HHS Public Access

Author manuscript

Rev Logop Foniatr Audiol. Author manuscript; available in PMC 2023 May 23.

Published in final edited form as:

Rev Logop Foniatr Audiol. 2021 ; 41(4): 164–171. doi:10.1016/j.rlfa.2021.02.002.

Elicited vs. Spontaneous Language as Methods for the Assessment of Grammatical Development: The DEME Assessment Tool

Anny Castilla-Earls,

Department of Communication Sciences and Disorders, 4455 Cullen Blvd, Room 100, University of Houston, Houston, TX 77204

Ana Teresa Pérez-Leroux,

Department of Linguistics and Spanish, University of Toronto, Toronto, Canada

Alejandra Auza

Hospital General Dr. Manuel Gea González, Mexico City, Distrito Federal, Mexico

Abstract

Introduction and Objectives.—This study compares data collection approaches in the assessment of grammatical development in Spanish-speaking children. Specifically, we compared error rates produced in data collected using samples from spontaneous language versus elicited production, using both broad (overall) and narrow measures (errors with noun phrases).

Methods and Participants.—Monolingual-Spanish-speaking five-year-olds (n=55) were divided into typical language development (TL) and at-risk (Risk) according to a preexisting test, *Tamiz de Problemas del Lenguaje*. All children completed an elicited production and a narrative task.

Results.—Children in the TL group outperform children in the Risk group in all measures used in this study. Statistically significant differences were found between children at Risk and TL children in both spontaneous and elicited language measures, although the effect size of the elicited language measures was considerably higher. Elicited and spontaneous tasks are more likely to produce results that are in accord than in disaccord. However, when results are in disaccord, the results almost always show low performance in elicited language but high performance in spontaneous language. Elicitation methods do not seem to have an impact on the type of error produced for neither narrow nor broad measures

Keywords

assessment; spontaneous language; elicited language; grammar

Children with Developmental Language Disorders (DLD), also known as Specific Language Impairment, have difficulties learning language in comparison to their same-age peers

Corresponding Author annycastilla@uh.edu 713-743-0488.

No conflict of interest to report

(Bishop et al., 2016; Leonard, 2019). Although difficulties in the areas of phonology, semantics, and pragmatics are all part of the disorder, the main characteristic is that children with DLD make significantly more grammatical errors than their peers without DLD. Therefore, measures of grammaticality are often included in language assessment practices for the identification of DLD. Our goal is to compare elicited language and spontaneous language approaches to collecting language samples.

Analysis of spontaneous language, which allows a naturalistic observation of the representative language skills of a child, is one of the standard strategies for the assessment of child language (Gutiérrez-Clellen et al., 2000; Restrepo, 1998; Rojas & Iglesias, 2013). Narrative samples are deemed particularly useful, as they can be used for the assessment of narrative and grammatical skills, and because they elicit more complex and grammatically diverse language compared to interviews or conversation (Wagner et al., 2000). Story-retelling tasks yield greater productivity in terms of sentence complexity, length, and completeness than other forms of sampling (Gazella & Stockman, 2003), a possible effect of the absence of a shared common ground (Masterson & Kamhi, 1992). Measures of language productivity such as mean length of utterance (MLU), subordination index (number of dependent clauses per sentence) and percentage of grammatical errors (PGU; Restrepo, 1998) or number of errors per utterance (Simon-Cereijido & Gutiérrez-Clellen, 2007) can be derived from narratives (Castilla-Earls & Eriks-Brophy, 2012; Restrepo et al., 2010). These measures are often used to examine language growth in children (Rojas & Iglesias, 2013; Castilla-Earls & Brophy, 2012) and/or to identify children with DLD (Restrepo, 1998; Simon-Cereijido & Gutiérrez-Clellen, 2007).

