



Butler University

Digital Commons @ Butler University

Undergraduate Honors Thesis Collection

Undergraduate Honors Thesis Collection

2020

English morpheme accuracy, diversity, and productivity measures in school-aged bilingual children

Gabrielle Raab
Butler University

Follow this and additional works at: <https://digitalcommons.butler.edu/ugtheses>



Part of the [Communication Sciences and Disorders Commons](#)

Recommended Citation

Raab, Gabrielle, "English morpheme accuracy, diversity, and productivity measures in school-aged bilingual children" (2020). *Undergraduate Honors Thesis Collection*. 524.
<https://digitalcommons.butler.edu/ugtheses/524>

This Thesis is brought to you for free and open access by the Undergraduate Honors Thesis Collection at Digital Commons @ Butler University. It has been accepted for inclusion in Undergraduate Honors Thesis Collection by an authorized administrator of Digital Commons @ Butler University. For more information, please contact digitalscholarship@butler.edu.

Running Head: ENGLISH MORPHEME MEASURES IN SCHOOL-AGED
BILINGUALS

**English morpheme accuracy, diversity, and productivity measures in school-aged
bilingual children**

A Thesis

Presented to the Department of Communication Sciences and Disorders

College of Communication

And

The Honors Program

of

Butler University

In Partial Fulfillment

Of the Requirements for Graduation Honors

Gabrielle Elizabeth Raab

May 9, 2020

Abstract

The identification of meaningful language measures for bilingual children has important clinical implications in the diagnosis of impairment. Given the shortage of bilingual speech-language pathologists, informative English measures are invaluable. This study extends current research to examine the utility of widely accepted English grammatical measures in bilingual school-aged children with differing levels of English exposure. The language sample analyses implemented have shown repeatedly to discriminate between typical and language impaired monolinguals, and more recently were considered in the assessment of developing bilinguals. Measures were taken from the language samples of thirteen school-aged bilingual Spanish-English speaking children who told a story using a wordless picture book. These thirteen participants were split into groups of high and low proficiency based on English preschool Language Assessment Scales (pre-LAS) scores and age of first exposure to English. There was no significant evidence found to support that these measures correlate with the broad measures of language (mean length of utterance and number of different words) or can differentiate between the two groups of varying proficiency.

Background

The number of multilingual and bilingual individuals in our society continues to grow rapidly and this raises concern that diagnosing impairment within these populations can be challenging, in particular in light of the shortage of bilingual speech-language pathologists in the United States. Assessing the morphosyntactic skills of bilingual children is made more difficult by the lack of clinical tools available for those of culturally and linguistically diverse backgrounds (Bedore & Peña, 2008). The benefit of using language samples in diagnosing language disorders is widely accepted throughout the field of speech and hearing. Dunn, Sliwinski, and Aram (1996) show the utility of using language samples over standardized tests. Both researchers and clinicians consider a language sample to be an essential component of a comprehensive assessment. Gutiérrez-Clellen, Restrepo, Bedore, and Anderson (2000) report that language sample analysis (LSA) is especially beneficial in cases of individuals of culturally and linguistically diverse backgrounds as standardized tests do not always lead to an accurate diagnosis because they are not standardized in a way that encompasses the language abilities of bilingual individuals. LSA allows clinicians to view a child's language as a whole, but in order to do so, discriminant measures must be created in order to distinguish between a language disorder and a language difference. These appropriate measures would allow LSA to be an appropriate and sensitive measure to assess a child's language skill.

In the present study, we explore the clinical utility of three measures of English morpheme use generated through spontaneous language samples of school-aged Spanish-English bilingual children with differing English proficiency levels. These investigations

are imperative as English is often the language used to assess the language skills of these bilingual children in the United States (Caesar & Kohler, 2007).

Broad Language Sample Measures

Broad language measures such as number of different words (NDW) and mean length of utterance (MLU) are used for both English and Spanish, but are considered clinically relevant and appropriate for use with clients of culturally or linguistically diverse backgrounds (Rojas & Iglesias, 2009). NDW reports the number of uninflected root words throughout the language sample and provides information regarding a child's lexical diversity (Golberg, Paradis, & Crago, 2008). For example, the words "run" and "running" would only count as one word despite the differing inflection. In contrast, "running" and "swimming" would account for two different words due to the use of two separate root words. Furthermore, MLU is a calculation of the number of morphemes or words from a sample and dividing it by the number of utterances produced. It is associated with morphosyntactic development, or the grammatical complexity of a child's oral language. MLU is generally calculated in morphemes for English, but in words for Spanish (MLUw) to ensure Spanish speakers are not given an unfair comparison due to the highly inflected nature of the language (Gutiérrez-Clellen et al., 2000). However, MLUw is the preferred measure for bilingual children because it is resistant to cross-linguistic differences in morpheme use. Both MLUw and NDW characterize a child's productive use of language. Bedore et al. (2010) studied MLUw and NDW in typically developing Spanish-English bilingual kindergarteners who demonstrated a mean MLUw of 5.04 and a mean NDW of 51.91 for English. Similarly, Simon-Cerejido & Gutiérrez-Clellen (2009) report a mean MLUw of 5.96 and a mean

