Comparing Methods for Assessing the English Grammatical Development

of Spanish-Speaking English Learners

Lisa Fitton

Mary Claire Wofford

Kristina N. Bustamante

Natasha De Novi

Bibiana Nuñez

Carla L. Wood

School of Communication Science & Disorders

Florida State University

Tallahassee, FL

Abstract

Identifying valid and informative approaches for assessing young English learners (ELs) is essential for school-based speech-language pathologists (SLPs). The present paper focuses on two distinct approaches for assessing English grammatical development, a key component and indicator of language development for ELs. The approaches addressed are standardized norm-referenced assessment and grammatical feature coding from story retell narratives. The authors review the utility of these approaches for evaluating the English grammatical development of Spanish-speaking ELs. A research example is provided to illustrate how a small sample ($n = 18$) of ELs performed on each of these English-based tasks. Findings reveal that children tended to demonstrate higher percent accuracies on the grammatical forms targeted within the narrative context compared to the norm-referenced contexts. These results, coupled with the background literature, indicate that the two approaches yield different information for individual children. Norm-referenced assessment allows SLPs to target specific grammatical structures and to obtain measures of children's mastery of those structures, whereas narrative-based assessment provides a naturalist measure of functional grammar use. Recommendations for practice are provided.

Keywords: Assessment; Bilingualism; Morphology, Language; Children

The number of Spanish-speaking English learners (ELs) being educated in the United States public school systems continues to grow (Kena et al., 2016). Consequently, identifying valid and reliable approaches to evaluate ELs' language abilities is an increasingly important topic among school-based speech-language pathologists. The purpose of this paper is to present and compare two alternative options for the valid and effective evaluation of the English grammatical development of young Spanish-speaking ELs.

Grammar, alternatively termed morphosyntax, is of interest clinically as an indicator of atypical language development in ELs (Goodwin, Huggins, Carlo, August, & Calderon, 2013; Gutiérrez-Clellen & Simon-Cereijido, 2007). Spanish-speaking ELs with speech-language impairment (SLI) have been observed to demonstrate deficits in grammatical development that are similar to their monolingual English-speaking peers and monolingual Spanish-speaking peers with SLI (Simon-Cereijido, & Gutiérrez-Clellen, 2009; Squires et al., 2014). In a comparison of Spanish-English bilingual children with SLI to a group of English-speaking monolingual children with SLI, both groups were observed to exhibit similar grammatical error patterns (Gutierrez-Clellen, Simon-Cereijido, & Wagner, 2008), suggesting that grammar is a consistent indicator of atypical language development among Spanish-English bilingual children.

ELs' grammatical abilities can also serve as a developmental marker for progress monitoring. The emergence of specific grammatical structures and the use of certain types of grammatical errors can provide insight into English language learning. However, ELs and English-speaking monolingual children do not share an entirely consistent language trajectory with identical language milestones (Gutierrez-Clellen, 2002). For example, incorrect word combinations or subject omissions are typical error patterns observed among Spanish-speaking ELs (Bedore & Pena, 2008; Gutierrez-Clellen, Simon-Cereijido, & Wagner, 2008). Additionally,

typically developing bilingual children are more likely to overgeneralize the past *-ed* form (e.g. *falled* for *fell*) than bilingual children with SLI, who are more likely to use the infinitive form (e.g. *fall* for *fell*; Jacobson & Schwartz, 2005). These examples are likely the result of the over-generalization of native language-specific grammatical rules English grammar. For SLPs with an understanding of these typical patterns of bilingual grammatical development, these markers can be useful benchmarks for monitoring the progression of English language development among ELs (Bedore et al., 2010).

To make appropriate inferences from assessment of ELs' English grammatical development, however, testing procedures that yield valid and informative results are essential. There are many options to assess grammar in English, but selecting the most appropriate tool for specific situations can be challenging. Options range from standardized, norm-referenced approaches to dynamic testing for individualized assessment. Given the unique characteristics and challenges of dual language development (see Kohnert, 2010), careful assessment planning is warranted.

In the following literature review, we review two approaches for assessing ELs' English grammatical development. The first is norm-referenced assessment. This approach is popular in schools due to its utility in providing a numeric value to represent a child's grammatical skills, which can be compared against other children of similar age or grade. The second approach discussed is narrative-based assessment using grammatical feature coding. Elicitation of story retell narratives is a method of language sampling that has become increasingly useful with the development of normative databases against which to compare measures obtained from the sample (e.g., Miller & Iglesias, 2015). Both approaches are clinically valuable, but each has distinct advantages and disadvantages.

Using Norm-Referenced Assessment to Evaluate Grammatical Development

Currently there are relatively few standardized, norm-referenced assessments specifically designed for evaluating the English grammatical development of Spanish-English bilingual children. For a test to be valid for a bilingual child, the measure must be normed on a bilingual population (Bedore, Peña, Gillam, & Ho, 2010). Given the differences in development noted between monolingual and bilingual children (August & Shanahan, 2006; Jacobson & Schwartz, 2005; Peña & Bedore, 2008), standard scores obtained based on a monolingual sample are not representative of ELs' overall language ability. Consequently, care is needed when selecting a norm-referenced assessment to evaluate an EL.

The *Bilingual English-Spanish Assessment* (BESA; Pena et al., 2014) is one of few standardized, norm-referenced measures designed specifically for assessing the language development of Spanish-English bilingual children. Importantly, the test includes a section for evaluating English grammatical development: the English morphosyntax subtest. Because of this feature and the use of a bilingual normative sample, the BESA was selected as an illustrative example of an appropriate standardized, norm-referenced tool for assessing grammar in young Spanish-speaking ELs.