The second main approach to assess the language of Spanish-speaking children relies on elicited language (Bedore & Leonard, 2001; Castilla-Earls & Eriks-Brophy, 2012; Morgan et al., 2013). Used as an assessment strategy, this approach focuses on creating felicitous (and/or obligatory) contexts for the production of constructions where errors are predictable. Most language assessment batteries include elicited language to control for the number of opportunities a child has to produce a given response, as some structures might be avoided or not used in a narrative (e.g., the subjunctive mood or clitic pronouns). In this study, we employ the *Desarrollo Morfologico del Español* (Castilla, 2008; Castilla-Earls et al., 2020), which is a task developed to elicit grammatical forms in Spanish, to investigate elicited language.

It is important to note that accuracy rates in elicited and spontaneous language seldom align in language development. Naturalistic language samples show far lower rates of ungrammatical productions relative to elicited samples (See Castilla-Earls et al., 2020 for a review). Errors that appear reliably under experimental paradigms might be vanishingly rare in spontaneous speech. At times these asymmetries have methodological sources. From one perspective, spontaneous speech might hide errors that are difficult to detect, or fail to elicit more difficult structures. For example, coding for missing direct objects--a clinically useful property-- is notoriously unreliable in spontaneous language ((Pirvulescu, 2006)). From another, more general perspective, children’s spontaneous production seems more grammatically conservative (i.e., faithful to the ambient language), showing few commission (as opposed to omission) errors. According to Snyder (2007), under the demands of

elicitation tasks children might venture beyond their established abilities; thus, in some instances, erroneous performance may be interpreted as task effects. Beyond issues of the theoretical interpretation of errors, these discrepancies raise questions about the congruence of the two approaches to sampling language. From an assessment perspective, what matters is whether one form of measurement offers better differentiation between typical and atypical populations.

There is an additional dimension to consider. While some approaches focus on broad measures of grammaticality, such as the percentage of grammatical utterances in a sample (Simon-Cereijido & Gutierrez-Clellen, 2007) or the number of grammatical errors per utterance in a sample (Restrepo, 1998),[1] other approaches focus on more narrow measures, scrutinizing a single property or set of properties (Gusewski & Rojas, 2017). Our goal is to explore these potential differences (spontaneous vs. elicited, and broad vs. narrow) in language assessment. Article production, a narrow measure of grammaticality, provides a perfect scenario to test the difference in elicitation approaches because articles are frequent and unavoidable in spontaneous language. Articles are also well-established clinical markers of DLD in Spanish (Anderson & Souto, 2005; Restrepo & Gutierrez-Clellen, 2001). For example, monolingual Spanish-speaking 5–6-year-olds in Morgan et al. (2013) have substantive differences in article accuracy according to clinical status (81% TD, 54% DLD). Developmentally, article acquisition has different timelines depending on whether the data is elicited or spontaneous. For monolingual children, articles emerge before the second birthday, when nominal phrases emerge, but production is not initially accurate. In spontaneous speech, articles first appear as an undifferentiated vowel (proto-article) segment that precedes the noun phrase, which eventually morphs into a full article paradigm and morphological accuracy (Lopez-Ornat 1997). By the second birthday, omission rates drop sharply (Serra et al. 2000, Pérez-Leroux and Battersby, 2009) and gender accuracy in articles is at ceiling. For instance, Snyder et al. (2005) find 97% accuracy on data extracted from the CHILDES database. In elicited language studies, article errors persist until school age on article choice (Solé, 1984; Castilla, 2008), in number marking (Miller & Schmitt, 2010), and article omission (Castilla, 2008). For monolingual children, Perez-Leroux, Castilla-Earls, and Brunner (2012) report target production of articles developing from 50% to 75% between the ages of three and five. By the age of six, both bilingual and monolingual children produce minimal article errors in elicited tasks, mostly of omission (Castilla-Earls et al., 2019). Relevant to our purpose, Castilla-Earls et al.'s (2020) comparison of effects sizes for article use in children with and without DLD showed that the effect size of the difference for articles was greater for elicited production than for spontaneous language.

