A Spanish Sentence Repetition Task and its relationship with spontaneous language in children aged 30 to 36 months

Natalia Bravo[1,2]*, Sonia Mariscal[2], Marta Casla[3] and Miguel Lázaro[4]

[1]Departmento de Pedagogía, Universidad de Castilla la Mancha, Talavera de la Reina, Spain, [2]Departmento de Psicología Evolutiva y de la Educación, Universidad Nacional de Educación a Distancia, Madrid, Spain, [3]Departmento de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid, Madrid, Spain, [4]Departmento de Psicología Experimental, Procesos Cognitivos y Logopedia, Universidad Complutense de Madrid, Madrid, Spain

Sentence Repetition Tasks (SRTs) have been convincingly established as a reliable tool for assessing child language development. However, there are important aspects of this task that deserve more attention. For example, few studies have explored their potential role for identifying language disorders in children under 4 years of age, as almost all evidence refers to children above this age. There is also scarce evidence regarding the relationship between the results of these tasks and measures of spontaneous language. To address this gap, we conducted a study with 24 Typically Developing (TD) monolingual Spanish speakers aged between 30 and 36 months. They performed a Spanish Sentence Repetition Task (SSRT), and their language was recorded and analyzed during spontaneous play with their parents. Variables such as Mean Length of Utterance (MLU), an index of lexical diversity (ILD) and the structure of the Noun Phrase were considered. The statistical analyses reflect a positive and significant correlation between the results obtained in the SSRT and both the MLU and Noun Phrase structure. A positive and significant relationship is also obtained between the MLU in repetition and the MLU of spontaneous language. However, no significant correlation is found between the ILD with either the SSRT or the other measures of spontaneous language. Based on these results, we interpret that the SSRT effectively mirrors the language development of children measured through spontaneous production and is suitable for assessing language skills of Spanish children under 4 years old.

KEYWORDS

sentence repetition task, Mean Length of Utterance, spontaneous language, lexical diversity, early language assessment

Introduction

Research has shown that sentence repetition is a good indicator of children's linguistic skills (Polišenská et al., 2015). Sentence Repetition Tasks (SRTs) have been widely used with children that present language difficulties and have been adapted to different languages (see Rujas et al., 2021 for a recent review). Despite its usefulness and the widespread use of the task both in research and clinical contexts, little is known about the relationship between children's performance in these tasks and their skills in spontaneous language during naturalistic interactions.

Usually, to measure children's linguistic abilities with sentence repetition tasks, researchers and clinicians present a list of sentences to the child, and the child is requested to repeat them. The accuracy of the repetitions, the Mean Length of the Utterances repeated, and the omission and commission errors are taken as indexes of children's linguistic level. Nevertheless, research devoted to analyze whether the same children present similar abilities in their spontaneous speech is very scarce. The aim of this work is to analyze to what extent the abilities that are displayed during sentence repetition are related to the linguistic skills needed for spontaneous speech during the interaction in naturalistic contexts. This evidence would add to concurrent validity of this kind of tasks.

This study analyzed the performance of Spanish-speaking children from 2;6 to 3;0 in a Spanish Sentence Repetition Task (SSRT) that has been previously tested (Bravo et al., 2020, 2023). In addition, we analyzed three characteristics of their linguistic development (MLU, linguistic diversity and the use of Noun Phrases) by taking a sample of spontaneous speech during the interaction with their parents. Finally, we analyzed the results that relate the measurements of both tasks.

The use of SRT

The SRT is an apparently very simple task, requiring children to repeat immediately the linguistic items presented to them by the examiner. However, the task is not simple, as when the child repeats a sentence, he or she does not do so merely mechanically and by rote. In order to respond to an SRT, after listening to a sentence, the listener creates a conceptual representation of the utterance and must activate a series of lexical and grammatical knowledge and processes involved in phonological production in order to subsequently reproduce it (Klem et al., 2014; Andreou et al., 2021).

These types of tasks have been designed and tested in a multitude of languages such as English (Stokes et al., 2006; Baddeley et al., 2009; Seeff-Gabriel et al., 2010; Riches, 2012– comparing English and Cantonese), French (Leclercq et al., 2014), Hungarian (Gábor and Lukács, 2012), Icelandic (Thordardottir, 2008), Mandarin Chinese (Wang et al., 2021), Italian (Devescovi and Caselli, 2007) and Catalan (Gavarró, 2017), among others.

In addition to being used as a language assessment tool with Typically Developing Children (TDC) (Devescovi and Caselli, 2007; Klem et al., 2014; Bravo et al., 2020, 2023), SRTs have also been used as a language measure in research with clinical populations, such as with people with Down syndrome (Koizumi and Kojima, 2022), Williams syndrome (Grant et al., 2002), Autism Spectrum Disorders (Botting and Conti-Ramsden, 2007; Harper-Hill et al., 2013) and with children with hearing aids or cochlear implants (Friedmann and Szterman, 2011; Ruigendijk and Friedmann, 2017). Furthermore, in recent years, SRTs have even been specifically designed for sign language users with hearing impairment, such as the one developed by Schönström and Hauser (2022). However, the most evidence for the value of this task as a tool for detecting language disorders is in relation to children with Developmental Language disorder (DLD) (Conti-Ramsden, 2003; Thordardottir and Brandeker, 2013; Peña et al., 2014; Polišenská et al., 2015; Auza et al., 2018; Simon-Cereijido and Mendez, 2018;

Pratt et al., 2021). Indeed, the effectiveness of SRTs in identifying children with DLD lies precisely in their multifactorial nature (Polišenská et al., 2015). As mentioned above, the task involves the activation of working memory, but also complex linguistic processing of sentence reconstruction and reproduction (Haug et al., 2020).