The BESA includes subtests of specific linguistic domains, including pragmatics, morphosyntax, and semantics. Each subtest has an English and Spanish version, with subtests designed to be sensitive to each language's unique linguistic features (i.e., the subtests are not based on a translation from one language to the other). The morphosyntax subtest is composed of two parts each requiring a different response from the child. In the cloze section, children are required to complete cloze tasks where they must produce a morphologically inflected word to complete a sentence. In English, the cloze task elicits eight

specific grammatical features (e.g., past tense *-ed*, plural *-s*, passives). In the sentence repetition section, children must repeat verbally presented sentences verbatim. Their responses are scored based on children's production of 2-5 specific word targets within each sentence (Peña et al, 2014).

The primary benefit of employing a norm-referenced approach to grammatical assessment is that norm-referenced tests provide a comparison of children's performance against a normative group. The scores derived from these tests can help communicate children's performance to non-experts, and can offer a benchmark against which to compare children's development as measured by the test. Norm-referenced scores are often required for decision-making in schools (Ireland & Conrad, 2016) and can facilitate the diagnosis of SLI when considered appropriately (Kaderavek, 2015).

For clinicians to leverage this advantage of norm-referenced assessment and make comparisons against the normative sample, they must follow the test's standardized procedures. Any adaptation to assessment procedures can invalidate normative comparisons (Zucker, 2004). If modifying the normative procedures would inform intervention planning, clinicians would therefore need to administer the test twice: once following standardized procedures to obtain a norm-referenced score (Zucker, 2004) and again using adapted procedures. However, this approach is not convenient, given the time constraints in school settings and the potential for learning effects with repeated administrations. Further, deviating from a test's standardized procedures is occasionally unavoidable. For example, standardized administration procedures may result in the child reaching ceiling performance, ending the assessment, before completing sufficient items to quantify performance. The current exemplar, the BESA, requires children to respond verbally to adult prompts given a picture context or decontextualized speech. Children

who have not been exposed to similar types of experiences may not be able to complete this task

sufficiently to provide a score that is valid and reliable (Laing & Kamhi, 2003).

Another option for clinicians seeking to use norm-referenced assessment results to inform

intervention planning is to conduct error analysis after administering the test according to

standardized procedures (American Speech-Language-Hearing Association, 2004). To conduct

informative error analysis using a test of grammatical development for an EL, recognition of the

grammatical structures that have been shown to distinguish EL children with and without SLI is

essential (see Gutiérrez-Clellen & Simon-Cereijido, 2009; Gutiérrez-Clellen et al., 2008;

Kohnert, 2010). Forms that have been shown to be indicators of SLI among Spanish-English

bilingual children include marking verb tense (e.g., past tense *-ed*, third person singular *-s*,

auxiliary *be* and *do*, and copula *be*; Gutiérrez-Clellen et al., 2008; Jacobson & Walden, 2013)

and other less-salient grammatical forms (e.g., possessive *–s*, plural *-s*; Bedore & Peña, 2008).

Using Narratives to Evaluate Grammatical Development

Narrative assessment allows children to produce authentic grammatical structures

spontaneously in the context of a culturally relevant task (Cleave et al. 2010). Narrative retells

are shown to be naturalistic, ecologically valid means of evaluating Spanish-English bilingual

children's grammatical knowledge in their two languages (Bedore et al., 2010; Cleave,

Girolametto, Chen, & Johnson, 2010; Miller, Gillam, & Peña, 2001).

Narrative retell tasks require several simultaneous processes of examinees, including

construction of a mental model of the narrative, inference-making pertinent to the narrative's

content, and reproduction of linguistic forms in a coherent retelling (Gutiérrez-Clellen, 2002).

Spontaneous oral narrative retells have limited verbal memory or attention demands, allowing

processing resources to be devoted to construction of complete narratives. Thus, the examinee is

also allowed greater linguistic agency over grammatical forms used in the retell. Notably, however, ELs experience the added difficulty of navigating appropriate grammatical structures in the developing second language while constructing a story retell.

Numerous measures used in oral narrative retell assessment analysis tap into grammatical development. Discussed subsequently are several commonly-used options, which can be considered as broader, global measures of grammar (e.g. mean length of utterance, percentage grammaticality, subordination index) or structure-specific measures that may be more clinically-meaningful for practitioners and researchers. Three measures discussed below are designated as global because they produce single-number values that aggregate grammatical performance at the transcript level; however, each of the global measurements can represent a distinct aspect of morphosyntactic development. In contrast, structure-specific measures results in multiple numeric values that represent occurrences of specific grammatical features within a narrative.

Global Measures of Grammatical Development. Mean length of utterance (MLU), a measure frequently used for grammatical analysis in language samples, is obtained by transcribing a child's narrative, segmenting it into utterances, and calculating the average number of units (e.g., morphemes, words) per utterance. This measure can be computed manually or obtained using software designed to facilitate language sample analysis, such as *Systematic Analysis of Language Transcripts* (SALT; Miller & Iglesias, 2015). MLU is widely used with young children because it can be sensitive to differentiating typically-developing bilingual children from those with language impairment (Bedore et al., 2010; Gutiérrez-Clellen & Simon-Cereijido, 2009). Researchers have employed several variants of MLU in practice, including the mean length of the longest five utterances in a sample (Hipfner-Boucher et al., 2015), mean

length of terminable unit (MLTU), and mean length of response--words (MLR-w; for review, see Gutiérrez-Clellen, 2002).

A second broad measure of grammaticality is the percentage of grammatically correct utterances produced in an EL's oral narrative retell. It is calculated by dividing the number of grammatically correct utterances by the total number of utterances in the sample. Grammaticality is considered a measure of overall narrative quality and has been suggested to be a less-biased predictor of language and grammatical ability for bilingual children (Bedore et al., 2010; Fiestas & Peña, 2004). However, considerations are needed when using this grammaticality measure to evaluate ELs' narrative retells. ELs' English grammatical development does not mirror monolingual grammatical development (Gutiérrez-Clellen, 2002), and young ELs' ability to construct complex narratives precedes their mastery over all grammatical structures (Bedore et al., 2010; Gutiérrez-Clellen, 2002). Typically-developing ELs therefore often exhibit decreased grammaticality in narrative retells compared to their monolingual peers (Bedore et al., 2010).