The present study compares a group of children at risk for DLD and a group of children with typical development. We investigate whether data from monolingual Spanish-speaking children reveal differences between the two approaches (elicited versus spontaneous), when using broad (overall grammatical accuracy) and narrow measures (article accuracy)? Is the

[1]There are differences between these broad measures. The number of grammatical error per utterance ris significantly higher than the percentage of ungrammatical utterances because it takes into account all errors present in an utterance, but both indices show similar trends in development (Castilla-Earls & Brophy, 2012).

distribution of errors observed comparable for children with typical language skills and those at risk of language disorders?

Methods

Participants

The participants in this study included monolingual Spanish-speaking children with typical language skills (TL; *n*=26) and at risk of developmental language disorders (Risk; *n*=29). There were 32 boys (Risk=19, TL= 13) and 23 girls (Risk=10, TL= 13). The average age for all children was 66 months (*SD*= 10 months) with no statistically significant differences between the TL and Risk groups (Age TL= 66 months, *SD*= 10 months; Age Risk= 66 months, *SD*= 10 months). All children passed an otoacoustic emission test and obtained a score of ≥70 on the Non-Verbal Scale of the KBIT-2 (Kaufman & Kaufman, 2014). Children were considered to be at risk of developmental language disorder if they failed the Tamiz de Problemas del Lenguaje (Auza, Marquez, Murata, & Morgan, 2018) using the cut-off scores by age groups provided in the testing manual. The TPL has acceptable diagnostic accuracy with at least 80% sensitivity and specificity for the age groups included in this study (Auza, Marquez, Murata, & Morgan, 2018; Plante & Vance, 1994).

Measures

Spontaneous language samples.—Three measures were derived from spontaneous language samples: Mean length of utterance in words (MLUw), Percentage of Grammatical Utterances (PGU), which is a broad measure of grammaticality, and Percentage of Correct Articles, which is the narrow measure of grammaticality used in this study. To calculate MLUw, all samples were segmented into communication units following the protocol from the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2018). PGU was calculated as the number of utterances with a grammatical error over all utterances in the analysis (e.g., errors of omission, errors with any grammatical structures, over regularization, etc.). To calculate the percentage of correct articles in the narrative, all instances with an article in the language samples were coded as either correct or incorrect (e.g., omission of articles, gender substitution, number substitutions). The percentage of correct articles in the narrative was calculated as the number of correct uses of articles in the sample over all uses (correct and incorrect) of articles. Article errors were classified either as omission or as substitution errors.

***Desarrollo Morfológico de Español* (DEME).—**The DEME was initially developed to test the production of articles, clitics, adjectives, and plurals in Spanish speaking children (Castilla, 2008). This task was used to examine language development in Spanish-speaking children ages 3 to 5 in Colombia. Using the DEME, Castilla (2008) showed developmental changes in the production of clitics, articles, adjectives and plurals between the ages of 3 and 4, but no statistically significant changes between 4 and 5, although these children were not performing as adults yet. With funding from the National Institutes of Deafness and Other Communication Disorders, the DEME was expanded to include the evaluation of verbs and the subjunctive mood. Castilla-Earls, Auza, Pérez-Leroux, Fulcher-Rood, and Barr (2020) examined the diagnostic accuracy of the DEME to identify children in Mexico with

Developmental Language Disorders. They found statistically significant differences between children with and without developmental language disorders for the production of articles, adjectives, clitics, verbs, and the subjunctive, but not for plurals. In addition, they found that a combination of verbs and clitics provided the best diagnostic potential between the two groups.

