The Non-word Repetition Task as a clinical marker of Specific Language Impairment in Spanish-speaking children

Dolors Girbau
Universitat Jaume I, Spain

Abstract
Forty native Spanish-speaking children (age 8;0–10;3), 20 with Specific Language Impairment (SLI) and 20 with Typical Language Development (TLD), received a battery of psycholinguistic tests, IQ, hearing screenings, and the Spanish Non-word Repetition Task (NRT). The children’s repetition of 20 non-words was scored. The percentage of correct non-words was significantly lower in children with SLI than in age-matched children with TLD. A length effect was found, with the subset of three-, four-, and five-syllable non-words leading to greater differences between groups. The NRT identified SLI accurately; likelihood ratios are reported with significant good sensitivity and specificity. Significant positive correlations were found, for all children together, between the overall NRT accuracy and the eight contemporary expressive/receptive language scores: PPVT-III, TTFC-2, and CEG tests; WISC-IV/Vocabulary subtest; and four ITPA subtests, including lexical fluency and comprehension of sentences/stories. The Spanish NRT can be used as a diagnostic marker for SLI. The clinical implications of phonological working memory links to psycholinguistic abilities are discussed.

Keywords
Assessment, diagnosis, expressive/receptive language abilities, phonological working memory deficit, Spanish Non-word Repetition Task, Specific Language Impairment

Corresponding author:
Dolors Girbau, Department of Basic, Clinical and Biological Psychology, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castelló, Spain.
Email: girbau@psb.uji.es
The identification of Specific Language Impairment (SLI) in elementary school children, especially in languages other than English, is still a challenge for professionals. There is a need for more instruments to diagnose Spanish-speaking children with SLI. This study discusses research on the Non-word Repetition Task (NRT) and presents new data on the Spanish NRT from a sample of elementary school children. In general, the task is considered a reliable diagnostic marker to identify SLI in children across several languages. The NRT has been normed in a few languages, but not yet in Spanish.

The key relationship between the NRT and other diagnostic language tests has not been analyzed in detail. Thus, the present research also explores the underlying key role of phonological working memory, as measured by the NRT, in respect of eight contemporary psycholinguistic tests/subtests at 8–10 years of age.

Identification of children with Specific Language Impairment

Specific Language Impairment (SLI) is usually defined as a deficit in language understanding and/or production, in the absence of other significant developmental/hearing deficits, autism, and severe neurological impairment (Leonard, 1998; Schwartz, 2009). That is, children with SLI have expressive and/or receptive language scores that are significantly below the age norms. The DSM-5, instead, uses the term ‘Language Disorder’ (within neurodevelopmental disorders) to define persistent difficulties in the acquisition and use of language due to deficits in comprehension or production (American Psychiatric Association, 2013). However, the DSM-5 does include the word ‘Specific’ for the category ‘Specific Learning Disorder’ (with possible reading impairment). Previously, the manual differentiated between either expressive or mixed receptive-expressive language disorders (American Psychiatric Association, 2000).

In the past, the term LI (Language Impairment) has been used when participants were not recruited through the standard procedures, as for example, not including any hearing screening or individual cutoff for the standardized language tests. Studies are progressively becoming more stringent, although the terminological and conceptual debate remains open. Recently, some authors have recommended the use of the term LI instead of SLI or ‘Language Disorder’ (see Reilly et al., 2014, for a critical review). However, others, such as Leonard, in a commentary section to Reilly’s paper, have argued that the alternative term LI would create more problems of boundary confusion than it would solve. Leonard also supports Bishop’s (2014) comments that SLI is used in far more instances in the published literature than alternative terms. According to Bishop, many of the labels employed (including ‘Language Disorder’) are too general to be useful, and, as changing a label can disrupt links with accumulated research evidence, it should not be undertaken lightly. While acknowledging that this debate is likely to continue, we subscribe here to the term SLI.

That said, identifying SLI in native speakers of languages other than English is a challenging process, especially due to the lack of tasks and standardized language tests. There is still a need for more research on SLI in other languages. For example, despite
the fact that Spanish ranks second in the world’s languages in terms of number of native speakers, SLI in Spanish has received relatively little attention (Girbau, 2010).

It is established that SLI is manifest in different ways depending on the specific characteristics of a target language (phonetics, morphology, syntax, or even semantics), despite the presence of common markers across languages (e.g., some working memory deficits). The study of SLI in languages other than English (e.g., Spanish, French, Italian, Cantonese, German, Finnish, Swedish, Hebrew, Icelandic) is contributing progressively to advancing our knowledge of this disorder.

For example, NRTs with specific phonological features have been created in different languages to better diagnose children with SLI, a disorder that usually involves a phonological working memory deficit (see next section). The NRT requires children to repeat non-words of increasing syllable length. It has been used as a diagnostic marker for identifying SLI. An MRI study (Girbau, Garcia, Marti, & Schwartz, 2014) used the Spanish NRT and two out of several Spanish oral language standardized subtests/tests as criteria for diagnosing elementary school children with SLI. In this study, each child within a comorbid subgroup with SLI and Reading Disabilities (RD) had a low score on the NRT. Furthermore, the authors found significant common brain volumetric markers for both the group with SLI and the comorbid group with SLI and RD (e.g., lower gray matter volume at the right postcentral parietal gyrus). This finding provides support to the SLI diagnostic category and the mentioned diagnostic markers based on the NRT.