SRTs as a tool for the assessment of linguistic skills

In relation to this multifactorial nature, Klem et al. (2014) conducted a longitudinal study with 216 TD monolingual Norwegian children aged 4–6 years. The children were assessed with a SRT, a test of vocabulary knowledge and a test of grammatical skills. The authors concluded that the "SRT is best seen as a complex linguistic task that reflects the integrity of language processing at many levels, speech perception, lexical (vocabulary) knowledge, grammatical skills and speech production to name but a few" (Klem et al., 2014, p. 7). As it has been previously stated, SRTs are a useful tool to identify and assess children that present linguistic difficulties. For example, Moll et al. (2015) assessed a group of children with and without dyslexia using the SRT and other measures of language and memory. The results indicated that the children with dyslexia scored worse on the repetition task than the group of children without dyslexia, but they found that these differences were specifically attributable to a subgroup of children who had a history of language development difficulties. The authors also noted that when controlling for memory-related skills, significant differences between the groups remained, which would indicate that the differences were not attributable to this variable as "the memory demands of sentence repetition should not be viewed as distinct from those involved in language production". Other recent works have highlighted the sensitivity of SRTs to measure lexical and grammatical aspects (Polišenská et al., 2015; Simon-Cereijido and Mendez, 2018; Fitton et al., 2019; Schönström and Hauser, 2022). Moreover, in studies where proficient language speakers were asked to repeat ungrammatical sentences, the results indicate that a very high percentage of participants grammatically correct the sentences when repeating them, suggesting that it is not so much memory but rather grammatical knowledge that guides repetition (Over and Gattis, 2010; Schönström and Hauser, 2022).

In Spanish, Bravo et al. (2020) conducted a study with a sample of 130 TD Spanish children aged 2 to 4 years who were assessed using a SSRT, revealing a clear developmental effect. Participants in the 3 to 4 year-old group scored better than those in the 2 to 3 year-old group, indicating that the SSRT is sensitive and reflects the changes that occur in language at this developmental stage. In addition, this SSRT achieved good concurrent validity results, obtaining a positive and significant correlation with a pseudoword repetition task. In a second study Bravo et al. (2023) designed a longitudinal study with children aged 2 to 4 years and analyzed to what extent the score obtained in this task at 33 months (T1) can predict language development 6 months later, at 39 months (T2). Results showed a positive and significant relationship between SSRT scores at T1 and scores

in the expressive language development scale from the Merrill-Palmer-Revised Scales of Development (Roid et al., 2004) applied 6 months later.

Another important issue to consider regarding SRT is that it's not only relevant to analyze the number of sentences repeated correctly or incorrectly, but also to conduct more qualitative analyses. For example, knowing the type of error that is made when the error occurs or the type of word the error affects (function or content words) can provide fundamental information regarding the linguistic and cognitive processes involved. Devescovi and Caselli (2007) showed that, particularly for the youngest group of their participants (aged 2 to 3 years), the omission of articles, prepositions and other modifiers was the most frequent error when repeating SRT items in Italian. From 3 years and 6 months onwards, the mean number of omission errors of function words decreased considerably. In the study conducted in Spanish by Bravo et al. (2020), similar results were found, as children omitted many more function words than content words in their repetitions, with this difference being significant especially in the younger group (2 years to 3 years and 6 months). These results have also been found in other languages, such as English (Seeff-Gabriel et al., 2010; Komeili and Marshall, 2013) and Hungarian (Novogrodsky et al., 2018). The higher rate of omissions in function words than in content words might indicate that SRTs are tasks more sensitive to the morphosyntactic than lexical development of children as young as 2 years old.

SRTs and spontaneous speech

Despite the numerous research studies conducted on SRTs, very few have focused on comparing their results with measures of spontaneous language. Most research has centered on comparing the performance of SRTs with other standardized tasks, ignoring the value of more natural assessments of children's language development, such as the analysis of spontaneous language samples. One of the few works in this line was conducted by Devescovi and Caselli (2007). They carried out a study comparing the results obtained in a sentence repetition task in Italian with some measures extracted from spontaneous language samples. They took a sample of 25 children aged 2–4 years and found that MLU in words, omission of articles and use of verbs in the SRT correlated significantly with the same measures obtained through the analysis of spontaneous language samples. In Spanish, the only study we have found in this line is a pilot study carried out by Moreno-Torres Sánchez et al. (2013) with a group of 10 children aged between 30 and 42 months, with bilateral profound deafness who received a cochlear implant between 12 and 24 months of age. They administered a SRT (PRO-24) comprising a total of 24 sentences (18 simple and 6 compound) and took spontaneous language samples to obtain, among other measures, MLU in spontaneous production. These authors found that most of the children tested scored very low in the sentence repetition test; only two children scored above 20% of correctly repeated sentences, and 5 did not produce any correct responses. However, they also found a significant correlation between the MLU of the spontaneous language sample and the MLU of the

repeated sentences, which seems to indicate a relationship between the morphosyntactic skills measured by the SRT and the skills displayed during spontaneous interaction. A third study is that of Wang et al. (2021), who administered a sentence repetition task to Mandarin Chinese-speaking children. They assessed 59 TD children aged 3.6–6.5 years and compared the results with some indexes extracted from spontaneous language samples such as the MLU, a measure of lexical diversity, the number of predicates (verbs) and a composite structural measure, designed for that study, which evaluates children's correct use of classifiers, aspect markers, passives, and relative clauses. These authors found significant correlations between the results obtained in the SRT and all the spontaneous language measures used, concluding that the SRT adequately reflects the linguistic ability of the children assessed. This conclusion is similar to that obtained by the authors of previous research, which is highly relevant, as it endorses the use of this type of task to infer the actual linguistic development of children. Nonetheless, despite the study conducted by Wang et al. (2021) finding interesting relationships between performance at the SRT and other measures, they presented data from a group with a very wide age- range in which developmental changes were not tracked. Further research is needed to explore the relationships between children's spontaneous speech and their performance in SRTs.