The DEME is a task designed to elicit specific grammatical structures known to be problematic for children with DLD: articles, direct object pronouns, adjectives, verbs, and the subjunctive mood. The test consists of 50 test items that provide a raw score between 0 and 50 that represent the number of correct items. The test consists of pictures accompanied by either a question or a sentence completion activity to elicit the grammatical structure of interest. There are 10 items for direct object pronouns, 10 items for articles, 10 items for adjectives, 10 items for verbs, 5 items for the subjunctive, and 5 items for plurals. It takes approximately 12 minutes to administer the task to a child. For the purpose of this study, we report raw scores for the DEME Total and the Article items only. We also presented information on substitution and omission errors of articles from the DEME.

Procedures—All children were tested individually. Children participated in two testing sessions of about 45 minutes each that included a variety of language measures. In the second session, children completed both the DEME and the language sample task. Children produced two stories: a story retell and a story generation. The story retell was always administered first. For the story retell, children heard a story script while looking at the pictures of one of two books (Frog Goes to Dinner or Frog on His Own). The story script used for the story retelling task can be found on the SALT website. Script selection was done randomly. After children listened to the story while looking at the pictures, children were asked to retell the story. After the story retell task was completed, children were given the second frog storybook and asked to tell a story while looking at the pictures. In this study, we use the combination of the two stories.

All stories were transcribed and segmented using SALT protocols by research assistants. The samples were coded for errors by native Spanish-speaking research assistants. The first author in this study reviewed all coding and disagreements were discussed and solved between the research assistant and the first author of the study. For the scoring of the DEME, a 96% coding inter-rater reliability was established using 20% of the sample.

Results

We first examine differences between TL children and children at risk of DLD in both spontaneous and elicited measures, using *t* tests. Descriptive statistics and the results of the group comparisons are included in Table 1. Regarding spontaneous language measures, there were no statistically significant differences between the number of utterances produced by children in the TL group and Risk group, but children in the TL group produced utterances that were longer on average than the children in the Risk group. Children in TL group also produced more articles on average and a higher proportion of these articles were correct in comparison with children in the Risk Group. For elicited language measures, we observed

that TL children outperformed children in the Risk group in both the total score of the DEME and the percentage of correct DEME articles.

We then examined the correlation between spontaneous and elicited measures. The correlation between the total score of the DEME and PGU was highly significant (r= .673, p<.001) as it was the correlation between the correct percentage of DEME articles and the correct percentage of articles in the spontaneous sample (r=.501, p<001). These correlations suggest a linear relationship between the spontaneous and elicited variables with moderate effect sizes. Figure 1 and Figure 2 show the details of this relationship.

Next, we compared the number of children who made errors in the language samples versus those who made errors in spontaneous language for our narrow and broad measures. We started first with an individual analysis of broad errors. Table 3 shows the distribution of children by group and by type of language elicitation. None of the children in this study produced 100% accurate responses in either the spontaneous language sample or the DEME. We used a cut-off of 80% to divide children into high accuracy and low accuracy inspired by the 80% cut-off proposed by Restrepo (1998) for identification of language disorders. Using this approach, we find substantive overall agreement in total, with 40 children who had results of the elicited and language sample which were in agreement: i) 11 children from the TL group were high performers in both tasks, and two in the Risk group; and ii) 23 children in the Risk group were low performers in both spontaneous and elicited tasks compared to 5 in the TL group. For a quarter of the children (n=14), the results of the two approaches were in disagreement. Importantly, there were fewer children in the Risk group (n=5) with disagreeing results (e.g. high performance in one task but low performance in the other task) than in the TL group (n=10).