NDW of 78.00 for this population as well. These measures may help identify impairment, but do not provide information regarding a child's morpheme use, a reliable indicator of language impairment (Bedore & Leonard, 1998).

Tense and Agreement Measures

Tense and agreement (T/A) morphemes include those that mark grammatical tense and those that alter words to make them agree with other components of the utterance. Difficulty with T/A is a distinctive feature of language impairment in English-speaking children (Leonard, 2014). Given this difficulty, the focus of LSA for English includes production of grammatical morphemes such as regular past tense -ed, present third person singular -s, the copula and auxiliary forms of is, are, am and the auxiliary *DO* (Simon-Cereijido, 2007). A problem arises in use of measures of accuracy of morpheme use in bilingual children because differing rates of morpheme marking are expected as compared to typically developing English monolinguals and in bilingual children with varying levels of English exposure and proficiency. Understanding morpheme marking in bilingual children, as well as ways to assess the use of different morphemes in these populations is vital in order to avoid mistakenly diagnosing impairment in children who are merely typically developing bilinguals (Paradis, 2005).

Hadley and Short (2005) examined the use of tense marker total and tense and agreement productivity score (TAP) in the language of children between 24 and 27 months of age with low language abilities and those at risk for language impairment. The former measures the diversity of the child's tense and agreement use, and the latter assesses the productive use of tense and agreement forms. However, both measures reflect contrastive uses of five different morpheme categories: (a) third-person singular (-

3s: jump/3s), (b) past tense (-ed: talk/ed), (c) forms of copula *BE* (cop *BE*: *He is fast*), (d) forms of auxiliary *BE* (aux *BE*: *She is going*), (e) forms of auxiliary *DO* (aux *DO*: *Does he like it?*). The tense marker total awards points for different surface forms for the five categories of interest: -3s, -ed, aux *DO* (*do, does, did*), cop *BE* (*is, am, are, was, were*), and aux *BE* (*is, am, are, was, were*). Therefore, a higher score indicates the ability of the child to use a higher number of different surface forms in each of the categories. The TAP score awards a speaker with points for each sufficiently different use for each of the T/A morphemes. Higher TAP scores indicate the speaker's ability to utilize these T/A morphemes in unique syntactic contexts. The researchers found that these measures were highly correlated with broad measures of language (MLU and NDW) in this population. Gladfelter and Leonard (2013) reported these measures, plus a composite measure of T/A accuracy, to differentiate typical from atypical language in monolingual English preschool-aged children.

Given that these had shown to be meaningful measures in the assessment of children developing language, Potapova, Kelly, Combiths, and Pruitt-Lord (2018) evaluated the measures in preschool age Spanish-English bilinguals. Their study included a group of bilinguals developing typically and bilingual children with low language skills as defined by parental concern regarding their child's language. The researchers found the measures to produce relevant information for this population, deeming them appropriate in the assessment of language skills in preschool aged bilingual children learning the English grammatical system. As anticipated, tense marker total and the productivity score converged with the broad measures of language. Additionally, the accuracy measure did not differentiate groups, but the diversity and productivity

measures were higher for the typical group than for the lower language group. Potapova et al. (2018) propose the lack of difference in the composite accuracy measure commonly used for clinical and research purposes may result from an overestimate of skill since the measure allows for repetitions and does not control for obligatory contexts. These are instances in which morphemes are required or obligatory in order to match that of a native-speaking adult's language. Contrastingly, the other measures were developed to account for these weaknesses, thereby showing utility with emerging T/A systems (Gladfelter & Leonard, 2013).

This study was motivated by the need to identify appropriate and meaningful measures of English language development in school-aged bilinguals. This study extends the current research to examine the grammatical measures of a finite verb composite score (FVM), tense marker total, and TAP score in bilingual school-aged children with differing levels of English exposure and proficiency levels. The identification of meaningful language measures for bilingual children has important clinical implications in the diagnosis of impairment.