Inferring the linguistic level through the analysis of spontaneous language measures

Regarding spontaneous speech, although the analysis of language samples can be time consuming, it proves to be a valuable method when combined with other external measures (Ambridge and Lieven, 2011). Furthermore, it is possible to obtain different measures from transcriptions of children's speech. The Mean Length of Utterance (MLU) is an index widely used across different contexts since Brown (1973) proposed the milestones associated with the number of words and morphemes that children produced during the first stages of early language acquisition (Parker and Brorson, 2005). MLU, measured as the mean number of words that are part of the utterances produced by the speaker, is a valuable indicator of children's grammatical level that is usually associated with age and is even an index of linguistic delay (Rice et al., 2010).

For Romance languages, with rich morphology, there are other relevant indexes of linguistic development. This is the case of determiners, which carry syntactic and morphological information (number and gender in Spanish and Italian) within Noun Phrases. Different studies in these languages carried out mainly in the nineties (Pizzuto and Caselli, 1992; Bottari et al., 1993; López-Ornat, 2003; Mariscal, 2009) have shown that the acquisition of determiners constitutes a key grammatical development that occurs between 2 and 4 years of age. They found a high frequency of determiner omissions in linguistic contexts of obligatory use during the initial phases of the process of Noun Phrase (NP) acquisition. More recent studies, as Guasti et al. (2008), confirm that children speaking Romance and non-Romance languages omit articles in their earlier productions. And regarding non-typical language development, Bottari et al. (2001) already shown that SLI

children omit determiners significantly more often than almost any other functional category or free morphemes. So, omission of articles and other determiners is a well-known phenomenon in child language. A decrease in this error type is a clear index of grammatical development, that can be good candidate for spontaneous production analysis in a Romance language as Spanish, particularly in very young age.

Children's linguistic variability is also associated with their expressive vocabulary development and therefore with their general linguistic level (Altenberg et al., 2018). The type/token ratio has been broadly used as a measure that is beyond the level of vocabulary (i.e., the number of words produced by the child) and it is employed on the assessment of children's expressive skills. Although it is easy to calculate this measure, it is very dependent on the size of the sample, since very small samples may result very high type/token ratios (Hindman et al., 2021). Another measure that has been proposed as an index of children's vocabulary diversity is the Vocabulary Diversity (VOCD or simply D), which uses an algorithm based on the whole sample to estimate the variety of words in samples with different size (MacWhinney, 2000). The comparisons between type/token ratios and the use of the VOCD showed that the latter was more sensitive detecting children with expressive language delay (Yang et al., 2022). However, further evidence is needed to answer the question of to what extent these indexes, taken from spontaneous language samples, are related to the measures obtained from sentence repetition tasks.

Thus, the aim of this study is to analyze the relationships between a Sentence Repetition Task developed in Spanish (SSRT) (Bravo et al., 2020) and linguistic measures of spontaneous speech, thus providing new evidence on the concurrent validity of the task. In order to analyze these relationships, this study takes measures of children's linguistic skills that have been proved as good indicators of their early grammatical knowledge. We seek to further analyze the characteristics of the SSRT as a tool to assess linguistic development in this language. Therefore, we contrast measures from the SSRT (such as, accuracy, Function Word Omission (FWO), and the MLU of repeated sentences in the task) that could be comparable to measures derived from spontaneous speech sample (such as, Determiner omission within NP, MLU in words and lexical diversity).

Following previous research reviewed above, we expect to find a significant correlation between the accuracy in the SSRT and the other two indexes obtained with the task: MLU of repeated sentences (MLU-r) and FWO. Regarding spontaneous speech, we also hypothesize that there will be significant relationships between children's use of NP, their MLU in words (MLU-w) and the lexical diversity (VOCD). Finally, we expect to find strong relationships between the accuracy of the SSRT and the measures derived from spontaneous speech samples. Specifically, we expect to find significant correlations between SSRT's accuracy, MLU-r and FWO, in relation to the proportion of Determiner omission within NP, the lexical diversity index and MLU-w taken from spontaneous speech samples.

We provide a sample of Spanish-speaking children who completed the SSRT and were video-recorded during interactions with their parents in a naturalistic setting. Furthermore, this study adds to the existing literature in the field by focusing on the age range from 2; 6 to 3 years.

Method

Participants

Twenty-four children (14 boys and 10 girls) aged between 30 and 36 months ($\bar{x}$ = 32.5; SD = 3.02) from different schools in Madrid and Toledo (Spain) participated in this study. All children came from a monolingual background, are Spanish speakers and have a medium socioeconomic family profile. No families expressed concerns about current difficulties in their children's development, nor did they report any previous history of hearing loss or problems in language development.

All families have expressed their willingness to participate in this research by signing the informed consent form, which was previously approved by the ethics committee of Universidad Autónoma de Madrid.

Procedure

The tasks were administered on different days and always in the same order: first the recording of spontaneous language in a natural interaction context and then the assessment with the SSRT.

The SSRT was administered individually by a trained evaluator at the nursery school, within a play-base scenario. This play scenario involves the child teaching a "non-speaking" puppet to speak by repeating sentences that the examiner says one at a time. The examiner presents each sentence clearly and with a marked rhythm, inviting the child to immediately repeat it back to the puppet. Every 5 or 6 sentences, to give the child a short break, a sticker is offered which can be stuck on a paper train. The first two sentences are for training, to make sure that the child understands what he/she is expected to do.

The spontaneous language samples consisted of 15 min of recordings of each child's interaction with his or her mother or father. The average duration of the sessions was 16.03 min (min. 14:11 and max. 20:10). The dyads were recorded at home or in a quiet room in the nursery school each child usually attends, depending on family preferences. The researcher provided the same set of toys to all children (a symbolic play set with cups, plates and spoons, several building blocks, a plastic ball, several dolls representing animals, and a book with pictures depicting actions in different contexts). Families were instructed to play with their child as they normally would. If the child or adult wished to use other toys in the room, they were also encouraged to do so.

