported are the standardized values.

\* $p = .057$  \*\* $p < .001$