A third global measure of grammar proficiency obtained in a narrative context is the subordination index (SI). SI is calculated by identifying the number of independent and subordinate clauses in a narrative and dividing it by the total number of utterances produced. It is intended to reflect the complexity of morphosyntactic structures used in narrative production (Heilmann, Miller, Nockerts, & Dunaway, 2010; Miller et al., 2001), whereas MLU and grammaticality respectively capture productivity and quality. Because grammatical development hinges on lexical knowledge SI may be a good indicator of more complex vocabulary use (Marchman et al., 2004). In a study of young EL narratives, SI was demonstrated to be correlated with lexical measures and predictive of macrostructural scores (i.e. NSS; Lucero, 2015), indicating that SI captures grammatical complexity beyond simple measurement of length.

Considerations for Use of Global Measures. All three of these global measures have been shown to be useful as measures of grammatical development in ELs, whether for distinguishing children with language impairment (Bedore et al., 2010; Gutiérrez-Clellen & Simon-Cereijido, 2009) or for predicting later language development (Lucero, 2015). Clinically, these global measures are appealing because they are relatively efficient in the context of language sample analysis (Heilman, 2010; Heilmann, Miller, & Nockerts, 2010). However, quick measures of grammar may not offer the richness that structure-specific analyses of grammar provide. MLU calculations, for example, necessarily exclude unintelligible utterances. For children who speak with many revisions and errors, MLU may be adversely affected and potentially meaningful data may be overlooked. Additionally, targeting MLU as a treatment goal may not provide sufficient guidance for clinicians; provision of additional short-term goals that illuminate specific structures for intervention can enhance a treatment plan. These global measures may be more meaningful if observed in concert with finely-grained analyses of grammatical structures in narratives, which can inform specific short-term and long-term intervention planning.

Grammatical Feature Coding. Another option for assessing grammar in ELs is direct observation of grammatical features, which mirrors error analysis as discussed previously. It is well established that ELs may overgeneralize, omit, or apply first language rules to emerging grammatical features, leading to errors in oral language (Bedore et al., 2010; Jacobson & Walden, 2013; Jacobson & Schwartz, 2005). Additionally, bilingual children with SLI are at increased risk for incorrect use of grammatical forms that have perceptually low salience, which in English heavily consists of verbal morphology (Restrepo & Gutiérrez-Clellen, 2004; Gutiérrez-Clellen et al., 2008). Coding for specific structures within the narrative retell that tend

to be problematic for ELs can illustrate the preferred and mastered structures or identify

structures produced in error. Analysis of those errors and the contexts in which they occur can

provide valuable information about EL language development not observable in standardized

assessments. Specific grammatical feature codes must be inputted into *SALT* manually per the

established transcription conventions to generate a Code Summary report.

A solution to enhancing understanding of grammatical ability is to supplement global

measures of grammar with grammatical feature coding (see Miller, 1981). Gutiérrez-Clellen and

Simon Cereijido (2009) combined MLU, grammatical accuracy, and several significant

morphosyntactic structures in a verb morphology composite score to assess grammar in both

English and Spanish narratives. The percentage of correct productions on select English verb

forms was compared to the number of obligatory contexts for the same verb forms. In Spanish,

targeted structures in the verb morphology composite included articles, verb agreements, clitics,

transitive, and ditransitive verbs. This strategy used converging grammatical evidence derived

from narratives that utilized both global measures of grammaticality and language-specific

structural analysis to identify language impairments.

Purpose of the Present Study

This literature review broadly focused on standardized norm-referenced assessment and

narrative-based grammatical feature coding for evaluating English grammatical development

among ELs. Overall, there is sufficient rationale for each approach. However, the specific results

that can be obtained from each assessment warrant further description. The purpose of the

present work is to present practical examples of similarities and differences between findings

obtained from norm-referenced testing and from the grammatical feature coding of ELs' story

retell narratives. The results obtained from error analysis of each are compared and

recommendations are provided for clinical practice.

Method

Participants

Participants were 18 preschool and kindergarten Spanish-speaking ELs. The children

were recruited from elementary schools in a rural county of Florida and ranged in age from 4.62

years to 6.26 years ($M = 5.58$, $SD = 0.53$). According to data from the Stanford Education Data

Archive, the county has a school-age population that is 76.14% African American, 18.50%

Hispanic, 3.7% White, and less than 1% Asian or Native American. Most of the children

(77.49%) receive free or reduced-price lunch, and 7.12% are identified as English Language

Learners (Reardon et al., 2016). All the participants' parents reported speaking primarily Spanish

at home, and the schools used English-only curriculums. All the children were reported to have

been born in the United States. Additional background information is shown in Table 1.

Table 1

Demographic Information

	n	%		n	%
Child Gender			**Parent Preferred Language**		
Male	10	55.56	Spanish	13	72.22
Female	8	44.44	Spanish and English	2	11.11
Child Preferred Language[1]			No Response	3	16.67
Spanish	14	77.78	**Parents' Country of Origin**		
Spanish and English	2	11.11	Mexico	7	38.89
English	2	11.11	Guatemala	2	11.11
Socioeconomic Status			El Salvador	3	16.67
Free lunch	18	100	Other/Multiple	6	33.33

[1]As reported by the child's parent during phone interviews.

Procedures

Upon receipt of informed consent from participants' parents, the children completed all testing during the 2016-2017 academic year. The test battery included hearing screenings (Armstrong, 2012), the *Primary Test of Nonverbal Intelligence* (PTONI; Ehrler & McGhee, 2008), the morphosyntax and semantics subtests of the *Bilingual English-Spanish Assessment* (BESA; Peña et al., 2014) in English and Spanish, and story retell narratives (Miller & Iglesias, 2015). Trained undergraduate research assistants administered the tests during four hour-long blocks conducted at the participants' schools under the supervision of a licensed, Spanish-English speaking speech-language pathologist. English-monolingual research assistants administered the English-based tests, and Spanish-English bilingual assistants administered the Spanish-based tests. Background interviews were conducted with participants' parents over the phone by research assistants. All procedures were reviewed and approved by the Human Subjects Committee Institutional Review Board at Florida State University (HSC 2017.20422).

