Published in final edited form as:

J Speech Lang Hear Res. 2013 February ; 56(1): 323–336. doi:10.1044/1092-4388(2012/11-0304).

Real-Word and Nonword Repetition in Italian-Speaking Children with Specific Language Impairment: A Study of Diagnostic Accuracy

Marco Dispaldro,

Università di Padova, Padova, Italy

Laurence B. Leonard, and

Purdue University, West Lafayette, IN

Patricia Deevy

Purdue University, West Lafayette, IN

Abstract

Purpose: Using two different scoring methods, we examined the diagnostic accuracy of both real-word and nonword repetition in identifying Italian-speaking children with and without specific language impairment (SLI).

Method: A total of 34 children aged 3;11 to 5;8 participated – 17 children with SLI and 17 typically developing children matched for age (TD-A children). Children completed real-word and nonword repetition tasks. The capacity of real-word and nonword repetition tasks to discriminate children with SLI from TD-A was examined through binary logistic regression and response operating characteristics curves.

Results: Both real-word and nonword repetition showed good (or excellent) sensitivity and specificity in distinguishing children with SLI from their typically developing peers.

Conclusions: Nonword repetition appears to be a useful diagnostic indicator for Italian, as in other languages. In addition, real-word repetition also holds promise. The contributions of each type of measure are discussed.

Keywords

Specific Language Impairment; Nonword Repetition; Real-Word Repetition; Clinical Marker

Introduction

Children with specific language impairment (SLI) show a significant deficit in language ability, yet score within average levels on nonverbal intelligence tests, exhibit normal hearing, and show no evidence of frank neurological impairment (Leonard, 1998). Epidemiological evidence indicates that SLI is present in approximately 7% of the kindergarten population (Tomblin et al., 1997). For many children with SLI, the disorder appears to have a genetic basis; twin studies show that the concordance rates for children with SLI and their identical twins are substantially higher than those seen for children with SLI and their same-sex dizygotic twins (Bishop et al., 1995). The phenotype seen in these

children is not uniform; some areas within language tend to be more adversely affected than other areas. For example, the area of morphosyntax is often more impaired than the areas of vocabulary and phonology. Moreover, often children's comprehension of language is more advanced than their production abilities.

Nonword Repetition

One especially revealing task used in SLI research is the nonword repetition task. In this task, children hear a nonsense word and repeat it immediately after. Many children with SLI show an extraordinary deficit in nonword repetition, especially as the length of the nonword is increased in number of syllables. This finding was first reported by Kamhi and colleagues (1986, 1988) and subsequently examined in great depth by Gathercole and Baddeley (1990). In these studies, children with SLI may lag behind same-age peers in the repetition of one- and two-syllable nonwords, but the gap between these groups becomes much larger when the nonwords are three and four syllables in length. This effect has been replicated in other languages, such as Swedish (Sahlén et al., 1999) and Spanish (Girbau & Schwartz 2008), though not for Cantonese (Stokes et al., 2006). In a study on Italian, Bortolini and colleagues (2006) found that children with SLI were less accurate than control children, but the effect of nonword length was similar for the two groups.

The fact that group differences are seen even at the one-syllable level (Graf Estes et al., 2007) suggests that children with SLI may have some difficulties with the discrimination, encoding, and/or production demands of the task. However, it seems clear that phonological short-term memory (PSTM) plays a major role in the nonword repetition findings (Gathercole, 2006). According to Baddeley (1986), working memory is a multi-component, capacity-limited system that comprises a controlling "central executive" as well as an articulatory loop system and visuo-spatial sketchpad. The central executive is thought to regulate information flow within working memory, retrieval of information from other memory systems, and the processing and storage of information. The articulatory loop includes a capacity-limited PSTM and an articulatory control process that acts to refresh and maintain speech material in the store for a brief period. The articulatory loop's function is to store verbal input temporarily, either novel phonological input or known words, while other cognitive tasks such as auditory comprehension take place. When temporary storage of known words is required, PSTM is supported by knowledge stored in long term memory. However for nonwords, it is PSTM that is primarily responsible for storage (Gathercole et al., 1994); crucially, this assumes that lexical characteristics of the words, such as neighborhood density (Dollaghan et al., 1993; Gathercole, 1995; Metsala & Chisholm, 2010) and phonotactic frequency (Coady et al., 2010; Edwards et al., 2004; Munson et al., 2005) are controlled. Impairments in the ability to temporarily store phonological information in working memory result in the relatively poor nonword repetition in children with SLI. As a consequence, the low scores in nonword repetition can be properly viewed as a PSTM deficit that limits the language acquisition of children with SLI (e.g., Gathercole & Baddeley 1990).