Non-word repetition tasks and language processes in SLI across languages

The NRT has been used occasionally either as a normed screening test or with likelihood ratios, and more often as a non-standardized task in small samples. Several studies have indicated that this task is, or has the potential to be, a reliable test for SLI/LI in preschool children (e.g., Dispaldro, Leonard, & Deevy, 2013; Gray, 2003; Roy & Chiat, 2004) and elementary school students (e.g., Dollaghan & Campbell, 1998; Girbau & Schwartz, 2007, 2008). It has also been found to be a clinical marker for adults with SLI (Poll, Betz, & Miller, 2010). Other studies indicate that poor non-word repetition in children is a marker of SLI, with high genetic heritability (Bishop, North, & Donlan, 1996; Newbury et al., 2009). Also suggestive of genetic links, poor performance in the NRT among children is related to a similarly poor level of performance in the NRT by their mothers (Girbau & Schwartz, 2008). Thus, the task has also the potential to be a good screening test for the identification of children at risk for SLI.

Two well-known NRTs for diagnosing SLI were created following the English phonotactic patterns; both were presented using voice recorded items (Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1996). The performances on these NRTs by children with SLI or Typical Language Development (TLD) were compared by Archibald and Gathercole (2006). The Children’s Test of Non-word Repetition included 40 pseudowords with English stress patterns, from two to five syllables in length (10 items at each length, e.g., *trumpetine*; Gathercole & Baddeley, 1996). It was normed on 612 children with TLD, on the basis of the number of correct responses from a minimum of 84
children in each age interval (ranging from 4;0 to 9;0). Their performance was found to improve progressively with age. On the other hand, the English NRT contained 16 pseudowords with non-English prosodic from one to four syllables in length (4 items at each length, e.g., doitauvab; Dollaghan & Campbell, 1998). It was administered to 40 children between age 6;0 and 9;9, half of whom were diagnosed as having LI only, on the basis of receiving language intervention. These children with LI were found to be less accurate than the age-matched children with TLD, especially on the three- and four-syllable items. Likelihood ratios revealed that a score of \( \leq 70\% \) phonemes correct identified a child as LI, and a percentage of 81% or above identified a child as having TLD. In a later study, the likelihood ratios were not as large in children with a mean age of 7;11 with a narrower age range (7;1–8;11); children with SLI were identified on the basis of standardized language tests/subtests (Ellis Weismer et al., 2000).

Non-word repetition tasks have been created in Spanish (Girbau & Schwartz, 2007, 2008), French (Le Foll, Godin, Jacques, & Taillant, 1995), Italian (Bortolini et al., 2006), Swedish (Bartholom & Åkesson, 1995; Sahlén, Reuter斯基öld-Wagner, Nettelbladt, & Radeborg, 1999), Cantonese (Stokes, Wong, Fletcher, & Leonard, 2006), Icelandic (Thordardottir, 2008), and Portuguese (Hage, Nicolielo, & Guerreiro, 2014). Usually, children with SLI/LI have performed more poorly than children with TLD, and length effects were found across several languages, especially in repeating non-words of three or more syllables (e.g., Bortolini et al., 2006; Dollaghan & Campbell, 1998; Girbau & Schwartz, 2007, 2008; Le Foll et al., 1995; Rodekohr & Haynes, 2001; see also Graf Estes, Evans, & Else-Quest, 2007, for a review).

NRT performance involves additional cognitive and language processes but it makes demands especially on phonological working memory. In this task, the individual listens to a made up word, temporarily stores the unfamiliar phonological representation, and then produces it. This phonological working memory process involves the phonological loop, which is a cognitive subsystem with the following two components: (a) phonological short-term storage, which maintains the incoming auditory sequence of segments/syllables as a phonological code, and (b) subvocal rehearsal or inner speech, which holds this phonological representation avoiding its decay (Baddeley, 1986, 1996, 2000, 2003a, 2003b). Thus, the term working memory refers to the system/s that may be needed to keep things in mind while performing a complex task (Baddeley, 2010). Usually, studies have explained the poor results on the complex NRT by children with SLI in terms of a phonological working memory deficit (e.g., Hage et al., 2014).

There are indications that this deficit affects also other language processes that involve phonological information processing, including, for example, the acquisition of new words and sentence comprehension (e.g., Briscoe, Bishop, & Norbury, 2001; Jackson, Leitao, & Claessen, 2015). According to some cognitive processing studies, which support the link between real-time processing and long-term knowledge, a more developed representational system (including large vocabulary) is needed for an accurate repetition of low-frequency non-words. In particular, 3- to 6-year-old children with/without Phonological Disorders (PD) produced the non-words with low-frequency syllables less accurately than those with high-frequency syllables; children who were diagnosed as having PD-only scored significantly low on an English test of articulation (Munson, Edwards, & Beckman, 2005). Thus, performance in the NRT is also affected by
the syllable frequencies. Alternative explanations to the phonological working memory deficit have also been investigated at different ages, as for example a possible deficit in phonological sensitivity. However, elementary school children (age 7;3–10;6) with SLI seemed to extract phonological regularities comparably to their age-matched children with TLD (Coady, Evans, & Kluender, 2010).