Standardized Measures

The *Primary Test of Nonverbal Intelligence* (PTONI; Ehrler & McGhee, 2008) is a picture-based test designed to quantify young children's nonverbal cognitive abilities. The test yields internal consistency reliability of .93 as measured using coefficient alpha, test-retest reliability of .97 and inter-rater reliability of .99. The test includes limited verbal instructions that can be presented in the child's dominant language and does not require a verbal response.

The *Bilingual English-Spanish Assessment* (BESA) is a standardized test designed to provide a metric of global language performance for young Spanish-English bilingual children (Peña et al., 2014). The test includes Spanish and English forms, and was created to allow children to respond in either language. It provides normative data for children ages 4;0-6;11

based on a sample from over 600 children living in the United States. Each subtest of the BESA

requires 15-20 minutes to complete. Scores from the morphosyntax subtest alone yield a

sensitivity rate of 92% and specificity of 86%. The semantics subtest has a sensitivity rate of

81% and a specificity rate of 83%. The overall composite provides a sensitivity of 92% and a

specificity of 88% (Peña et al., 2014).

The BESA English cloze is designed to target eight specific grammatical structures.

These include possessive -*s*, third person singular present -*s*, regular past tense -*ed*, plural -*s*,

present/past progressive (auxiliary + progressive -*ing*), copula, negatives, and passive forms.

Each specific structure is targeted by three sequential items presented in a cloze format.

Story retell narratives were elicited using standardized procedures for Systematic

Analysis of Language Transcripts (SALT; Miller & Iglesias, 2015) software. The *Frog, Where*

Are You? (Mayer, 1969) story was first read aloud by an examiner, accompanied by the wordless

picture book as a visual aid, using an English script (Miller & Iglesias, 2015). The script includes

49 total utterances, with 512 total words and 163 different words. The mean length of utterance

in morphemes is 11.53. The child was then asked to tell the story back to the examiner. This

story retell was recorded on an audio recording device.

Narrative transcription and coding. Two research assistants transcribed the audio files

into electronic transcripts using SALT formatting and the standard conventions for c-units

(Miller & Iglesias, 2015). The assistants then identified the obligatory contexts for each of the

grammatical features targeted on the BESA English cloze subtest (Peña et al., 2014) within each

narrative. These grammatical features were selected to provide a direct comparison against the

BESA. The assistants coded the child's response to the obligatory contexts as either correct or

incorrect uses of the targeted feature (see Table 2).

Table 2

Example Coded Excerpt from Narrative Retell

$ Child

C The frog jump/ed [GSP:RPT] out the jar.
C The dog got his head stuck in the jar.
C His head was [GSP:COP] stuck.
C He no [GSM:NEG] move it.
C The kid call/ed [GSP:RPT] the frog.
C The dog was bark/ing [GSP:PP] at the beehive.
C A beaver walk [GSM:RPT] and bit (the) the boy in the nose.
C The bee/s [GSP:PM] were chase/ing [GSP:PP] the dog.

Notes. GSP = Grammatical structure present, GSM = Grammatical structure missing, RPT =

Regular past tense, COP = Copula, NEG = Negatives, PP = Present/past auxiliary + progressive,

PM = Plural marker.

Narrative reliability. Reliability for the narrative transcription and coding was

established between two research assistants. The assistants completed transcription training and

demonstrated at least 95% reliability with the first author on practice transcripts before

transcribing and coding the samples. They cross-checked all the transcripts for both consistency

with the audio recordings and coding accuracy. Reliabilities were computed by dividing the total

agreements by the sum of agreements plus disagreements. Word-level transcription reliability

was measured to be at 99.26%, c-unit segmentation was at 100%, and grammatical coding

reliability was at 99.39%. These discrepancies (less than 1% of coded forms) were resolved by

the first author.

Language Sample Analyses

Several measures of English grammaticality were computed from the BESA and story

retell narratives results. In addition to the norm-referenced standard scores provided for the

BESA morphosyntax subtests, overall percent accuracy was computed for the cloze section. Percent accuracy was obtained by summing the total number of correct grammatical productions within the cloze task, and then dividing this value by the total number of items included on the measure. Structure-specific grammatical accuracy was also obtained for each of the structures targeted on the BESA English cloze. The total number of correct structure-specific productions were summed and then divided by the total number of items in each section.

Similar to the measures computed from the children's item-level responses to the BESA, several supplemental measures were obtained from the narratives. Mean length of utterance (MLU) in morphemes was obtained from the SALT standard measures report (Miller & Iglesias, 2015) as a global measure of grammatical productivity. Additionally, overall percent grammatical accuracy was computed by summing the total number of correctly-produced grammatical forms coded, and then dividing this value by the total number of obligatory contexts for the eight targeted grammatical structures (Miller, 1981). Then, structure-specific grammatical accuracy was computed by summing the total correct productions for each grammatical form, and dividing these by the total obligatory contexts for that specific form.

Results

The participants exhibited nonverbal cognitive abilities within normal limits as measured by the PTONI ($M = 94.88$, $SD = 15.56$). The children scored 100.59 ($SD = 12.51$) on average on the Spanish semantics portion of the BESA and 99.18 ($SD = 16.00$) on average on the English semantics. Twelve children (66.67%) obtained higher standard scores on the Spanish portions of the BESA than on the English portions, suggesting Spanish language dominance. Of the remaining children, four (22.22%) received higher scores on the English portions than on the

Spanish portions, suggesting English language dominance. The final two children (11.11%)

received the equivalent scores for Spanish and English, indicating balanced language dominance.

Grammar as Measured by the BESA

On average, the participating children scored 91.00 (SD = 16.29) on the Spanish

morphosyntax subtest of the BESA. For English morphosyntax, children scored 86.82 on

average (SD = 15.09). The participants' grammatical accuracy on each of the cloze subtests,

broken down by each of the targeted features, is shown in Table 3.