Materials and task

Spanish Sentence Repetition Task

This task has been designed by Bravo et al. (2020) to assess the language development of children aged between 24 to 48 months. The SSRT includes a list of 33 sentences of varying length and

morphosyntactic complexity, ranging from 2 to 9 words. It is designed to elicit verbal production of specific morphosyntactic structures, that children typically produce at the ages tested. In developing the task, vocabulary was controlled by extracting words frequently used in the Spanish acquisition process from the database of the Spanish version of the MacArthur Communicative Development Inventory (López-Ornat et al., 2005). Moreover, we considered syllable structure, including words with a simple syllabic structure (consonant-vowel). The complete list of sentences is shown in Appendix A.

Coding

Spanish Sentence Repetition Task

The evaluation with the SSRT was audio-recorded for subsequent orthographic transcription. Next, the following aspects were coded, and the following scores were obtained:

- Accuracy in the child's repetition of sentences. Each correctly reproduced sentence is scored with 1 point; if it is not repeated correctly, 0 points are assigned. The maximum score in the "Accuracy" dimension is, therefore, 33, and the minimum is 0. Children's articulation errors in their utterances are not penalized.
- Mean Length of Utterance in repetition (MLU-r) refers to the number of words the child is able to repeat correctly, divided by the number of sentences he/she repeats during the completion of the task.
- Function word omissions refer to the frequency of FWO errors (determiners, pronouns, conjunctions, adverbs) made by the child when repeating each sentence. We use this value because it is the most frequent error found in sentence repetition and is considered a good index of morphosyntactic development (i.e., fewer omission errors occur as grammatical development progresses) (Devescovi and Caselli, 2007; Bravo et al., 2020).

Spontaneous language samples

A trained researcher transcribed the verbal production of the child and his or her interlocutor using the Child Language Data Exchange System (CHILDES) transcription system (MacWhinney, 2000). Subsequently, all Noun Phrases produced by each child were coded. In order to obtain a reliability index of this coding, another researcher coded the Noun Phrases produced in 25% of the language samples collected. An inter-rater reliability analysis was carried out, reaching a Kappa value of 0.847.

The coding of Noun Phrases was carried out following Mariscal (2009) according to the codes shown in Table 1. Grammatical omissions of determiners and the use of the structure Determiner+Noun indicate more advanced morphosyntactic knowledge, while errors of omission of the determiner indicate lower morphosyntactic knowledge.

In order to investigate the relationship between spontaneous language and performance on the sentence repetition task, the following measures were obtained from the spontaneous language samples:

TABLE 1 Noun phrase coding and examples.

Code	Description	Example	Translation
0N	Grammatical omission of the determiner	*CHI: quiero *agua*	*CHI: (I) need *water*
0N	Agrammatical omission of the determiner	Responding to "qué es esto?" *CHI: *pe* (instead of "un pez")	In response to "what is this?" *CHI: *fi* (instead of "a fish")
DN	Determiner + Noun (any determiner, whether it be an article, demonstrative, possessive …)	*CHI: y dónde está *el lobo*?	*CHI: and where is *the wolf*?
UN	Uncertain	*CHI: este es *te… tete* (instead of "chupete" or "el chupete")	*CHI: this is *te… tete* (instead of "dummy" or "the dummy")–in this case, the syllable "te" in front of the Noun could stand for the article "el (the)" or the first syllable of the Noun.

*CHI stands for "child".

- Vocabulary Diversity Index (VOCD) from the CLAN program of the CHILDES project.
- Mean Length of Utterance in words (MLU-w). We took the MLU-w index using the CLAN program of the CHILDES project.
- Proportion of correct (0N and DN) and incorrect (0*N) Noun Phrases, as morphosyntactic measures.

Results

All the data analyses were conducted with the SPSS program, version 25.0.

Relationship between measures obtained from SSRT

To analyze the relationship between the three SSRT measures, we calculated raw scores for accuracy (number of correctly repeated sentences), MLU-r (calculated across the total number of repeated sentences) and the number of function words that were omitted within the repeated sentences. Then, we conducted a series of bivariate Pearson correlations among these scores.

Table 2 displays the descriptive statistics obtained from the scores in the SSRT. It can be observed that out of the 33 SSRT items, children accurately repeated an average of 13.08 sentences. Regarding MLU-r, we found an average of 4.3 words, with 2.3 being the minimum and 6 being the maximum. Note that the total number of words in the longest sentence is 9. The omission of function words in the repeated sentences averages at 18, with the SSRT including a total of 86 function words across its 33 items.

TABLE 2 Means, standard deviations and ranges of the SSRT measures.

	N	Min	Max	Mean	SD
Accuracy	24	1	30	13.08	7.46
MLU-r	24	2.3	6.03	4.37	0.86
FWO	24	0	50	18	15.18

TABLE 3 Means, standard deviations and ranges of the spontaneous language measures.

	N	Min	Max	Mean	SD
MLU-w	24	1.45	2.52	2.05	0.32
VOCD	24	32.25	80.76	52.89	14.02
Total proportion of noun phrases	24	0.17	0.61	0.31	0.10
Proportion of correct noun phrases	24	0.56	1	0.89	0.11
Proportion of incorrect noun phrases	24	0	0.44	0.11	0.12

In relation to the previous scores, we found a positive and significant correlation between the variable Accuracy and MLU-r $[r_{(24)} = 0.923, p < 0.001]$, a significant and negative correlation between Accuracy and FWO $[r_{(24)} = -0.712, p < 0.001]$, and also a significant and negative correlation between FWO and MLU-r $[r_{(24)} = -0.686, p < 0.001]$. These correlations show, as expected, that children who perform better on the sentence repetition task are those who produce longer sentences and omit fewer function words. Furthermore, children who produce longer utterances tend to make fewer omission errors.