Some studies have analyzed the relation between the non-word repetition performance and some standardized language tests scores in children with SLI/LI and/or TLD, mostly in English speakers, with somewhat mixed findings. Non-word repetition was significantly related to receptive vocabulary in English-speaking children with TLD (Metsala, 1999; Roy & Chiat, 2004). In children with TLD, too, superior non-word repetition performance was related to a broader vocabulary and the production of longer and syntactically more varied utterances, when compared with children who performed poorer in the non-words task (Adams & Gathercole, 2000). Similarly, a group of children with larger vocabularies performed more accurately in the NRT than a smaller vocabulary group (Edwards, Beckman, & Munson, 2004). Montgomery (2004) reported that non-word repetition did not correlate significantly with sentence comprehension in 12 children with SLI or their 12 age-matched peers with TLD (6;4–10;5). Finally, subgroups of children with SLI (10;11) who had the lowest and the highest non-word repetition performance differed significantly in oral sentence comprehension as measured by the Test for Reception of Grammar (TROG; Bishop, 1982), but not in their expressive/receptive vocabulary scores (Botting & Conti-Ramsden, 2001).

In other languages, Icelandic non-word repetition scores correlated significantly with morphological accuracy in spontaneous language production for 22 school children with SLI and TLD together (Thordardottir, 2008). Percentage consonants correct on Swedish non-word repetition (Barthelom & Åkesson, 1995) correlated positively with four receptive language measures, including oral comprehension of words, stories, and sentences (the last showing the strongest link), in 27 preschool children with LI (Sahlén et al., 1999).

Finally, Spanish non-word repetition scores correlated highly with auditory association and grammatical integration in completing oral sentences, but not with verbal expression and auditory comprehension of stories, for the overall sample of 22 children with SLI and TLD from Spain (Girbau & Schwartz, 2007). The same was true for the 22 Spanish/English-speaking children with SLI and TLD together from New York City, except for auditory comprehension, which also correlated significantly with the non-word performance (Girbau & Schwartz, 2008). In short, findings are somewhat mixed and do seem to vary among languages. Hence, at this stage of our knowledge, more correlational studies are needed and it is important that we build up more detailed information on performance in a variety of languages.

**Purpose of the study**

The present study aimed to determine the potential use of the auditory Spanish NRT (Girbau & Schwartz, 2007, 2008) as a diagnostic marker to identify Spanish-speaking children with SLI from 8 to 10 years of age. Our main goal was to determine the accuracy with which this task can distinguish children with and without SLI, including the calculation of likelihood ratios, sensitivity, and specificity values.
We wanted also to explore the relationships between non-word repetition performance and scores from a variety of contemporary tests of expressive/receptive psycho-linguistic abilities. At present, findings across the literature are somewhat mixed. The present study could contribute further evidence on the development of production and comprehension abilities in respect of both typical and atypical children. Finally, elucidating the reciprocal synergies between the diagnostic markers of SLI can provide better guidelines for assessment and clinical intervention or educational language programs.

**Method**

**Participant recruitment and testing**

Following ethical and administrative consent, children were recruited from middle or middle-low socio-economic status public schools in Castelló, in the Valencia region of Spain. School instruction was provided in Spanish language, but the students also understood Catalan, according to their teachers and a parent questionnaire. In the Valencia region, there is diglossic bilingualism, with Spanish as the formal language and Catalan as a non-formal language, so that many children are proficient in the Spanish language and understand Catalan (two close Romance languages). Twenty children with SLI (8;0–9;11) and 20 age-matched children with TLD (8;1–10;3) participated in the present study. The mean age for the group with SLI was 8;9 (SD = 8.47 months) and for the group with TLD it was M = 9;1 (SD = 7.07 months). The group of children with SLI included 15 boys and 5 girls. The group with TLD included 9 boys and 11 girls. Several previous clinical studies found higher rates of males with SLI at different ages (e.g., Cuperus, Vugs, Scheper, & Hendriks, 2014; Tomblin et al., 1997; Whitehouse, 2010).

All children passed an individual hearing screening in our sound booth at 20 dB (500, 1000, 2000, 3000, and 4000 Hz), before every testing session, following the American National Standards Institute (2004a, 2004b). Parents were individually interviewed to complete our parent questionnaire and a Social-Economic Status (SES) Scale (Hollingshead, 1975). The parent questionnaire was used to determine the family and child’s history for language deficits, the child’s primary language, whether the child had a history of neurological disorders, behavior characteristics of autism, and other developmental/medical information. None of the participants had a phonological or neurological disorder, or autistic symptoms.

Participants came mostly from middle SES homes, and they had a clear preference for mass media in Spanish. The average SES (based on the SES Scale, Hollingshead, 1975) for the 20 children with SLI was 25.70 (SD = 9.74, range from 10.5 to 52.5); for the 20 children with TLD it was 40.30 (SD = 13.98, range from 21 to 63.5). A SES score between 30 and 39 is considered to be middle class; 20–29 is lower-middle SES, 40–54 is upper-middle SES (the maximum score is 66). Thus, the two groups have an average SES close to middle class, though leaning towards lower-middle SES for the group with SLI and upper-middle class for the group with TLD.

All children performed within normal limits on the Test of Nonverbal Intelligence, which administration lasted around 20 minutes (TONI-2; Brown, Sherbenou, & Johnsen, 2000). For children with SLI, M = 101.20 and SD = 8.92 (ranging from 89 to 118), and for children with TLD, M = 106.65 and SD = 10.46 (ranging from 89 to 120).
The two language status groups (SLI/TLD) did not differ significantly in age (months), $F(1, 38) = 2.50, p = .12$. The groups did not differ significantly in their performance IQ, $F(1, 38) = 3.14, p = .08$. Thus, the age and IQ variables were not considered in any of the subsequent analyses.