Moving on now to the narrow measures, we found 11 children in the TL group who made no errors of articles in either the language sample or the DEME; none of the at-risk group fell in this category. The distribution of children who made errors in articles is presented in Table 4. A child was considered to have high performance in articles if they were at least 90% accurate. Seven children had low performance in both spontaneous language and the DEME and eight children had high performance in the Risk group. Similarly, 19 children in the TL group showed high performance in both tasks, indicating an accord between their results. In total, the results of the DEME and the spontaneous language were in agreement for 62% of the children. Disagreeing results were found for seven children in the TL group and 14 children in the Risk group. For these children, the performance in the DEME was low while the performance in the sample was high, which amounts to 38% of the sample with disagreement in their results. To further explore the type of errors made by children, we compared the percentage of omission and substitution errors in both types of elicitation techniques. Using a Two-Way Repeated measure ANOVA design, we explored the effect or error (omission or substitution) and sampling context (spontaneous or elicited) with Group (TL or Risk) as a between-subject variable. Descriptive statistics are presented in Table 5. Our results revealed a significant main effect for sampling (F=8.671, p=0.005), a significant effect for group (F=12.035, p=001), but a non-significant effect of error type (F=.543, p=465). None of the interactions were statistically significant. These results fit in with standard predictions: the DEME elicited a higher percentage of errors than the spontaneous

sample and children at Risk made significantly more errors than children in the TL group in both elicited and spontaneous sampling techniques.

Discussion

The purpose of this study was to examine the language data collected using two main approaches commonly used in child language assessment: Spontaneous and elicited language tasks. We used both broad and narrow measures of grammaticality to explore how children with typical language skills and children at risk of language disorders perform on spontaneous and elicited measures.

Our results suggested that children with typical language skills outperformed children at Risk in both broad and narrow spontaneous and elicited measures. The language sample collected included a comparable number of utterances between the groups, but children in the Risk group produced utterances that were shorter on average and produced fewer articles spontaneously in general. Children produced lower accuracy rates in elicited language in comparison to spontaneous language for both broad and narrow measures. The largest effect size for group differences was found for the DEME total score and PGU, which are both broad measures of grammaticality. These results are in agreement with those of Castilla-Earls and colleagues (2015, 2018, 2019), Restrepo (1998), Simon-Cereijido, and Gutierrez-Clellen (2007), who have documented significant differences between typically developing children and children with DLD for both spontaneous and elicited measures. The results are also in alignment with Castilla-Earls et al., (2020) who reported higher effect sizes for elicited measures for articles in their review of the literature. These findings are novel because, to our knowledge, it is the first time that the group difference between TL children and children at risk are simultaneously reported for spontaneous and elicited measures.

Our results also showed moderate correlations between elicited and spontaneous measures for all children in the study. This finding suggests that although elicited and spontaneous measures are positively associated (i.e., they go in the same direction), this association is moderate at best. The association between spontaneous and elicited measures was stronger for broad measures of grammaticality (see Figure 1) than for narrow measures of grammaticality (See Figure 2). This difference might be explained by ceiling effects for articles, while broad measures allowed for more variability in performance. This highlight the relevance of using broad measures that target a variety of grammatical structures, such as the DEME. It is also important to note there is variability of the grammatical errors in children with DLD, so broad measures are better to identify children with DLD and to select grammatical strctures for remediation (Morgan et al., 2009).

Regarding the distribution of children who made errors, we found that both broad and narrow measures were more likely to yield concord than discord between the two sampling approaches. We also found differences in broad and narrow measures. Disagreements between the outcomes of the narrative sample analyses and the elicitation test results were found for 25% of children using broad measures, and 38% of children when the grammatical assessment was narrowly focused on errors with articles. These results suggest

that an examiner is more likely to find children with similar results between spontaneous and elicited measures, but that the possibility remains of finding children with disagreeing results for both TL children and children at risk because of the well-known heterogeneity found in the language manifestations of children with language disorders. However, the disagreement in the results for both groups of children are such that performance on the language sample might be high while performance in the DEME, our elicited task, is low.

Last, in our examination of elicitation methods on the type of errors produced (omissions versus substitutions), we found no evidence of an effect of type of error. Children produced similar error types during the elicitation and spontaneous language sample task. Nonetheless, more errors were attested during the elicited language given that, in a well-designed elicitation task, the items target the most vulnerable structures in the language. These findings suggest that elicitation methods do not have an impact on the type of error produced, independent of the type of measure (broad or narrow).