Relationship between measures from spontaneous language

For the purposes of this study, we recorded the total frequency of NP (*number of NP*), summing up instances of correct NP (*DN*), grammatical omissions of determiner *(0N)* and agrammatical omissions of determiner *(0*N)*. Additionally, we computed the number of correct NP by considering *DN* instances along with grammatical omissions of determiner. Lastly, we determined the frequency of incorrect NP, those with agrammatical determiner omissions. For the analyses in this study, we did not take into account the so-called "uncertain Noun Phrases" (see Table 1), due to their ambiguous categorization, and they also constitute only 3.25% of the total sample of NP produced.

To analyze the relationship between the indexes derived from the spontaneous speech sample we initially computed the MLU-w and the diversity index VOCD as it was stated in the method section. We also calculated each child's proportion of NP over the total number of transcribed utterances, along with the ratio of correct and incorrect NP over the total number of NP produced by each child (see Table 3).

As can be seen in Table 3, regarding the MLU-w, we found an average production of 2.05 words, with a minimum of 1.4 and a maximum of 2.52. The Vocabulary Diversity Index (VOCD) shows an average of 52.8, ranging from a minimum of 32.2 to a maximum of 80.7. Note that we report here the values for D optimum average, since the command offers different values. This values usually range from 10 to 100, and higher values indicate higher diversity (McCarthy and Jarvis, 2010). We also found that, out of all the Noun Phrases used by children in their spontaneous language, 89% are produced accurately, while only 11% constituted determiner omission errors.

Regarding the statistical analyses, we found that MLU-w positively and significantly correlates with the proportion of correct NP structures $[r_{(24)} = 0.685, p < 0.001]$ and negatively and significantly correlates with the proportion of determiner omission errors in NP $[r_{(24)} = -0.685, p < 0.001]$. We did not find a significant relationship between VOCD and child-produced MLU-w $[r_{(24)} = 0.378, p = 0.068]$, nor between the VOCD index and the correct usage of NPs $([r_{(24)} = 0.231, p = 0.276]$.

Relationship between SSRT and spontaneous language measures

To analyze the relationship between the SSRT scores and the measures obtained from the spontaneous language samples we performed further bivariate Pearson correlations. Table B1 in the Appendix B show the results of these analyses. We found a positive and significant correlation between Accuracy on the SSRT and MLU-w in spontaneous language $[r_{(24)} = 0.435, p < 0.05]$, indicating that children who perform better on the SSRT tend to produce longer utterances in their spontaneous language. Regarding the relationship between MLU-w and MLU-r, the results also show a positive and significant correlation between these 2 variables $[r_{(24)} = 0.460; p < 0.05]$. On the other hand, the VOCD index has a positive, but not significant, relationship with the SSRT Accuracy score $[r_{(24)} = 0.319, p = 0.129]$.

Regarding the structure of Noun Phrases and their relation to SSRT, a positive and significant correlation is obtained between the proportion of incorrect NP (i.e. determiner omission errors) produced in spontaneous language and FWO in the SSRT $[r_{(24)} = 0.463; p = 0.023]$ and a negative and significant correlation between the proportion of such determiner omission errors and the total score in the SSRT $[r_{(24)} = -0.490, p = 0.015]$. That is, children who make more determiner omission errors in their spontaneous language achieve lower scores in the SSRT. On the contrary, a positive and significant correlation is found between the correct use of NP (sum of DN and 0N) and the total score in the SSRT $(r = 0.490; p = 0.015)$ and the MLU-r $(r = 0.444; p = 0.023)$.

The correlation analyses conducted between the various measures of spontaneous language considered in our study and the SSRT indicate that the child's language MLU-w and measures related to NP usage are closely linked to the child's performance in the SSRT. To assess the extent to which each of these variables explains the performance in the SSRT, we conducted a multiple linear regression analysis considering that this kind of analysis is more suitable than other options as ANOVA for example,

to specifically explore accuracy measures (Jaeger, 2008). The dependent variable was the accuracy in the SSRT, and the predictors were linguistic measures obtained from the spontaneous language sample (MLU-w and the proportion of correct NP). We found that the model fits both variables (proportion of correct NP and MLU-w) as follows: $F_{(3,20)} = 3.6$, $p = 0.043$. We also found that these two variables account for a significant portion of the variance in SSRT accuracy ($R^2 = 0.26$).

Discussion

Despite the extensive research on SRTs, few studies have compared their outcomes with children's spontaneous language measures. Most research has focused on comparing SRT performance with standardized tasks (see Hesketh and Conti-Ramsden, 2013; Thordardottir and Brandeker, 2013; Aguado et al., 2018; Bravo et al., 2020), overlooking the value of assessing children's language development through natural language sample analyses. In this context, the aim of this study was to examine the relationship between different scores obtained in a sentence repetition task developed in Spanish (Bravo et al., 2020, 2023) for children from 2 to 4 years of age and other well-attested measures taken from spontaneous speech samples. Taken together, our results show that children's performance at the SSRT is related to their grammatical skills expressed during their spontaneous interactions with their caregivers.

We first found a strong relationship between the three measures of the SSRT; children with higher accuracy also exhibited lower levels of function words omission and higher scores in the MLU-r. Other studies, such as Leclercq et al. (2014) in French, have also analyzed the relationship between internal measures of the SRT (or sub-measures) and the total score obtained from the task, finding similar results. The authors concluded that the overall morphosyntactic measure in the SRT strongly correlates with function word measures, verbal morphology, and grammatical accuracy in repetition, affirming that the task effectively mirrors its intended purpose. In their study with Italian-speaking children, Devescovi and Caselli (2007) also found significant correlation between the different measures of the SRT, number of repeated verbs, FWO, and MLU-r. Although our study does not use identical measures, it complements previous research by demonstrating consistency in the measures obtained using the SSRT.