We also administered the following language tests in Spanish to each child: Peabody Picture Vocabulary Test (PPVT-III; Dunn & Dunn, 2006); Token Test for Children (TTFC-2; McGhee, Ehrler, & DiSimoni, 2007); Test de Comprensión de Estructuras Gramaticales (CEG; Mendoza, Carballo, Muñoz, & Fresneda, 2005); Vocabulary subtest from the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2007); and four subtests from the Illinois Test of Psycholinguistic Abilities (ITPA; Kirk, McCarthy, & Kirk, 2001). All of these tests have norms from Spain, except for the TTFC-2 test, which was translated into Spanish but the norms were from the USA. The administration of the oral language tests/subtests lasted for around two hours.

Receptive and productive vocabulary was mostly measured through the PPVT-III and the Vocabulary subtest of WISC-IV (a word definition task), respectively. For the latter, the raw score was converted into a $z$-score, as we did for the ITPA (see below). The PPVT-III, TTFC-2 (a test of auditory orders about tokens), and CEG (in which the child points to the picture that matches a particular oral sentence) scores are given in percentiles.

The four contemporary subtests we selected from the ITPA were: (a) Auditory Comprehension, the child listens to brief stories and responds to questions about them by pointing to pictures; (b) Auditory Association, the child completes sentences spoken by the examiner (e.g., ‘The father is big, the child is… small.’); (c) Verbal Expression, a lexical fluency task in which the child says as many items in a stated category as possible in one minute (e.g., animals); and (d) Grammatical Integration, the child completes oral sentences spoken by the examiner according to related pictures (e.g., ‘This man is painting. He is a… painter.’). Each ITPA subtest raw score was converted into a $z$-score (i.e., standard deviations in relation to mean scores from the norms), which was calculated with respect to the $M$ and $SD$ from the corresponding age norms [(raw score – $M$) / $SD$].

The diagnosis criteria for SLI were based on the oral language tests/subtests (PPVT-III, TTFC-2, CEG, the Vocabulary subtest from WISC-IV, and the four selected ITPA subtests). All children with SLI scored at least $\leq -1$ SD ($z$-score) or $\leq 16$th percentile, in at least two out of the eight Spanish language standardized subtests/tests. Sixteen children with SLI out of 20 qualified because of their significant low scores in at least three of the eight referred language tests/subtests. Furthermore, 19 children with SLI scored at/below the referred cutoff on at least one out of the four expressive language standardized subtests. All children with SLI were receiving intervention at school or through private services. All children with TLD scored within normal limits for all these oral language tests/subtests; no child with TLD had any language test/subtest score close to/below the cutoff scores of $-1$ SD or 16th percentile. For each group, we calculated a mean and standard deviation of either the $z$-scores or the percentiles for each of the tests/subtests (see Table 1). Additionally, 13 children with SLI had also Reading Disabilities (RD). Their scores in the PROLEC-R subtests (Cuetos, Rodríguez, Ruano, & Arribas, 2007) were below $-1$ SD on at least three out of the nine reading subtests. Five other children with SLI did not have any RD; two children with SLI could not be tested
for RD. In a previous MRI study (Girbau et al., 2014) each child within the comorbid subgroup with SLI and RD had a significant low score on the NRT.

**Non-word repetition task**

The NRT followed Spanish phonotactic patterns (see Girbau & Schwartz, 2007, for more details). There were 20 pseudowords, four at each of five syllable lengths, i.e., one, two, three, four, and five syllables (e.g., /flín/, /múntiɾ/, /konsценbɾál/, /giɾenflónis/, /kleptasmaθóɾfun/). Sixty different medium-low frequency syllables were combined in different orders within the 20 non-words. They were selected from a sample of 1148 syllables produced by 6- to 10-year-old children (Justicia, Santiago, Palma, Huertas, & Gutiérrez, 1996) and 1156 syllables produced by 6- to 13-year-olds (Justicia, 1995), and from a list of more than 2500 Spanish syllables (Armario Toro, 2001). All non-words began with consonants and had no diphthongs; each syllable contained only one vowel. Twelve non-words included at least one cluster, which were distributed across the five syllable lengths. The stress in the non-words varied across four different syllable positions; only one syllable in each non-word was stressed.

The non-words and the instructions were digitally audio-recorded by the author using Cool Edit Pro (Syntrillium Software Corp., 2002), with an interstimulus interval after each item that was increased as the length of the non-word increased. This allowed the child’s repetition during the task administration. The average duration of a non-word at each length was: one-syllable non-words, $M = 445$ ms ($SD = 51$ ms); two-syllable, $M = 1015$ ms ($SD = 110$ ms); three-syllable, $M = 1332$ ms ($SD = 130$ ms); four-syllable, $M = 1621$ ms ($SD = 137$ ms); and five-syllable, $M = 1890$ ms ($SD = 143$ ms).