The current study has two limitations that are important to acknowledge. First, the score used to separate the children between high and low percentage of errors was arbitrary. The results of this study are interpreted using an 80% cut-off scores, but different cut-off scores might yield different results. Second, the sample size of the study is relatively small. It is possible that higher samples might reveal more variability in performance. Future studies should increase the sample size to examine the difference between elicited and spontaneous language samples.

Conclusion

Our comparison of the two approaches to sampling show results that are in agreement with previous studies, with at-risk children producing more errors and typical children producing less, and elicitation showing more errors than spontaneous language (Bedore & Leonard, 2001; A. Castilla-Earls et al., 2020). The magnitude of the discrepancy between our two approaches is substantive, so we conclude that researchers should favor elicited language approaches when the goal is to find differences between the groups. Although most of the children showed results that were in agreement between elicited and spontaneous language, a substantive number of children showed results in disagreement. When errors were in disagreement, we found low performance in elicited language and high performance in spontaneous language. The results of this study suggest that elicited and spontaneous measures do not completely overlap and might provide different information about the language skills of a child, particularly when narrow measures of grammaticality are considered.

References

Auza A, Murata M, Márquez ME, & Morgan GP (2018). Tamiz para detectar problemas del lenguaje (TPL). Mexico: Manual Moderno.

Bedore LM, & Leonard LB (2001). Grammatical Morphology Deficits in Spanish-Speaking Children with Specific Language Impairment. Journal of Speech, Language, and Hearing Research, 44(1–4), 905–924. 10.1044/1092-4388(2001/072)

Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T, Adams C, Archibald L, Baird G, Bauer A, Bellair J, Boyle C, Brownlie E, Carter G, Clark B, Clegg J, Cohen N, Conti-Ramsden G, Dockrell

J, Dunn J, Ebbels S, … Whitehouse A. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. PLoS ONE, 11(7), 1–27. 10.1371/journal.pone.0158753

Bishop, Dorothy VM, Snowling MJ, Thompson PA, Greenhalgh T, Adams C, Archibald L, Baird G, Bauer A, Bellair J, Boyle C, Brownlie E, Carter G, Clark B, Clegg J, Cohen N, Conti-Ramsden G, Dockrell J, Dunn J, Ebbels S, … house A. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. Journal of Child Psychology and Psychiatry and Allied Disciplines, 58(10), 1068–1080. 10.1111/jcpp.12721 [PubMed: 28369935]

Castilla-Earls A, Auza A, Pérez-Leroux AT, Fulcher-Rood K, & Barr C. (2020). Morphological Errors in Monolingual Spanish-Speaking Children With and Without Developmental Language Disorders. Language, Speech, and Hearing Services in Schools, 51(2), 270–281. 10.1044/2019_LSHSS-19-00022 [PubMed: 32255752]

Castilla-Earls AP, & Eriks-Brophy A. (2012). Spontaneous language measures in monolingual Spanish-speaking preschool children. Revista de Logopedia, Foniatria y Audiologia, 32(3). 10.1016/j.rlfa.2012.04.002

Fiestas CE, & Peña ED (2004). Narrative Discourse in Bilingual Children. Language, Speech, and Hearing Services in Schools, 35(2), 155–168. 10.1044/0161-1461(2004/016)

Gazella J, & Stockman IJ (2003). Children's Story Retelling Under Different Modality and Task Conditions. American Journal of Speech-Language Pathology, 12(1), 61–72. 10.1044/1058-0360(2003/053) [PubMed: 12680814]

Gutiérrez-Clellen VF, Restrepo MA, Bedore L, Peña E, & Anderson R. (2000). Language Sample Analysis in Spanish-Speaking Children. Language, Speech, and Hearing Services in Schools, 31(1), 88–98. 10.1044/0161-1461.3101.88 [PubMed: 27764338]