Table 3

Grammatical Accuracy on the BESA Cloze

	M (%)	SD
English		
Possessive -*s*	37.04	37.73
Third Person Singular -*s*	48.15	44.61
Regular Past -*ed*	57.41	39.28
Plural Noun -*s*	72.22	26.20
Present/Past Auxilary + Progressive -*ing*	31.48	38.73
Copula	51.85	30.73
Negatives	38.89	38.35
Passives	12.96	25.92
Spanish		
Articles	79.17	31.21
Present/Past Auxilary + Progressive	64.81	38.73
Direct Object Clitics	48.61	31.47
Subjunctive	47.22	44.46

Grammar in Story Retell Narratives

The children produced an average of 13.44 (SD = 10.94) utterances that were fully

intelligible during the story retells. All the children produced at least one intelligible utterance,

and 93.89% (*SD* = 17.52%) of all the utterances produced were fully intelligible. The children

used an average of 80.61 (*SD* = 75.71) total words and 33.89 (*SD* = 24.97) different words. Their

mean length of utterance (MLU) in morphemes was 5.43 (*SD* = 2.21). The children used 6.72

(*SD* = 7.39) mazes (i.e., filled pauses, repetitions, or revisions) during the retells. Only one of the

narratives included an utterance in Spanish, which was excluded from the grammaticality

analyses. Table 4 provides the participants' grammatical accuracy and the average number of

obligatory contexts produced for each of the features of interest within the story retell narratives.

Table 4

Grammatical Accuracy within Story Retell Narratives

Target Form	n^1	Accuracy (%)	*SD*	Occurrences[2]	*SD*	Freq Script[3]
Possessive -*s*	5	60.00	54.77	0.33	0.59	1
3rd Singular -*s*	6	19.44	40.02	1.50	3.26	0
Regular Past -*ed*	11	81.52	26.56	1.89	2.14	33
Plural Noun -*s*	10	97.50	7.91	1.44	1.69	10
Aux + Prog -*ing*	11	93.18	16.17	2.00	2.38	6
Copula	9	100.00	--	1.06	1.55	12
Negatives	10	86.67	32.20	0.78	0.88	3
Passives	0	--	--	--	--	2

[1]The number of children who produced at least one obligatory context for the grammatical

feature within their story retell narrative.

[2]The mean number of obligatory contexts produced for the feature across the children

[3]The number of times each form was presented in the script for the *Frog Where Are You?* story.

Comparisons Between Narratives and the BESA

Figure 1 displays the participants' performance on individual grammatical forms across

tasks (i.e. narrative retell and the BESA). The children achieved a higher percent grammaticality

on the targeted structures within the narratives (81.28%, *SD* = 29.49%) compared to the English

BESA cloze (43.75%, *SD* = 24.85%). Similarly, the children achieved greater accuracy for most

of the targeted grammatical forms when produced in the narrative context as compared to those

elicited during the English BESA cloze. One exception was the third person singular *-s*, which

was produced with 19.44% (*SD* = 40.02%) accuracy in the narratives and 48.15% (*SD* =

44.61%) accuracy on the cloze subtest. None of the children's narratives included an obligatory

context for a passive form. Therefore, no comparison was possible for that structure.

Figure 1

Grammatical Accuracy in Narrative Context Compared to the BESA Cloze Task
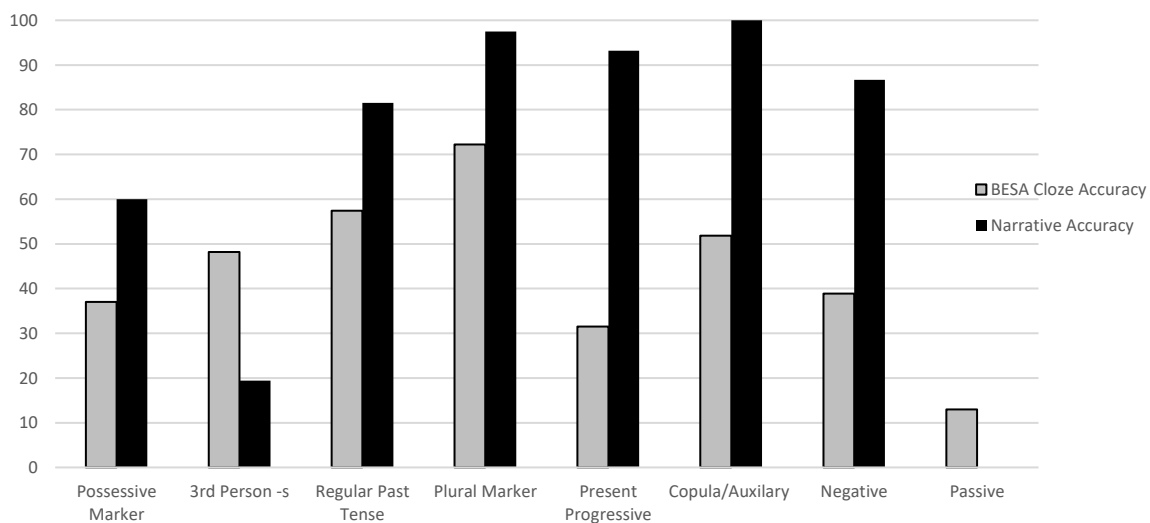
Correlations obtained between the children's grammatical accuracy on the BESA cloze as

compared to the narratives revealed few between-measure relations. The children exhibited some

similarities in their accuracy on the various structures targeted within the BESA (see Table 5),

but no significant relations existed between grammaticality on the BESA and grammaticality

within the narrative context. Further, children did not demonstrate consistent performance on the

separate structures as measured within the narrative context.