**Table 1.** Language tests percentiles/z-scores: means and (standard deviations) for children with TLD/SLI.

<table>
<thead>
<tr>
<th>Language tests</th>
<th>Children with TLD$^c$</th>
<th>Children with SLI$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-III test$^a$</td>
<td>69.55 (18.98)</td>
<td>13.56 (14.09)</td>
</tr>
<tr>
<td>TTFC-2 test$^a$</td>
<td>66.15 (23.29)</td>
<td>23.90 (20.26)</td>
</tr>
<tr>
<td>CEG test$^a$</td>
<td>65.80 (26.27)</td>
<td>11.85 (12.64)</td>
</tr>
<tr>
<td>WISC-IV: Vocabulary subtest$^b$</td>
<td>0.93 (0.67)</td>
<td>−0.77 (0.74)</td>
</tr>
<tr>
<td>ITPA test$^b$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditory Association</td>
<td>0.37 (0.64)</td>
<td>−1.39 (0.96)</td>
</tr>
<tr>
<td>Grammatical Integration</td>
<td>0.70 (0.47)</td>
<td>−1.10 (1.04)</td>
</tr>
<tr>
<td>Verbal Expression</td>
<td>0.88 (0.76)</td>
<td>−0.83 (0.89)</td>
</tr>
<tr>
<td>Auditory Comprehension</td>
<td>0.21 (0.30)</td>
<td>−0.58 (0.80)</td>
</tr>
</tbody>
</table>

Notes: $^a$ The scores for PPVT-III, TTFC-2, and CEG tests are given in percentiles. $^b$ The Vocabulary subtest from WISC-IV and the ITPA subtests scores are given in z-scores or SDs (in relation to mean scores from the norms). $^c n = 20$.

PPVT-III = Peabody Picture Vocabulary Test; TTFC-2 = Token Test for Children; CEG = Test de Comprensión de Estructuras Gramaticales [Test of Grammatical Structures Comprehension]; WISC-IV = Wechsler Intelligence Scale for Children; ITPA = Illinois Test of Psycholinguistic Abilities.
Procedure

The non-words were presented to the children individually via earphones through a computer. They had to repeat each item immediately after listening to it. Each non-word was presented only once. The entire task lasted approximately three minutes. Children were videotaped during the task performance. Children responded to all items. The 20 repetitions of non-words were transcribed for each participant. The reliability for segment transcription was 99.2% (percentage of agreement between two judges), on the basis of 10% of the sample.

The analysis focused on the accuracy for the NRT or number of correct non-words, i.e., number of repeated non-words with the same sequence of segments as the auditory model item. Thus, if the child repeated exactly the same non-word that he/she heard, a score of 1 was given to this particular item. Otherwise, a score of 0 was given (even if only one sound of the non-word was omitted/substituted or if only a new sound was wrongly added).

Results

Accuracy, i.e., percentage of correct non-words, was analyzed for the Spanish NRT. As expected, most of the errors occurred on the subset of non-words that were three, four, and five syllables in length. Both groups performed rather well in the one- and two-syllable non-words; no individual statistics are reported due to ceiling effects at these non-word lengths. Therefore, in addition to the total scores typically examined in the NRTs, we analyzed a combined score for three-, four-, and five-syllable non-words. We also examined the likelihood ratios of the NRT.

Children’s non-word repetition accuracy

We analyzed the overall accuracy of non-word repetition and also the accuracy for the subset of longer non-words. Percentages of correct non-words (overall and for the subset of 3-4-5 syllable non-words) were arc-sine transformed prior to ANOVA analyses. The two language status groups differed significantly in the percent of total number of non-words correct, $F(1, 38) = 32.14, p < .00005$, and in the 3-4-5 composite percent correct, $F(1, 38) = 12.15, p < .002$. Notably, the children with SLI made more errors in the task, especially in the subset of 3-4-5 syllable non-words, than the children with TLD (see Table 2).

Likelihood ratios

To determine the extent to which the Spanish NRT can accurately distinguish children with SLI and TLD, we calculated the likelihood ratios (Sackett, 1992; Sackett & Haynes, 2002). They are the likelihood that a given test result would be expected in a child with SLI compared to the likelihood that this same result would be expected in a child without SLI. The positive predictive value is the probability of having SLI following a positive test result, that is, the proportion of children identified as positive on the screening test
who actually have SLI. The negative predictive value is the probability that a child does not have SLI following a negative test result. A sensitive screening test provides a maximum proportion of true positives or children with SLI correctly identified, and a minimum of false negatives (i.e., children with SLI who are not detected by the test). A specific screening test provides a maximum proportion of true negatives or children with TLD (without SLI) correctly identified, and a minimum of false positives (i.e., children with TLD who are identified as having SLI). A good diagnostic test should display both acceptable sensitivity and specificity, so that performance on the NRT may distinguish children with SLI and TLD (see Lalkhen & McCluskey, 2008). According to the authors, most clinical tests fall short of this ideal.

We calculated the likelihood ratios for the 40 children, using the 3-4-5 syllable composite percentage (see Table 3). For children with TLD, if the percentage of correct non-words in the 3-4-5 subset is greater than 50.0%, the likelihood ratio is 0.00, with a negative predictive value of 100%. For children with SLI, if the percentage of correct non-words in the 3-4-5 subset is lower than or equal to 50.0%, the likelihood ratio is 6.67 with a positive predictive value of 86.96%. The sensitivity (1.00 or 100%) and the specificity (0.85 or 85.00%) of the Spanish NRT are good. Thus, these results indicate that the task is a good diagnostic test for initial identification of children at risk of SLI, with a slightly greater sensitivity than specificity on the basis of this sample. Children testing

Table 2. Mean and (standard deviation) accuracy for the total/subset Non-word Repetition Task: percentage of correct non-words for children with TLD/SLI.