1. Do morphological measures of accuracy, diversity and productivity relate to culturally sensitive broad measures of language sample analysis?
2. Do morphological measures of accuracy, diversity and productivity capture group differences?

Given that the measures of diversity and productivity were developed to assess early stages of English acquisition, I hypothesized that these measures would demonstrate a relationship with the broad language measures. Based on previous findings, I expected that the measure of accuracy would not relate to the broad measures. Additionally, I

hypothesized that the children with less English experience would not differ from those with more experience in the measure of accuracy, but would differ in the measures of diversity and productivity.

Method

Participants

Participants were recruited from schools and daycare centers primarily in the areas of Miami, Florida and Indianapolis, Indiana as part of a previous research project. The data set includes the language samples of bilingual Spanish-English school-aged children. The bilingual children were in English-only classrooms. Inclusion in the study required parental report via a detailed questionnaire of at least 30% Spanish input (Pearson, Fernandez, Lewedeg, & Oller, 1997). Additionally, parents indicated the age at which the child had begun to learn English. However, although parents indicated the amount of exposure at the time of testing, it is uncertain what exposure levels were like at the point of first exposure to English. Each child was administered the preschool Language Assessment Scales (pre-LAS) to estimate fluency in each language. To be included in the study, participants must have obtained a score of 4 or 5 in both English and Spanish. In total, 22 children met these criteria and were initially included in the present study. Age of first exposure to English and pre-LAS scores were utilized to divide children into higher and lower English proficiency groups. Next, the groups were age-matched by including in each group a child who is within at least four months of age of another child. The high proficiency group all had pre-LAS scores of 5 and had been exposed to English for a greater number of years. The final groups included six children in the high proficiency group ($M = 69.33$, $SD = 3.56$), and 7 children in the lower

proficiency group ($M = 66.57$, $SD = 6.21$). The groups were comparable in their amount of time in the United States as well as the amount of exposure to their native language. Therefore, it can be assumed that each group also had comparable amounts of exposure to English.

Each sample consists of a child telling a story based on a wordless picture book, with an adult conversation partner encouraging the child to move the story along. The children were randomly assigned one of two wordless picture books which they used to tell a story, *Frog Goes to Dinner* (Mayer, 1974) or *One Frog Too Many* (Mayer, 1975). Children were asked to tell the story while looking through the pictures. When necessary, the experimenter would cue them to continue the story or elaborate saying things such as, “Tell me about the man.” or “What happened next?”, but refrained from helping the child come up with words or aiding them in specific tasks throughout the language sample. This story-telling context encourages participants to produce utterances that are more complex than they may have been in conversation. With more complex utterances, it may be clearer if a child has difficulties with a specific feature of language.

Data Analysis

Each language sample was transcribed and coded using the *Systematic Analysis of Language Transcripts* (SALT; Miller & Chapman, 2002) software. Once coded for grammatical morphemes and errors in production, SALT produced the following measures: MLU in morphemes and words, NDW, as well as number of different verbs (NDV). Additional grammatical features that were examined include past tense –ed, present third person singular –s, auxiliary is, are, am, copula is, are, am, noun plural –s, and genitive –s. The following language measures were calculated using different

combinations of the data listed above: tense marker total, TAP score, and FVM. Tense marker totals and TAP scores were generated following the protocols outlined by Hadley and Short (2005), awarding points for contrastive uses of -3s, -ed, cop *BE*, aux *BE*, and aux *DO*. The tense marker total awards 1 point for each possible form of the five morpheme categories, for a maximum score of 15. The TAP score awards the participants up to 5 points for each sufficiently different use of each of the morphemes, for a maximum score of 25. Given the resulting small N, analyses were limited to descriptive statistics and t-tests.

Results

Means and standard deviations were calculated for each of the measures for both the high and low proficiency groups. MLUw for the high proficiency group ($M = 5.83$, $SD = .93$) and low group ($M = 6.19$, $SD = .55$) is consistent with previous research about typically developing Spanish-English bilingual children (Simon-Cerejido & Gutierrez-Clellen, 2009). Furthermore, NDW for the high proficiency group ($M = 95.17$, $SD = 5.60$) and the low proficiency group ($M = 97.14$, $SD = 15.60$) are similar to previously reported norms for this measure. An independent samples t-test showed that there was not a significant difference in the MLUw score for the high proficiency ($M = 5.83$, $SD = .93$) and low proficiency ($M = 6.19$, $SD = .55$) conditions; $t(11) = -.88$, $p = .40$. Additionally, scores demonstrated that both groups are performing at a level that is considered typical for this population (Bedore et al., 2011; Simon-Cerejido & Guitérrez-Clellen, 2009). Furthermore, there was not a significant difference in the NDW for the high proficiency ($M = 95.17$, $SD = 5.60$) and low proficiency ($M = 97.14$, $SD = 15.60$) conditions; $t(11) = -.29$, $p\text{-value} = .78$ showing that both groups are performing at a level consistent with

other typically developing bilingual children. A summary of the scores for each of the measures across both groups can be found in Table 1.