Table 5

Correlations within the BESA English Cloze Task

Variable	1	2	3	4	5	6	7	8	9
1. Total Raw Score	1								
2. Possessive -*s*	**.715****	1							
3. Third person -*s*	**.936****	**.742****	1						
4. Past tense -*ed*	**.699****	**.774****	**.792****	1					
5. Plural -*s*	**.521***	0.176	**.485***	0.18	1				
6. Progressive	**.701****	0.318	**.622****	0.246	0.204	1			
7. Copula	**.556***	0.106	0.36	0.15	0.027	**.580***	1		
8. Negative	**.802****	0.256	**.637****	0.34	**.618****	**.579***	**.629****	1	
9. Passive	**.539***	0.416	0.389	0.125	0.369	0.221	0.255	0.449	1

*p < .05

** *p* < .01

Discussion

Overall, the participants obtained higher percent accuracies on each of the grammatical forms within the narrative context compared to the English BESA Cloze task. An exception was the third person -*s*, which children produced with greater accuracy on the norm-referenced task. None of the children attempted to produce a passive form in the narrative context. Consequently, comparison of passive tense accuracy across tasks was not possible. Among the grammatical forms that were produced in both contexts, no significant correlations were obtained between the BESA Cloze and narrative-based measures of accuracy.

None of the participants produced obligatory contexts for all the grammatical forms targeted by the English BESA Cloze task in their story retell narratives. At least half of the children did produce obligatory contexts for the regular past tense -*ed*, plural -*s*, copula, and negative forms. However, the children produced an average of 2.00 obligatory contexts or fewer for each of the targeted forms, resulting in relatively few opportunities for assessing grammatical accuracy. By contrast, the English BESA Cloze task provides 3 obligatory contexts in which to assess children's use of each of the grammatical forms of interest.

Considerations in Comparing Differences in Tasks and Measures

One advantage to utilizing formal standardized tasks such as the BESA cloze subtest is the elicitation of specific, discrete grammatical forms. Although the open-ended nature of narrative language samples may be advantageous for other reasons, one disadvantage is the potential absence of obligatory contexts for all desired grammatical forms, particularly among young children who may tend to produce shorter and less complex narratives (Heilmann, Nockerts, & Miller, 2010; Miller & Chapman, 1981). This weakness was noticeable in the current results. The children did not produce as many of the targeted grammatical forms in their

narrative retells as were elicited through the norm-referenced assessment. Children may avoid

generating utterances in a narrative retell that contain complex forms. The overreliance on earlier

developing forms in narrative samples does not provide information about whether the more

complex forms are emerging or developed but not used. For example, conditional forms are

rarely observed in narrative retells but the absence of the form may not indicate that the child has

not mastered it. Rather, the children simply did not demonstrate its use.

        In response to this potential limitation of narrative retells, clinicians and researchers have

utilized available tools, such as the *Narrative Assessment Protocol* (NAP), when formulating

narrative retell scripts to ensure specific grammatical forms are modeled and therefore likely to

be elicited by the story model. Additionally, the NAP is used to systematically examine the

specific structural conventions present in the child's retell (NAP; Pence, Justice, & Gosse, 2007).

The *NAP* includes elements in the categories of sentence structure (e.g. compound, complex,

negative, interrogative), phrase structure (e.g. elaborated noun phrase, compound noun,

prepositional phrase), modifier (e.g. adverbs, advanced modifiers), nouns (e.g. plural, possessive,

tier two), and verbs (e.g. auxiliary + main, irregular past, regular past, tier two, compound). It

has been adapted and validated for Spanish-language narrative samples in both a long- and short-

form (Gorman et al, 2016). The NAP score sheet can be completed in real time as the examinee

produces unique NAP elements during the oral narrative retell (Justice et al., 2010).

Clinical Recommendations

        Given the differences in results obtained from the norm-referenced assessment of English

grammar compared to those obtained through grammatical feature coding in narratives, it is

unlikely that these two approaches to grammar assessment are interchangeable in clinical

practice. Considering previous research and the results obtained from the current sample of

children, the approaches are likely appropriate for different purposes and contexts. Specifically,

norm-referenced assessment may be most beneficial when conducting a baseline assessment. As

long as the normative sample is appropriate for the child being evaluated, a norm-referenced test

can provide a valuable numeric comparison for considering the child's skill level. Additionally,

this approach allows clinicians to target specific grammatical structures to assess children's

mastery of those forms. With the inclusion of error analysis, a norm-referenced grammatical

assessment can be informative for preliminary evaluation reporting and intervention planning.

Although some standardized, norm-referenced language measures include ELs in their

normative samples and therefore offer valid comparisons, the available standardized measures

are limited in that they elicit decontextualized on-demand responses. Consequently, narrative-

based assessment is sometimes valuable for obtaining measures of grammaticality in a natural

communication context. Inclusion of grammar coding may provide supplemental information

within language sample analysis. Although limited in that specific grammatical forms may not be

prompted frequently, if at all, in a narrative context, the administrator can modify procedures to

elicit structures of interest (e.g., the NAP). Further, when clinicians are seeking efficient, brief

measures of language, a narrative assessment may be ideal because of the rich information that it

holds. Although a narrative requires only a few minutes to elicit (Miller & Iglesias, 2015),

analyses of the resulting sample can produce multiple measures that provide insight into

language productivity, diversity, and grammaticality.

Additionally, the narrative-based approach to assessment of specific grammatical features

may provide useful information for progress monitoring. Narrative retells are quick measures

that can provide multiple benchmarks simultaneously. Consequently, clinicians may find these

retells highly valuable. In addition to being valid and reliable, progress monitoring tools must be

efficient, sensitive to growth, and robust to test-retest effects (Petersen & Spencer, 2012). Story

retell narratives meet these criteria, particularly when different stories are employed for the

retells (Peterson & Spencer, 2012). To obtain a measure of progress on English grammar use for

an EL, SLPs may elicit a narrative and keep track of which forms are attempted, whether

correctly or incorrectly. These narrative-based measures of grammar would likely reveal

incremental growth over time (e.g., Miller, 1981; Gutiérrez-Clellen & Simon Cereijido, 2009)

with minimal influence of testing effects, although the longitudinal analysis of these assessment

approaches is beyond the scope of the present paper.