<table>
<thead>
<tr>
<th></th>
<th>TLDa</th>
<th>SLIa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total non-words</td>
<td>76.50 (9.19)</td>
<td>53.75 (12.34)</td>
</tr>
<tr>
<td>3-4-5 syllable non-words</td>
<td>67.08 (11.30)</td>
<td>31.14 (17.09)</td>
</tr>
</tbody>
</table>

Note: The percentages for the total non-words are based on 20 items. The subset of 3-4-5 syllable non-words included 12 items.

Table 3. Likelihood ratios for the percentage of correct non-words from the subset of three-, four-, and five-syllable non-words in children with SLI and TLD.

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>SLIa Number of children</th>
<th>Proportion of children</th>
<th>TLDa Number of children</th>
<th>Proportion of children</th>
<th>Likelihood ratioa</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 50.0%</td>
<td>20 (20/20 = 1)</td>
<td>3 (3/20 = 0.15)</td>
<td>1/0.15 = 6.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 50.0%</td>
<td>0 (0/20 = 0)</td>
<td>17 (17/20 = 0.85)</td>
<td>0/0.85 = 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: a n = 20 children. b Likelihood ratios were calculated by dividing the proportion results for the group with SLI by the proportion for the group with TLD. c Number of children whose score in the 3-4-5 syllable composite percentage is either above or equal/lower than 50%. d Proportion of children out of 20 participants in the SLI/TLD groups whose score in the 3-4-5 syllable composite percentage is either above or equal/lower than 50%.
positive for possible SLI will need to receive a further comprehensive evaluation for the
diagnosis of SLI. The NRT can also be part of a language test battery.

Relation of non-word repetition accuracy to expressive and receptive
language abilities

We analyzed the relations between the non-word repetition scores and the standardized
expressive/receptive language tests/subtests scores. For all Pearson product moment cor-
relations (Table 4) we set the significant alpha level at \( p = .006 \) (i.e., .05/8). Bold font indicates significance following the Bonferroni correction.

Table 4. Pearson correlations between children’s language tests scores and the percentage of
correct non-words in the overall Non-word Repetition Task: overall sample and children with
SLI/TLD.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Overall samplea</th>
<th>TLDb</th>
<th>SLIb</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-III/NRT</td>
<td>.68**</td>
<td>.33</td>
<td>-.03</td>
</tr>
<tr>
<td>TTFC-2/NRT</td>
<td>.71**</td>
<td>.41</td>
<td>.40</td>
</tr>
<tr>
<td>CEG/NRT</td>
<td>.66**</td>
<td>.18</td>
<td>.21</td>
</tr>
<tr>
<td>WISC-IV: Vocabulary subtest/NRT</td>
<td>.65**</td>
<td>.07</td>
<td>.12</td>
</tr>
<tr>
<td>ITPA: Auditory Association subtest/NRT</td>
<td>.73**</td>
<td>.22</td>
<td>.51*</td>
</tr>
<tr>
<td>ITPA: Grammatical Integration subtest/NRT</td>
<td>.65**</td>
<td>.24</td>
<td>.19</td>
</tr>
<tr>
<td>ITPA: Verbal Expression subtest/NRT</td>
<td>.57**</td>
<td>.10</td>
<td>.08</td>
</tr>
<tr>
<td>ITPA: Auditory Comprehension subtest/NRT</td>
<td>.63**</td>
<td>-.06</td>
<td>.58*</td>
</tr>
</tbody>
</table>

Notes: *p < .05, **p < .0001. Using the Bonferroni correction, the significant \( p \)-value begins at \( p = .006 \) (i.e., .05/8). Bold font indicates significance following the Bonferroni correction.

| PPVT-III = Peabody Picture Vocabulary Test; TTFC-2 = Token Test for Children; CEG = Test de Compre-
| nsión de Estructuras Gramaticales; WISC-IV = Wechsler Intelligence Scale for Children; ITPA = Illinois Test of
| Psycholinguistic Abilities. |

Discussion

The present study adds to previous literature concerning the phonological working mem-
ory task and related psycholinguistic abilities in children with SLI. We provide additional
data for the NRT diagnostic accuracy in 40 elementary school children to test the risk for SLI, expanding research at older ages. We also expand the literature concerning the NRT and diagnostic markers for Spanish-speaking children with SLI.

**Children’s non-word repetition accuracy**

This study provides evidence that the Spanish NRT identifies language status (SLI/TLD) with a high degree of accuracy. Particularly, children with SLI obtained a significantly lower percent of total number of non-words correct than children with TLD (54% vs. 77%). This significant difference in the average accuracy of non-word repetition was more than twice as large for the 3-4-5 composite percent correct (31% vs. 67%). Thus, the difference between the two groups (SLI/TLD) was greater in longer non-words, beginning with the three-syllable length. In sum, although significant differences between the two language groups were found in both non-word sets of the NRT, the subset of 3-4-5 syllable non-words identifies even better 8- to 10-year-old children with SLI.

Our results agree with previous studies from Spain ($n = 22$) and New York City ($n = 22$), which also found this length effect and poorer performance in children with SLI (Girbau & Schwartz, 2007, 2008). These findings on the language status group difference and the length effect beginning with the three-syllable non-words have been reported also in English-speaking children with SLI/LI and TLD (e.g., Dollaghan & Campbell, 1998; Rodekohr & Haynes, 2001) and in smaller samples for other languages, including French (Le Foll et al., 1995) and Italian (Bortolini et al., 2006; Dispaldro et al., 2013).