The analyses concerning the indexes derived from spontaneous language samples revealed robust associations between MLU-w and the accurate usage of Noun Phrase. Children with longer MLU-w exhibited fewer omission errors of Determiners in NP. Previous research has found that NP is commonly acquired early and is a frequent structure both in experimental and naturalistic contexts (Mariscal, 2009). The correlation between NP measures used in this study and MLU suggests that both measures are good indicators of children's grammatical knowledge at 2.6 (Rice et al., 2010).

Regarding the relationship between the measures derived from the SSRT and the spontaneous speech samples, our findings indicate that children's SSRT accuracy and MLU-r are strongly related to both the MLU-w and the correct use of NP when they produce the sentences in naturalistic situations. This result suggests that the measures obtained with the SSRT serve as reliable indicators of children's grammatical development and that the mechanisms involved to produce elicited sentences during repetition in clinical or experimental contexts reflect the skills necessary to produce sentences spontaneously. Previous research, such as Devescovi and Caselli (2007) in Italian, also identified a correlation between the SSRT and specific measures derived from spontaneous language: MLU, FWO, and verb usage. However, the participants' age range was broad (2 to 4 years), and, when age was controlled, the associations between SSRT accuracy and certain aspects of spontaneous speech were weaker and sometimes non-existent. Moreover, Wang et al. (2021) established a connection between SRT, MLU, VOCD, and a composite structural measure of spontaneous language in their study involving Mandarin-speaking children aged between 3.6 and 6.5 years. Compared to these studies, our research focuses on a narrower age range and results show the relationships between the ability to repeat sentences (typically used in clinical and experimental settings) and spontaneous speech during the earliest phases of syntactic development.

In our study, we did not find a relationship between the SSRT and the Vocabulary Diversity Index (VOCD), nor did we find a correlation between this index and the rest of the measures used for the analysis of spontaneous language samples (neither MLU-w nor the use of determinants in Noun Phrase). There could be several reasons for this. Despite the recent presentation of VOCD as a promising measure for assessing children's lexical diversity, it is not without issues. In a recent study, Yang et al. (2022) tested 4 measures used in clinical practice to calculate the lexical diversity index: the Type-Token Ratio (TTR) index, the Number of Different Words (NDW), the Moving Average Type Token Ratio (MATTR) and the Vocabulary Diversity (VOCD). They found that, if these indexes are used as a measure of lexical richness, the VOCD and NWD are the ones that best reflect the vocabulary of the participants, but did not show correlation with traditional measures of syntax, like MLU in words or morphemes.

Although numerous studies have found continuity between lexical and grammatical skills (Mariscal and Gallego, 2013), this relationship fundamentally refers to the early stages. A critical mass of vocabulary is necessary to support the construction of the first multiword utterances and morphological variations (Marchman and Bates, 1994). However, at later ages this relationship is far from direct, and is closely related to contexts and input (Brinchmann et al., 2019). On the other hand, the children in our sample are very young, between 30 and 36 months, and it is possible that at these ages the lexical diversity is not varied enough to reflect the grammatical level of each child. In fact, the words used in the repetition task are frequent, the lexical diversity indexes allow us to reflect the use of rare words, which, in other studies, is related, for example, to access to literacy (Hindman et al., 2021). Therefore, to solve the SRT it is not necessary to have a very high variety of word types, although it is necessary to have a minimum vocabulary that the children in this sample, being typically developing, clearly achieve.

It is interesting to note that this fact allows us to think that the SRT is a very specific assessment tool for grammatical development that can be administered to children whose lexical skills follow a typical course of development but who may

begin experiencing specific grammatical difficulties from an early age.

Further research with spontaneous language samples in diverse contexts could provide more insight into this matter, adding new and more fine-grained evidence that contributes to concurrent validity of SSRT.

Taken together, the results of the present study with the Spanish Sentence Repetition Task (SSRT) are in line with previous research that has shown the usefulness of sentence repetition tasks in assessing children aged between 2 and 4 years old (Devescovi and Caselli, 2007; Gábor and Lukács, 2012; Novogrodsky et al., 2018; Bravo et al., 2020, 2023). Its ability to discriminate between different developmental levels has been demonstrated, and concurrent and predictive validity results support its use (Bravo et al., 2020, 2023). Establishing that this new tool, simple and easy to apply, may provide results mirroring those obtained through the study of spontaneous language, could represent a valuable opportunity in the field of child language assessment. Usually, the analysis of spontaneous speech is complex and time consuming and therefore is not suitable for clinical contexts. The correlations shown in this study suggest that the scores obtained with the SSRT reflect children's linguistic knowledge.

While Bravo et al. (2020) study involved 130 children, the current research was carried out with a smaller sample, allowing us to focus on the earliest stages of grammatical development in younger children. Although the sample size is a limitation to consider in this study, analyzing spontaneous language samples is very time and resource-consuming. We believe that the results could provide more valuable information if we had a larger sample. Similarly, it would have been interesting to include children with different ages, and to study the relationship between language measures and the SSRT at different development stages. In fact, this is the first study of this type conducted with a monolingual Spanish-speaking population at such an early developmental phase. Therefore, it would be possible to continue analyzing measures of the SSRT that may reflect more sophisticated changes in morphological and syntactic development.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Universidad Autónoma de Madrid. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2024.1368035/full#supplementary-material

References

Aguado, G., Ripoll, J. C., Tapia, M. M., and Gibson, M. (2018). Marcadores del trastorno específico del lenguaje en español: Comparación entre la repetición de oraciones y la repetición de pseudopalabras. *Revista de Logopedia, Foniatría y Audiología* 38, 105–112. doi: 10.1016/j.rlfa.2018.04.003

Altenberg, E. P., Roberts, J. A., and Scarborough, H. S. (2018). *Young Children's Structure Production: A Revision of the Index of Productive Syntax (world) [Brief-report]*. Rockville MA: ASHA Wire; American Speech-Language-Hearing Association.

Ambridge, B., and Lieven, E. V. M. (2011). *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press.