Limitations

Critically, because the focus of the present paper is the valid evaluation of Spanish-

speaking ELs, caution is recommended in extending these results and recommendations to ELs

from diverse linguistic backgrounds (see Pitoniak et al., 2009). Although some similar patterns

of English grammatical development results have been observed among groups of bilingual

children who speak other languages, such as French and English (Paradis et al., 2003), it is

unlikely that this pattern is consistent across all English-learning bilingual children. Different

growth trajectories, and therefore assessment results, would be anticipated for children whose

native language does not include similar inflectional patterns to English (Blom, Paradis,

Sorenson Duncan, 2012; Paradis, 2016). Further, the small sample size and relatively young age

of the children limit generalization of these results to more diverse groups of ELs.

Additionally, the present paper focused primarily on two examples of approaches for

assessing the English grammatical development of ELs. Other standardized assessments, such as

those that do not include cloze tasks, are likely to yield different information than that which was

obtained from the BESA Cloze task (e.g., Ebert & Pham, 2017). Results from these tests may be

influenced by factors not addressed in this paper. These outside factors include translation effects, test sensitivity and specificity, and task effects (Abedi, 2011).

For naturalistic assessment, such as the narrative grammatical coding presented in the present paper, additional options exist as well. Clinicians may determine that specific error coding provides unnecessary detail for their specific situation, and may opt for global measures of grammatical development. More general language sampling, which can be obtained from a variety of contexts (e.g., free play, oral reporting, etc.), can also provide different insights into English grammar use for ELs at different points in their development (e.g., Cleave et al., 2010; Heilmann, 2010). Oral rubrics, asking questions, and written samples are alternative methods that may be beneficial for specific grammatical assessment needs (e.g., Pence, Justice, & Gosse, 2007; Gorman et al., 2016).

Conclusions

English grammar is a key component and indicator of language acquisition for Spanish-speaking ELs. In this review of two alternative approaches for evaluating the English grammatical development of Spanish-speaking ELs, the authors found that norm-referenced assessment yields noticeably different results than those obtained from the narrative-based approach to grammar assessment. Although norm-referenced assessment can be informative for identifying specific grammatical features as targets for therapy and for baseline evaluation of grammar, grammatical feature coding of a narrative language sample can be a quick tool useful for monitoring progress. With these options for assessment and informed expectations of EL grammatical development, clinicians can make appropriate decisions for supporting and identifying both typically- and atypically-developing Spanish-speaking ELs.

References

Abedi, J. (2011). Assessing English language learners: Critical issues. In M Basterra, E.

Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing*

linguistic and cultural diversity (pp. 49-71). New York, NY: Routledge.

American Speech-Language-Hearing Association. (2004). *Preferred practice patterns for the*

profession of speech-language pathology [Preferred Practice Patterns]. Retrieved from

www.asha.org/policy.

Armstrong, J. H. (2012). *Florida school health administrative guidelines*. Retrieved from

www.floridahealth.gov

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners:*

Report of the national literacy panel on language-minority children and youth. Mahwah,

NJ: Lawrence Erlbaum Associates.

Bedore, L. M., Peña, E. D., Gillam, R. B., & Ho, T. H. (2010). Language simple measures and

language ability in Spanish English bilingual kindergarteners. *Journal of Communication*

Disorders, 43(6), 498-510. doi: 10.1016/j.jcomdis.2010.05.002

Blom, E., Paradis, J., & Sorenson Duncan, T. (2012). Effects of input properties, vocabulary size

and L1 on the development of third person singular -s in child L2 English. *Language*

Learning, 62, 965-994. doi: 10.1111/j.1467-9922.2012.00715.x

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in

monolingual and dual language learner children with specific language impairment.

Journal of Communication Disorders, 43(6), 511-522. doi:

10.1016/j.jcomdis.2010.05.005

Ebert, K. D., & Pham, G. (2017). Synthesizing information from language samples and

    standardized tests in school-age bilingual assessment. *Language, Speech, and Hearing*

    Services in Schools, 48, 42-55. doi: 10.1044/2016_LSHSS-16-0007

Ehrler, D. J., & McGhee, M. L. (2008). *Primary Test of Nonverbal Intelligence* (PTONI).

    Austin: TX. Clocktower Hill Research and Publishing Group, LLC.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children. *Language,*

    Speech, and Hearing Services in Schools, 35, 155-168. doi: 10.1044/0161-

    1461(2004/016)

Goodwin, A. P., Huggins, A. C., Carlo, M. S., August, D., & Calderon, M. (2013). Minding

    morphology: How morphological awareness relates to reading for English language

    learners. *Reading and Writing*, *26*(9), 1387-1415. doi: 10.1007/s11145-012-9412-5

Gorman, B. K., Bingham, G. E., Fiestas, C. E., & Terry, N. P. (2016). Assessing the narrative

    abilities of Spanish-speaking preschool children: A Spanish adaptation of the narrative

    assessment protocol. *Early Childhood Research Quarterly, 36*, 307-317. doi:

    https://doi.org/10.1016/j.ecresq.2015.12.025

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual

    children. *Linguistics and Education, 13*, 175-197.

Gutiérrez-Clellen, V. F. & Simon-Cereijido, G. (2007). The discriminant accuracy of a

    gramatical measure with Latino English-speaking children. *Journal of Speech, Language,*

    and Hearing Research, 50, 968-981. doi: 1092-4388/07/5004-0968

Gutiérrez-Clellen, V. F. & Simon-Cereijido, G. (2009). Using language sampling in clinical

    assessments with bilingual children: Challenges and future directions. *Seminars in*

    Speech and Language, 30(4), 234-245. doi: 10.1055/s-0029-1241722

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with

      language impairment: A comparison with monolinguals and second language learners.