**Likelihood ratios for the Spanish Non-word Repetition Task**

The present study found good likelihood ratios for the NRT, including good sensitivity and specificity, as previous studies found in smaller samples (Girbau & Schwartz, 2007, 2008). This indicates that the Spanish NRT can be a valuable and reliable diagnostic test of SLI in Spanish-speaking children, and may be used together with other tests. The percentage of correct non-words identifies SLI with good sensitivity and specificity in elementary school children. Particularly, a score lower or equal to 50% correct in the 3-4-5 syllable non-words indicates that it is significantly likely that the child has SLI. Conversely, if this score is greater than 50%, it is significantly likely that this diagnosis can be ruled out; no child with SLI from this sample scored above 50% in the NRT. If we use the NRT, additional testing can be done in children screening positive in the NRT, in order to determine the diagnosis of SLI. For example, this NRT was used within a language test battery to identify elementary school children with SLI in an MRI study (Girbau et al., 2014). Particularly, the diagnostic criteria for diagnosing a child with SLI were based on having a score $\leq 50\%$ in the Spanish NRT subset and $< 16$th percentile in at least two of the Spanish oral language standardized subtests/tests we used here. Girbau et al. also found significant brain volumetric differences in children with SLI, including lower gray matter volume at the right postcentral parietal gyrus and greater cerebrospinal fluid volume, when compared to children with TLD. These brain markers support the valuable diagnostic use of our language test battery, including the NRT, to diagnose Spanish-speaking children with SLI.
The use of this NRT as a diagnostic marker on the basis of the percent of non-words correct involves a simple scoring procedure. Thus, complex segment-by-segment transcription is not required to use this NRT as a diagnostic marker. Some previous studies administering the English NRT used the percentages of correct total phonemes as cutoffs; for example, above 81% for preschool children with TLD and below 70% for SLI (Dollaghan & Campbell, 1998). Another study found that the percentage of whole non-words correct with a 65% cutoff identified differences between preschool children with SLI and TLD better than the percentage of phonemes correct (Dispaldro et al., 2013).

Scientists are improving the approach to the diagnosis of children with SLI. Early studies recruited these participants on the basis of receiving language intervention (Dollaghan & Campbell, 1998), but nowadays they are identified more according to individual cutoffs for the standardized language tests/subtests scores, as these measures become more accurate. In the future, treatment programs for children with SLI may also benefit from including phonological working memory tasks with non-words of increasing length.

Relation of non-word repetition accuracy to expressive and receptive language abilities

The overall Spanish NRT accuracy correlated significantly with each of the eight oral expressive and receptive language tests/subtests scores in the overall sample of 40 children. Children’s language status (SLI/TLD) did not seem to be mediating any of the eight positive correlations, though significance might be reached with bigger language group sizes. The more recent previous studies analyze the two groups together. Nevertheless, including the correlational data for each language group separately would be helpful for further comparisons with other investigations.

All correlations (for the full sample) were positive: the higher the non-word repetition accuracy the higher the score in each language test/subtest. The associations were moderately strong to strong, indicating the extent to which production and comprehension language processes are interrelated through phonological working memory ability in 8-to 10-year-old children with/without SLI.

Relation of non-word repetition accuracy to expressive language abilities. The four expressive language subtests, i.e., the Vocabulary subtest from WISC-IV, and three ITPA subtests (Verbal Expression, Auditory Association, and Grammatical Integration) correlated significantly with NRT accuracy. A previous study in typical 4-year-old children found that a better non-word repetition performance was related with the production of a wider repertoire of words, and longer and syntactically more varied utterances (Adams & Gathercole, 2000). We estimate that this may also be somehow associated with the length effect beginning with the three-syllable non-words that is usually found in the NRT. Furthermore, Grammatical Integration and Auditory Association subtests involve some auditory working memory demands similar to the NRT. Both subtests request the child’s oral completion of a sentence presented orally by the examiner with or without visual support respectively. These strong positive correlations with the two subtests were also found in two additional samples from Spain/NYC, each one including 22 children with SLI and TLD together (Girbau & Schwartz, 2007, 2008).
some morphological knowledge demands involving the proper use of suffixes in completing the sentence. A previous study found that Icelandic non-word repetition performance correlated with morphological accuracy in spontaneous language production for school children with SLI and TLD together (Thordardottir, 2008).

Concerning the expressive vocabulary, significant correlations of the NRT with the Vocabulary subtest (a word definition task) and the Verbal Expression subtest (production of words about particular categories) are noteworthy. Previous studies found that lower non-word repetition accuracy was related to smaller vocabulary in younger children with TLD (Adams & Gathercole, 2000; Edwards et al., 2004). In addition, a large vocabulary in preschool children with/without phonological disorders was also related to higher accuracy in repeating low-frequency non-words (Munson et al., 2005). In the present study, syllables with low and medium frequency were combined within a non-word, so that we found a similar association in older children. These NRT features were also included in the previously referred studies with two smaller samples, in which the correlation with Verbal Expression failed to reach significance (Girbau & Schwartz, 2007, 2008). In general, children seem to benefit from a larger/stronger semantic network to perform the NRT. A theoretical review supported that the abilities to repeat non-words and to learn new words with new phonological forms are closely linked, since both require phonological short-term store, in typical individuals and those with disorders of language learning (Gathercole, 2006). However, in a previous study, 28 older children with SLI (M = 10;11) who had the lowest and the highest non-word repetition performance did not differ significantly in their expressive/receptive vocabulary scores, although an association was found (Botting & Conti-Ramsden, 2001).