Andreou, M., Tsimpli, I. M., Masoura, E., and Agathopoulou, E. (2021). Cognitive mechanisms of monolingual and bilingual children in monoliterate educational settings: evidence from sentence repetition. *Front. Psychol.* 11:613992. doi: 10.3389/fpsyg.2020.613992

Auza, A., Kapantzoglou, M., and Murata, C. (2018). Two Grammatical tasks for screening language abilities in Spanish-speaking children. *Am. J. Speech-Lang. Pathol.* 27:1. doi: 10.1044/2017_AJSLP-17-0052

Baddeley, A. D., Hitch, G. J., and Allen, R. J. (2009). Working memory and binding in sentence recall. *J. Mem. Lang.* 61, 438–456. doi: 10.1016/j.jml.2009.05.004

Bottari, P., Cipriani, P., and Chilosi, A. M. (1993). Protosyntactic devices in the acquisition of italian free morphology. *Lang. Acquis.* 3, 327–369. doi: 10.1207/s15327817la0304_1

Bottari, P., Cipriani, P., Chilosi, A. M., and Pfanner, L. (2001). The Italian determiner system in normal acquisition, specific language impairment, and childhood aphasia. *Brain Lang.* 77, 283–293. doi: 10.1006/brln.2000.2402

Botting, N., and Conti-Ramsden, G. (2007). Autism, primary pragmatic difficulties, and specific language impairment: can we distinguish them using psycholinguistic markers? *Dev. Med. Child Neurol.* 45, 515–524. doi: 10.1111/j.1469-8749.2003.tb00951.x

Bravo, N., Lázaro, M., and Mariscal, S. (2020). A sentence repetition task for early language assessment in Spanish. *Span. J. Psychol.* 23:e39. doi: 10.1017/SJP.2020.43

Bravo, N., Lázaro, M., Rujas, I., and Mariscal, S. (2023). A sentence repetition task in Spanish language: a valid tool for early language assessment. *CoDAS* 35:e20220164. doi: 10.1590/2317-1782/20232022164en

Brinchmann, E. I., Braeken, J., and Lyster, S.-A. H. (2019). Is there a direct relation between the development of vocabulary and grammar? *Dev. Sci.* 22:e12709. doi: 10.1111/desc.12709

Brown, R. (1973). Development of the first language in the human species. *American Psychologist* 28, 97–106. doi: 10.1037/h0034209

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *J. Speech Lang. Hear. Res.* 46, 1029–1037. doi: 10.1044/1092-4388(2003/082)

Devescovi, A., and Caselli, M. C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *Int. J. Lang. Commun. Dis.* 42, 187–208. doi: 10.1080/13682820601030686

Fitton, L., Hoge, R., Petscher, Y., and Wood, C. (2019). Psychometric evaluation of the bilingual English–Spanish assessment sentence repetition task for clinical decision making. *J. Speech Lang. Hear. Res.* 62, 1906–1922. doi: 10.1044/2019_JSLHR-L-18-0354

Friedmann, N., and Szterman, R. (2011). The comprehension and production of WH-questions in deaf and hard-of-hearing children. *J. Deaf Stud. Deaf Educ.* 16, 212–235. doi: 10.1093/deafed/enq052

Gábor, B., and Lukács, Á. (2012). Early morphological productivity in Hungarian: Evidence from sentence repetition and elicited production. *J. Child Lang.* 39, 411–442. doi: 10.1017/S0305000911000110

Gavarró, A. (2017). A sentence repetition task for catalan-speaking typically-developing children and children with specific language impairment. *Front. Psychol.* 8:1865. doi: 10.3389/fpsyg.2017.01865

Grant, J., Valian, V., and Karmiloff-Smith, A. (2002). A study of relative clauses in Williams syndrome. *J. Child Lang.* 29, 403–416. doi: 10.1017/S030500090200510X

Guasti, M. T., Gavarró, A., de Lange, J., and Caprin, C. (2008). Article omission across child languages. *Lang. Acquis.* 15, 89–119. doi: 10.1080/10489220801937346

Harper-Hill, K., Copland, D., and Arnott, W. (2013). Do spoken nonword and sentence repetition task discriminate language impairment in children with ASD? *Res. Autism Spectr. Disord.* 7, 265–275. doi: 10.1016/j.rasd.2012.08.015

Haug, T., Batty, A. O., Venetz, M., Notter, C., Girard-Groeber, S., Knoch, U., et al. (2020). Validity evidence for a sentence repetition test of Swiss German Sign Language. *Lang. Test.* 37, 412–434. doi: 10.1177/0265532219898382

Hesketh, A., and Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of specific language impairment. *PLoS ONE* 8:e56314. doi: 10.1371/journal.pone.0056314

Hindman, A. H., Farrow, J. M., Anderson, K., Wasik, B. A., and Snyder, P. A. (2021). Understanding child-directed speech around book reading in toddler classrooms: evidence from early head start programs. *Front. Psychol.* 12:719783 doi: 10.3389/fpsyg.2021.719783

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards Logit Mixed Models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A., Gustafsson, J.-E., and Hulme, C. (2014). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Dev. Sci.* 18:12202. doi: 10.1111/desc.12202

Koizumi, M., and Kojima, M. (2022). Syntactic development and verbal short-term memory of children with autism spectrum disorders having intellectual disabilities and children with down syndrome. *Autism Dev. Lang. Impairm.* 7:239694152211096. doi: 10.1177/23969415221109690

Komeili, M., and Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clini. Linguist. Phonet.* 27, 152–162. doi: 10.3109/02699206.2012.751625

Leclercq, A.-L., Quémart, P., Magis, D., and Maillart, C. (2014). The sentence repetition task: A powerful diagnostic tool for French children with specific language impairment. *Res. Dev. Disabil.* 35, 3423–3430. doi: 10.1016/j.ridd.2014.08.026

López-Ornat, S. (2003). Problemas ficticios y problemas reales sobre el desarrollo gramatical/real and false problems on grammatical development. *Cognitiva* 15, 177–185. doi: 10.1174/021435503769204694

López-Ornat, S., Gallego, C., Gallo, M. P., Karousou, A., Mariscal, S., and Martínez, M. (2005). *MacArthur: Inventario de desarrollo comunicativo*. Madrid: TEA Ediciones.