Nonword Repetition as a Clinical Marker

Nonword repetition tasks have received much attention in investigations of SLI because of their potential to identify children with language impairment. This is due in large part to the reliance of nonword repetition on language processing rather than accumulated language knowledge (Campbell et al., 1997); moreover, nonwords can be created to take into account the phonotactic properties of any language or dialect (Oetting & Cleveland, 2006).

In order to be considered a good clinical marker, performance on a task must be accurate in distinguishing children with SLI from their typically developing (TD) peers – they must

display both acceptable ‘sensitivity’ and ‘specificity’. ‘Sensitivity’ refers to the accuracy of a test in detecting individuals with the disorder, whereas ‘specificity’ is the accuracy in identifying children without the disorder. Sensitivity and specificity values of at least 80% are regarded as acceptable (Meisels, 1988), whereas levels of at least 90% are regarded as good (Plante & Vance, 1994).

Findings regarding the sensitivity and specificity of nonword repetition tasks in distinguishing children with SLI from TD children are somewhat mixed. Promising results have been found for preschool children (Deevy et al., 2010; Gray, 2003). Recently, Redmond et al., (2011) showed both sensitivity and specificity above 90% in distinguishing seven-year-old children with and without language impairment. However, some studies report acceptable values for specificity but not for sensitivity both in preschool children (Conti-Ramsden, 2003; Conti-Ramsden & Hesketh, 2003), and in six- and seven-year-old children (Conti-Ramsden et al., 2001; Oetting & Cleveland, 2006; Oetting et al., 2008). Archibald and Joanisse (2009) reported relatively low values (below 70%) for both sensitivity and specificity in school-age children.

Research on the effectiveness of nonword repetition as a potential diagnostic tool in other languages is limited. In a search for clinical markers of SLI in Cantonese, Stokes et al. (2006) did not find significant differences in repetition performance between preschoolers with SLI and age-matched TD children. Bortolini et al. (2006) found acceptable sensitivity (82%) and specificity (82%) using a nonword repetition task with Italian-speaking preschoolers. Dispaldro and colleagues (2009, 2011) have examined both nonword and real-word repetition in typically developing three- and four-year-old Italian-speaking children. Interestingly, they found that real-word repetition predicted grammatical ability better than nonword repetition in children with typical language development. It is currently an open question whether real-word repetition could be as effective, or more effective, than nonword repetition as a clinical marker for children with SLI learning Italian.

Real-Word Repetition

Repeating real words involves activation of the phonological form of a lexical representation in long-term memory; this lexical representation reflects not just phonological knowledge, but also semantic knowledge. The effect of long-term knowledge on the temporary storage of real words has been confirmed by the fact that overall accuracy is higher in repetition of real words than in repetition of nonwords in English (Chiat & Roy, 2007; Metsala & Chisholm, 2010; Roy & Chiat, 2004), Italian (Casalini et al., 2007; Dispaldro et al., 2009, 2011) and Swedish (Sahlén et al., 1999).