Relation of non-word repetition accuracy to receptive language abilities. The overall percentage of correctly repeated non-words correlated significantly with the three receptive language tests (PPVT-III, TTFC-2, CEG), and the Auditory Comprehension subtest from ITPA. A recent study in English concluded that receptive vocabulary (PPVT-IV) and phonological short-term memory (NRT), in 5-year-old children with SLI and TLD together, were significant predictors of fast mapping, which is the ability to map phonological and semantic information in word learning (Jackson et al., 2015). Previous research has also supported the significant correlation between NRT measures and PPVT in preschoolers with LI (Sahlén et al., 1999) and typical toddlers (Brandeker & Thordardottir, 2015).

As for the TTFC-2 test, it is strongly dependent on working memory and speed of processing, because the examiner can only read once each auditory linguistic command about tokens, which the child must identify/manipulate immediately. The CEG test (a grammar understanding test based on the TROG test) also demands working memory abilities, since the examiner can only say the oral sentence once, after which the child needs to point to a matching picture among four drawings. Similar to the present study, work in SLI (Botting & Conti-Ramsden, 2001) found a significant difference, for TROG scores, between the lowest and the highest non-word repetition performance groups. It is also noteworthy that the strongest correlation in the Swedish NRT study, in children with LI of around 5 years of age, was between percentage consonants correct on the NRT and a non-standardized translation of TROG (Sahlén et al., 1999). However, comprehension
of long sentences in another non-standardized picture-pointing task did not correlate significantly with an English NRT for either 8-year-old children with SLI or TLD separately (Montgomery, 2004). The smaller sample size of 12 children in each group may be behind the non-significant correlation; data on both groups together were not reported.

Finally, the Auditory Comprehension subtest from the ITPA clearly involves immediate auditory verbal memory. In this subtest, the child answered to questions by pointing to pictures after listening to a short story once. The present significant result agrees with the significant correlation that was found between a non-standardized oral comprehension task of fables and percentage consonants correct on the Swedish NRT in preschool children with LI (Sahlén et al., 1999). The correlational result in the ITPA subtest was also significant for the Spanish NRT in the overall sample from NYC (Girbau & Schwartz, 2008), though it failed to reach significance in the overall group from Spain (Girbau & Schwartz, 2007). The present study based on a larger sample confirms the significant correlation. Indeed, auditory verbal working memory processes are needed for auditory story comprehension.

Clinical implications. The present correlational outcomes help to understand to what extent phonological working memory abilities need to be targeted in any individualized speech-language therapy program to ameliorate SLI. Moreover, on the basis of the current and previous significant correlational findings across languages, children may also benefit from expanding receptive/productive vocabulary, lexical fluency, oral completion of sentences including morphology markers, and auditory comprehension of commands, stories, and sentences with different length/syntax complexity. Obviously, some visual support may also be effective to achieve these clinical/educational goals.

The fact that all the present correlations were positive could mean that expanding the phonological working memory may lead to improving any of the language abilities examined here, and vice versa. Further research is needed on this point. That said, more intensive speech-language therapy at the earliest age possible is recommended in children with poor non-word repetition accuracy. Otherwise, we estimate that this deficit may hinder the rehabilitation process and even trigger some additional deficits, as it probably happened, in the present sample, to more than 50% of children with SLI, who exhibited comorbidity with RD. Prevention through educational programs targeting these abilities is crucial.

Conclusions

To sum up, the Spanish Non-word Repetition Task appears to be an accurate diagnostic marker for identifying children with Specific Language Impairment from 8 to 10 years of age. The present study provides likelihood ratios with good sensitivity and specificity for a 50% correct non-word repetition cutoff in the preselected groups. We also found a length effect, similar to previous findings in several languages. With a set of norms, the Spanish NRT can be an accurate screening test as part of a comprehensive evaluation for the diagnosis of SLI.

Future research on this topic would need to report correlations for the overall sample of children with SLI/TLD together and also separately, to avoid lack of power and for
further comparisons across investigations respectively. In the present study, the better the phonological working memory abilities (as measured by non-word repetition) were, the better the diverse contemporary expressive and receptive language abilities were. Particularly, this significant association was evident in relation to lexicon/word definition; lexical fluency with limited time; oral completion of sentences; and auditory comprehension of words, stories, commands, and sentences with different length/syntax complexity. Thus, early intensive educational and clinical language programs may need to focus on expanding phonological working memory abilities with NRTs, and on the cited expressive/receptive language abilities, to strengthen and improve this reciprocal synergy.

Although a phonological working memory deficit is not sufficient to diagnose SLI, it is arguably a necessary core deficit to diagnose SLI in children. Previous studies strongly suggest its high genetic heritability (e.g., Bishop et al., 1996; Girbau & Schwartz, 2008). Thus, further NRTs across languages with good sensitivity/specificity and appropriate items need to be developed. Future research is also needed to improve our knowledge of how rehabilitation programs including phonological working memory tasks, similar to the NRT with items of increasing length, may improve working memory and other related language skills in children with SLI.

Acknowledgements

I would like to thank R. G. Schwartz for his generous and valuable advice in planning the study and data analyses.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partly funded by a grant from Spain, Instituto de Salud Carlos III – Ministerio de Sanidad y Consumo, PI041733, D. Girbau, P.I.

References


Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript, Department of Sociology, Yale University, New Haven, CT.