Table 1. Performance on measures for high proficiency and low proficiency groups

Group		MLUw	NDW	FVM	TAP	TenseMarkerTotal
High Proficiency (n = 6)	<i>M</i>	5.83	95.17	83.55	13.33	6.83
	<i>SD</i>	.93	5.60	28.28	1.97	1.72
Low Proficiency (n = 7)	<i>M</i>	6.19	97.14	70.83	12.57	6.14
	<i>SD</i>	.55	15.60	17.08	2.57	1.07

Note. MLUw = mean length of utterance in words; NDW = number of different words; FVM = finite verb morphology composite, TAP = T/A productivity score

Do morphological measures of accuracy, diversity and productivity relate to culturally sensitive broad measures of language sample analysis?

Pearson correlations were calculated to explore the relationship between MLUw and NDW with the other measures for both groups. No significant correlations were found between MLUw and the FVM, productivity score, and tense marker total for the high and low proficiency groups. Furthermore, there was no evidence of correlation between NDW and the T/A morpheme measures. Table 2 summarizes the results found in these analyses.

Table 2. Summary of Pearson Correlation p-values

Group		FVM	TAP	TenseMarkerTotal
High Proficiency	MLU _w	.47	.74	.46
	NDW	.63	.67	.76
Low Proficiency	MLU _w	.17	.85	.78
	NDW	.63	.25	.97

*. Correlation is significant at the 0.05 level

Note. MLU_w = mean length of utterance in words; NDW = number of different words; FVM = finite verb morphology composite, TAP = T/A productivity score

Do morphological measures of accuracy, diversity and productivity capture group differences?

In order to address my second research question, independent samples t-tests were conducted for all of the T/A measures (FVM, TAP score, and Tense Marker Total) between the high proficiency and low proficiency groups. There was no significant difference found between the groups for FVM (high proficiency: $M = 83.55$, $SD = 28.28$; low proficiency: $M = 70.83$, $SD = 17.08$), conditions; $t(11) = .10$, p -value = .34. The TAP scores were also comparable across groups (high proficiency: $M = 13.33$, $SD = 1.97$; low proficiency: $M = 12.57$, $SD = 2.57$) conditions; $t(11) = .59$, p -value = .57. Additionally, the groups did not differ for Tense Marker Total (high proficiency: $M = 6.83$, $SD = 1.72$; low proficiency: $M = 6.14$, $SD = 1.07$) conditions; $t(11) = .88$, p -value = .40. Given the lack of significant difference between groups, the results here provide no evidence that these measures are able to differentiate between the of higher and lower proficiency groups.

Discussion

The present study sought to extend the current literature to address the use of measures of accuracy, diversity, and productivity in school-aged Spanish-English bilingual children. Both MLUw and NDW have been shown to be culturally sensitive measures when examining a bilingual's language skills as they are able to characterize the productive use of language. Furthermore, both measures have been found to be useful in identifying language impairment in both monolingual and bilingual children (Bedore et al., 2010; Simon-Cereijido & Gutiérrez-Clellen, 2009). I had two primary hypotheses for this study. First, I hypothesized that the FVM, productivity score, and tense marker total would demonstrate a relationship between the broad measures, MLUw and NDW. Second, I hypothesized that the measures of diversity and productivity would show differences between those with differing amounts of exposure to English.

A relationship between the broad measures of language and the measures of accuracy, diversity, and productivity was not revealed in the data. This contrasts with the findings of Potapova et al. (2018) who found that measures of diversity and productivity were correlated with MLUw and NDW in preschool Spanish-English bilinguals. Additionally, the groups in the current study demonstrated comparable performance across each of the measures of interest and did not differentiate the two groups as expected. Potapova et al. (2018) found that these measures captured differences between the typically developing and low language groups in their study. However, no information regarding the children's exposure to either of the languages was included. It is possible that the nature of the children's bilingualism could differ from those included in the present study. For example, it is unknown whether the children were simultaneous

bilinguals, where both languages were present in the child's environment since birth, or if they were only recently exposed to English at the time of testing. As a result, the role of type of bilingualism is uncertain.