Kapantzoglou M, Fergadiotis G, & Restrepo MA (2017). Language Sample Analysis and Elicitation Technique Effects in Bilingual Children With and Without Language Impairment. Journal of Speech, Language, and Hearing Research, 60(10), 2852–2864. 10.1044/2017_JSLHR-L-16-0335

Kaufman AS, & Kaufman NL (2014). Kaufman Brief Intelligence Test, Second Edition. In Encyclopedia of Special Education. 10.1002/9781118660584.ese1325

Leonard LB (2019). SLI across Languages. In Children with Specific Language Impairment. 10.7551/mitpress/9152.003.0007

Masterson JJ, & Kamhi AG (1992). Linguistic Trade-Offs in School-Age Children With and Without Language Disorders. Journal of Speech, Language, and Hearing Research, 35(5), 1064–1075. 10.1044/jshr.3505.1064

Morgan GP, Restrepo MA, & Auza A. (2013). Comparison of Spanish morphology in monolingual and Spanish-English bilingual children with and without language impairment. Bilingualism, 16(3), 578–596. 10.1017/S1366728912000697

Pérez-Leroux AT, Castilla-Earls AP, & Brunner J. (2012). General and specific effects of lexicon in grammar: Determiner and object pronoun omissions in child Spanish. Journal of Speech, Language, and Hearing Research, 55(2). 10.1044/1092-4388(2011/10-0004)

Pirvulescu M. (2006). Theoretical implications of object clitic omission in early French: Spontaneous vs. elicited production. Catalan Journal of Linguistics, 5, 221–236. 10.5565/rev/catjl.70

Plante E, & Vance R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. Language, Speech, and Hearing Services in Schools. 10.1044/0161-1461.2501.15

Restrepo María Adelaida. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. Journal of Speech, Language, and Hearing Research, 41(6), 1398–1411. 10.1044/jslhr.4106.1398

Restrepo Maria Adelaida, Castilla AP, Neuharth-pritchett S, Hamilton CE, & Arboleda A. (2010). Effects of a supplemental Spanish oral language program on sentnece lenght, complexity, and gramaticality in Spanish-speaking children attending English-only preschools. Language, Speech, and Hearing Services in Schools, 41(January), 3–13. [PubMed: 19901038]

Reuterskiöld Wagner C, Nettelbladt U, Sahlén B, & Nilholm C. (2000). Conversation versus narration in pre-school children with language impairment. International Journal of Language and Communication Disorders, 35(1), 83–93. 10.1080/136828200247269 [PubMed: 10824226]

Rojas R, & Iglesias A. (2013). The Language Growth of Spanish-Speaking English Language Learners. In Child Development (Vol. 84, Issue 2, pp. 630–646). WileySociety for Research in Child Development. 10.1111/j.1467-8624.2012.01871.x [PubMed: 23075314]

Simon-Cereijido G, & Gutiérrez-Clellen VF (2007). Spontaneous language markers of Spanish language impairment. Applied Psycholinguistics, 28(2), 317–339. https://doi.org/10.1017.S0142716407070166

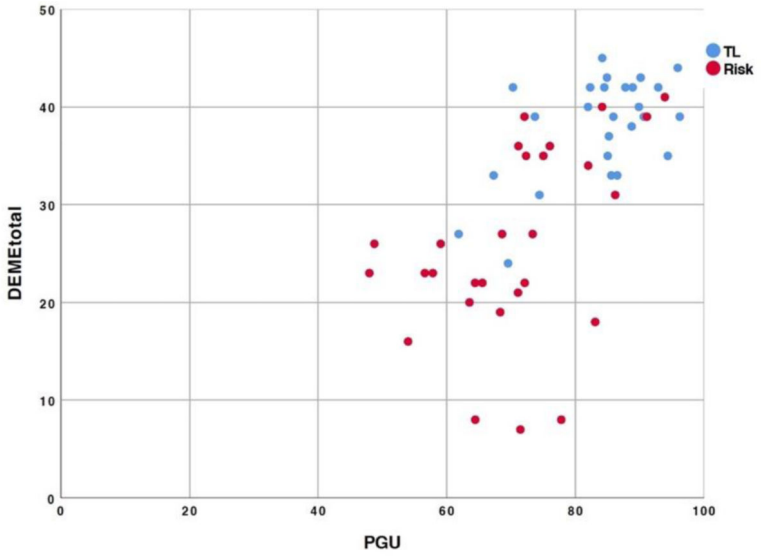

Figure 1.