MacWhinney, B. (2000). The CHILDES project: tools for analyzing talk. *Child Lang. Teach. Ther.* 8:211. doi: 10.1177/026565909200800211

Marchman, V. A., and Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *J. Child Lang.* 21, 339–366. doi: 10.1017/S0305000900009302

Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *J. Child Lang.* 36, 143–171. doi: 10.1017/S0305000908008908

Mariscal, S., and Gallego, C. (2013). *La imitación como herramienta para investigar y evaluar el desarrollo lingüístico temprano: Un estudio piloto de repetición de palabras y pseudopalabras*.

McCarthy, P. M., and Jarvis, S. (2010). MTLD, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. *Behav. Res. Methods* 42, 381–392. doi: 10.3758/BRM.42.2.381

Moll, K., Hulme, C., Nag, S., and Snowling, M. J. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Appl. Psycholinguist.* 36, 203–221. doi: 10.1017/S0142716413000209

Moreno-Torres Sánchez, I., Madrid Cánovas, S., and Moruno López, E. (2013). Prueba repetición de oraciones para niños de 24 a 48 meses (PRO-24). Estudio piloto con niños típicos y niños sordos con implante coclear. *Revista de Logopedia, Foniatría y Audiología* 33, 25–35. doi: 10.1016/j.rlfa.2012.10.003

Novogrodsky, R., Meir, N., and Michael, R. (2018). Morphosyntactic abilities of toddlers with hearing impairment and normal hearing: evidence from a sentence-repetition task. *Int. J. Lang. Commun. Dis.* 53, 811–824. doi: 10.1111/1460-6984.12386

Over, H., and Gattis, M. (2010). Verbal imitation is based on intention understanding. *Cogn. Dev.* 25, 46–55. doi: 10.1016/j.cogdev.2009.06.004

Parker, M. D., and Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Lang.* 25, 365–376. doi: 10.1177/0142723705059114

Peña, E. D., Gillam, R. B., and Bedore, L. M. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *J. Speech Lang. Hear. Res.* 42, 381–392. doi: 10.1044/2014_JSLHR-L-13-0151

Pizzuto, E., and Caselli, M. C. (1992). The acquisition of Italian morphology: implications for models of language development. *J. Child Lang.* 19, 491–557. doi: 10.1017/S0305000900011557

Polišenská, K., Chiat, S., and Roy, P. (2015). sentence repetition: what does the task measure? *Int. J. Lang. Commun. Dis.* 50, 106–118. doi: 10.1111/1460-6984.12126

Pratt, A. S., Peña, E. D., and Bedore, L. M. (2021). Sentence repetition with bilinguals with and without DLD: differential effects of memory, vocabulary, and exposure. *Biling.: Lang. Cogn.* 24, 305–318. doi: 10.1017/S1366728920000498

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., and Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *J. Speech Lang. Hear. Res.* 53, 333–349. doi: 10.1044/1092-4388(2009/08-0183)

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *Int. J. Lang. Commun. Dis.* 47, 499–510. doi: 10.1111/j.1460-6984.2012.00158.x

Roid, G. H., Sampers, J. L., Anderson, G., Erickson, J., and Post, P. (2004). *Merrill-Palmer-Revised Scales of Development.* Stoelting Company (Spanish version: TEA Ediciones, 2011).

Ruigendijk, E., and Friedmann, N. (2017). A deficit in movement-derived sentences in german-speaking hearing-impaired children. *Front. Psychol.* 8:689. doi: 10.3389/fpsyg.2017.00689

Rujas, I., Mariscal, S., Murillo, E., and Lázaro, M. (2021). Sentence repetition tasks to detect and prevent language difficulties: a scoping review. *Children* 8:578. doi: 10.3390/children8070578

Schönström, K., and Hauser, P. C. (2022). The sentence repetition task as a measure of sign language proficiency. *Appl. Psycholinguist.* 43, 157–175. doi: 10.1017/S0142716421000436

Seeff-Gabriel, B., Chiat, S., and Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *Int. J. Lang. Commun. Dis.* 45, 691–702. doi: 10.3109/13682820903509432

Simon-Cereijido, G., and Mendez, L. (2018). Using language-specific and bilingual measures to explore lexical–grammatical links in young latino dual-language learners. *Lang. Speech Hear. Serv. Schools* 49:1. doi: 10.1044/2018_LSHSS-17-0058

Stokes, S., Wong, A., Fletcher, P., and Leonard, L. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: the case of Cantonese. *J. Speech Lang. Hear. Res.* 49, 219–236. doi: 10.1044/1092-4388(2006/019)

Thordardottir, E. (2008). Language-specific effects of task demands on the manifestation of specific language impairment: a comparison of English and Icelandic. *J. Speech Lang. Hear. Res.* 51, 922–937. doi: 10.1044/1092-4388(2008/068)

Thordardottir, E., and Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J. Commun. Disord.* 46, 1–16. doi: 10.1016/j.jcomdis.2012.08.002

Wang, D., Zheng, L., Lin, Y., Zhang, Y., and Sheng, L. (2021). Sentence repetition as a clinical marker for mandarin-speaking preschoolers with developmental language disorder. *J. Speech Lang. Hear. Res.* 65:401. doi: 10.1044/2021_JSLHR-21-00401

Yang, J. S., Rosvold, C., and Bernstein Ratner, N. (2022). Measurement of lexical diversity in children's spoken language: computational and conceptual considerations. *Front. Psychol.* 13:905789. doi: 10.3389/fpsyg.2022.905789