The lack of significance in the current study may also be explained by the limitations of the measures themselves. Though the tense and agreement measures have been useful in assessing the grammatical morphology use of children whose skills are emerging, these are highly impacted by the obligatory contexts, or instances requiring the use of these grammatical forms (Souto, Leonard & Deevy, 2013). For example, the utterance, "The frog is happy," requires the use of copula *is* in order for the sentence to be grammatical. Some morphemes may have more obligatory contexts than others, causing them to weigh heavier on the score. For example, in the FVM score, the number of T/A morphemes accurately produced is divided by the total number of obligatory contexts for the morphemes of interest. Morphemes with a higher number of obligatory contexts (e.g. copula *is*) will have more influence on the child's overall score than those that have fewer obligatory contexts (e.g. past tense *-ed*).

Hadley and colleagues' (Hadley & Holt, 2006; Hadley & Short, 2005) measure of productivity focuses on the diversity of forms used, and requires a minimum number of five unique instances of use of each morpheme of focus in order to obtain the highest score possible. As a result, this measure is similarly impacted by the number of opportunities to produce grammatical forms with the added requirement of different subjects for each instance. For example, the third person singular and auxiliary *DO* morphemes have the lowest number of different productions among both groups' language samples in this study. There were few, if any productions of these morphemes

of interest demonstrating that the children in this study did not produce sufficient obligatory contexts in the sample to obtain the points for these forms toward the maximum 25.

A final explanation for the findings of this study may lie in the instruments used to elicit the spontaneous language samples. Although Mercer Mayer's (Mayer, 1974; Mayer, 1975) wordless picture books are frequently used to elicit language samples and are even endorsed by the makers of SALT software for use with monolingual English speakers as well as bilingual Spanish-English speaking children, it is possible the two stories utilized did not provide sufficient opportunities for obligatory contexts for the morphemes of interest for this population.

In summary, although previous research demonstrated the utility of T/A measures in assessing the language skills of preschool bilinguals as well as diagnosing SLI in monolingual children, the results of this study did not reveal a correlation between the measures of accuracy, diversity, and productivity and the broad measures of language in school-age Spanish-English bilinguals. Furthermore, the results provide no evidence that the measures of interest differentiate between groups of higher and lower proficiency since members of each group demonstrated comparable performance across each of the measures. Future studies of this type with this population might consider language sample elicitation procedures that lend themselves to increased obligatory contexts for the morphemes of interest, possibly to include a different type such as play. Additionally, of course, future research could be strengthened by utilizing a larger sample size.

References

- Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*(5), 1185–1192.
- Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*(1), 1–29.
- Bedore, L. M., Peña, E. D., Gillam, R. B., & Ho, T. H. (2010). Language sample measures and language ability in Spanish–English bilingual kindergarteners. *Journal of Communication Disorders, 43*(6), 498–510.
- Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools, 38*(3), 190–200.
- Dunn M., Flax J., Sliwinski M., & Aram D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research, 39*, 643–654.
- Gladfelter A., & Leonard L. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. *Journal of Speech, Language, and Hearing Research, 56*, 542–552.
- Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29*(1), 41–65.

- Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*, 88-98
- Gutiérrez-Clellen, V. F., & Simon-Cerejido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language, 30*, 234–245.
- Hadley, P. A., & Holt, J. K. (2006). Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research, 39*, 984–1000.
- Hadley P. A., & Short H. (2005). The onset of tense marking in children at-risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 1344–1362.
- Mayer, M. Frog goes to dinner. New York, NY: Penguin books; 1974.
- Mayer, M. One frog too many. New York, NY: Penguin Books; 1975.
- Paradis, J. (2005). Grammatical morphology in children learning English as a second language: Implications of similarities with specific language impairment. *Language, Speech, and Hearing Services in Schools, 36*, 182-187.
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18(1)*, 41–58.
- Potapova I., Kelly S., Combiths P. N., & Pruitt-Lord S. L. (2018). Evaluating English

morpheme accuracy, diversity, and productivity measures in language samples of developing bilinguals. *Language, Speech, and Hearing Services in Schools*, 49, 260–276.

Rojas, R., & Iglesias, A. (2009). Making a case for language sampling: Assessment and intervention with (Spanish-English) second language learners. *The ASHA Leader*, 14(3), 10–13.

Simon-Cerejido G., Gutiérrez-Clellen V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics*, 28, 317-339.

Souto S. M., Leonard L. B., & Deevy P. (2014). Identifying risk for specific language impairment with narrow and global measures of grammar. *Clinical Linguistics & Phonetics*, 28, 741-756