Correlation between broad measures of grammaticality, Percentage of Grammatical Utterances, and Total DEME score, by group.

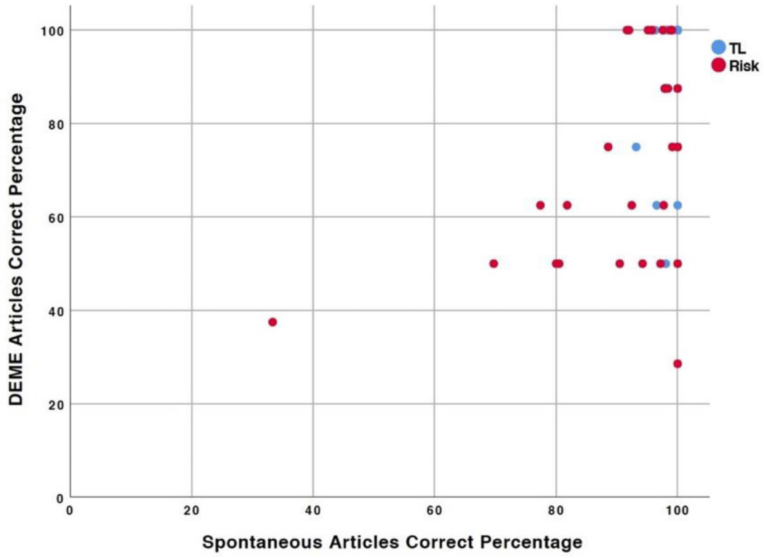

Figure 2.

Correlation between narrow measures of grammaticality, Percentage of correct articles in spontaneous language sample, and article accuracy in DEME, by group.

Table 1.

Descriptive statistics

Measure	TL Group	Risk Group	*t*	*p*	*g*
Utterances	84.5 (20.4)	80.0 (37.2)	.589	.559	0.181
MLUw	6.6 (0.9)	5.4 (1.5)	3.755	.000	0.957
Number of Articles Sample	86.2 (30.6)	60.1(34.8)	2.932	.005	0.796
PGU	83.7 (9.0)	70.1 (11.6)	4.822	.000	1.300
Articles Correct Sample	98.7 (1.7)	91.2 (13.7)	2.778	.008	0.748
Articles Correct DEME	92.8 (14.2)	73.0 (22.4)	3.864	.000	1.043
Total DEME	38.0 (5.3)	25.8 (9.8)	5.638	.000	1.525

Notes: TL= Typical Language Group; g= Hedge's Effect Size

Table 3.

Distribution of children who made broad errors by group

Group	Low DEME Low PGU	Low DEME High PGU	High DEME Low PGU	High DEME High PGU
TL	5	9	1	11
Risk	23	4	0	2

Table 4.

Distribution of children who made article errors by group

Group	Low DEME Low Sample	Low DEME High Sample	High DEME Low Sample	High DEME High Sample
TL	0	7	0	19
Risk	7	14	0	8

Table 5.

Descriptive information for Article Errors

Group	Percentage of Errors of Omission in Sample	Percentage of Errors of Substitution in Sample	Percentage of Errors of Omission in DEME	Percentage of Errors of Substitution in DEME
TL	0.6(1.2)	0.5(1.0)	1.0(3.3)	2.9(8.9)
Risk	6.4(12.2)	1.9(2.4)	7.8(9.7)	7.8(11.3)