      Applied Psycholinguistics, 29(3), 3-19. doi: 10.1017/S0142716408080016

Heilmann, J. J. (2010). Myths and realities of language sample analysis. *SIG 1 Perspectives on*

      Language Learning and Education, 17, 4-8. doi: 10.1044/lle17.1.4

Heilmann, J., Miller, J. F., & Nockerts, A. (2010). Using language sample databases. *Language,*

      Speech, and Hearing Services in Schools, 41, 84-95. doi: 10.1044/0161-1461(2009/08-

      0075)

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative

      scoring scheme using narrative retells in young school-age children. *American Journal of*

      Speech-Language Pathology, 19(2), 154-166. doi: 10.1044/1058-0360(2009/08-0024)

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the

      transcript matter? *Language, Speech, and Hearing Services in Schools, 41*, 393-404. doi:

      10.1044/0161-1461(2009/09-0023)

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pellitier, J., & Girolametto, L.

      (2015). Narrative abilities in subgroups of English language learners and monolingual

      peers. *International Journal of Bilingualism, 19*(6), 677-696.

Ireland, M., & Conrad, B. J. (2016). Evaluation and eligibility for speech-language services in

      schools. *Perspectives of the ASHA Special Interest Groups, 1*(SIG 16), 78-90. doi:

      10.1044/persp1.SIG16.78

Jacobson, P. F., & Schwartz, R. G. (2005). English past tense use in bilingual children with

      language impairment. *American Journal of Speech Language Pathology, 14*(4), 313-323.

      Doi: 10.1044/1058-0360(2005/030)

Jacobson, P. F., & Walden, P. R. (2013). Lexical diversity and omission errors as predictors of

    language ability in the narratives of sequential Spanish-English bilinguals: A cross-

    language comparison. *American Journal of Speech-Language Pathology, 22*(3), 554-565.

    doi: 10.1044/1058-0360(2013/11-0055)

Kaderavek, J. N. (2015). *Language disorders in children: Fundamental concepts of assessment

    and intervention*. Boston, MA: Pearson.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., . . . Dunlop

    Velez, E. (2016). *The Condition of Education 2016* (NCES 2016-144). Washington, DC:

    U.S. Department of Education, National Center for Education Statistics.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence, and

    implications for clinical actions. *Journal of Communication Disorders, 43*(6), 456-473.

    doi: 10.1016/j.jcomdis.2010.02.002

Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally

    and linguistically diverse populations. *Language, Speech, and Hearing Services in

    Schools, 34*, 44-55. doi: 10.1044/0161-1461(2003/005)

Lucero, A. (2015). Cross-linguistic lexical, grammatical, and discourse performance on oral

    narrative retells among young Spanish speakers. *Child Development, 86*, 1419-1433. doi:

    10.1111/cdev.12387

Marchman, V. A., Martínez-Sussman, C., & Dale, P. F. (2004). The language-specific nature of

    grammatical development: Evidence from bilingual language learners. *Developmental

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

Mayer, M. (1969). *Frog, Where Are You?* New York, NY: Dial Press.

Miller, J. (1981). *Assessing language production in children: Experimental procedures*.

   Needham Heights, MA: Allyn & Bacon.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in

   morphemes. *Journal of Speech and Hearing Research, 24*, 154-161. doi:

   10.1044/jshr.2402.154

Miller, J., & Iglesias, A. (2015). *Systematic Analysis of Language Transcripts (SALT), Version

   16* [Computer software]. SALT Software, LLC.

Miller, L., Gillam, R. B., & Peña, E. D. (2001). *Dynamic assessment and intervention:

   Improving children's narrative skills*. Austin, TX: Pro-Ed.

Paradis, J. (2016). The development of English as a second language with and without specific

   language impairment: Clinical implications. *Journal of Speech, Language, and Hearing

   Research, 59*, 171-182. doi: 10.1044/2015_JSLHR-L-15-0008

Petersen, D. B., & Spencer, T. D. (2012). The narrative language measures: Tools for language

   screening, progress monitoring, and intervention planning. *SIG 1 Perspectives on

   Language Learning and Education, 19*, 119-129. doi: 10.1044/lle19.4.119

Pitoniak, M. J., Young, J. W., Martiniello, M., King, T. C., Buteaux, A., & Ginsburgh, M.

   (2009). *The guidelines for the assessment of English-language learners*. Princeton, NJ:

   Educational Testing Service.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2014)

   Bilingual English-Spanish Assessment. San Rafael, CA: AR-Clinical Publications.

Peña, E. D., & Bedore, L. M. (2008). Bilingualism and language impairment in children. In R.

   Schwartz (Ed.), *Handbook of child language disorders* (pp. 281-307). New York, NY:

   Psychology Press.

Pence, K., Justice, L. M., & Gosse, C. (2007). *Narrative Assessment Protocol*. Columbus, OH:

   Preschool Language and Literacy Lab, The Ohio State University.

Reardon, S. F., Kalogrides, D., Ho, A., Shear, B., Shores, K., & Fahle, E. (2016). *Stanford

   Education Data Archive* [Data file]. Retrieved from

   http://purl.stanford.edu/db586ns4974.

Restrepo, M. A., & Gutiérrez-Clellen, V. F. (2004). Grammatical impairments in

   Spanish/English-speaking children. In B. Goldstein (Ed.), *Language development: A

   focus on Spanish-English speakers* (pp. 213-256). Baltimore, MD: Brookes.

Simon-Cereijido, G., & Gutiérrez-Clellen,V. F. (2009). Spontaneous language markers of

   Spanish language impairment. *Applied Psycholinguistics, 28*(2), 317-339. doi:

   https://doi.org/10.1017/S0142716407070166

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B.

   (2014). Story retelling by bilingual children with language impairments and typically-

   developing controls. *International Journal of Language & Communication Disorders,

   49*(1), 60-74. doi: 10.1111/1460-6984.12044.

Zucker, S. (2004). *Administration practices for standardized assessments*. San Antonio, TX:

   Pearson Education.