Word length may affect real-word repetition accuracy differently in typical and clinical groups. In children with typical development, real-word repetition is less affected by length than nonword repetition (Chiat & Roy, 2007; Dispaldro et al., 2009, 2011; Metsala & Chisholm, 2010; Roy & Chiat, 2004). However, Chiat and Roy (2007) compared young typically developing and language delayed children on a real-word and nonword repetition task and found differences between the groups in the effect of word status on their repetition performance. While both groups showed the familiar effects of length, the interaction between length and word status was not significant for the language delayed group. This seems to suggest that although real-word repetition places fewer demands on PSTM than nonword repetition, these demands may be enough to affect the performance of children with SLI. In addition, fewer children in the clinical sample showed any difference in performance based on word status; in other words, real word status did not provide an advantage for repetition performance for this group. Although the differences were subtle, the authors concluded that there was a “greater vulnerability of real words in the clinical sample compared to the typical sample”.

To our knowledge, no studies have thus far examined the utility of real-word repetition, either alone or in combination with nonword repetition, as a clinical marker. The first aim of this study is to evaluate the potential of real-word repetition as a clinical marker of SLI in Italian. We will also test the diagnostic accuracy of a related set of nonwords.

Comparison of Scoring Methods

As described above, several studies using different nonword repetition tasks (e.g., the Preschool Repetition test (PSRep), Chiat & Roy, 2007; the Children's Test of Nonword Repetition (CNRep), Gathercole et al., 1994; the Nonword Repetition Test (NRT), Dollaghan & Campbell, 1998), have shown that children with SLI repeat nonwords significantly less accurately than children with TD. However, in a meta-analysis of 23 nonword repetition studies, Graf Estes et al. (2007) found that the magnitude of this difference varied across studies. Children with SLI performed between 0.62 and 4.34 standard deviations below children with TD. Using a moderator analysis, they showed that the magnitude of the nonword repetition deficit in children with SLI was significantly associated with the type of task used to assess it. For example, the CNRep task yielded a larger effect size ($d = 1.94$) than the NRT ($d = 0.90$). The different effect sizes associated with the various nonword repetition task may have reflected test characteristics, such as syllable length, articulatory complexity, and wordlikeness. A fourth possible source of difference could have been the scoring method. The CNRep is typically scored using the whole-word method; that is, each repeated item was evaluated as a whole and scored as either entirely correct or as an incorrect production of the target, without regard to the number of phonological errors. In contrast, the NRT is typically scored as the percentage of phonemes produced correctly. If children with SLI have even a few more phonological errors than children with TD in repeating nonwords, the whole-word method could penalize them disproportionately compared to scoring percentage of phonemes correct. To investigate this hypothesis, Graf Estes et al. (2007) examined the NRT performance of children with SLI and with TD in which responses were scored using both methods. Contrary to their hypothesis, the results showed that the difference between groups in the whole-word method was smaller ($d = .48$) than in the percentage of phonemes correct method ($d = 1.17$) (however, as the authors noted, this result is due in part to the larger standard deviation in the whole-word method).

A second study seems to suggest that different scoring methods should not contribute to different results. Deevy et al. (2010) hypothesized that differences in articulatory accuracy could contribute to group differences, given that preschoolers with SLI are at higher risk for speech delay. They used two scoring methods; the first treated out-of-inventory phonemes as errors and the second made allowances for these errors by treating them as unscorable. They found that group differences in accuracy of nonword repetition ($d = 2.26$ vs. 2.12) and diagnostic accuracy (Sensitivity = 86% (25/29) vs. 79% (23/29); Specificity = 91% (43/47) vs. 89% (41/47)) were comparable for the two scoring methods.

While these results seem to be concordant, further investigation into the impact of different scoring methods is needed. To our knowledge, no studies have thus far examined whether scoring with the "whole-word" method or the "percentage of phonemes correct with allowances for developmental phonological errors" method produce (i) differences between real-word and nonword repetition scores, (ii) differences between the scores of children with SLI and TD children and (iii) differences in diagnostic accuracy.

In summary, the present study had four aims designed to inform current practice with Italian-speaking children. First, we wanted to determine whether repetition performance would be better for real words than for nonwords. Our second aim was to look for group differences: we expected that the children in the SLI group would perform at lower levels

than children in the TD-A group. Assuming that both of these expectations were confirmed, the third aim of this experiment was to use diagnostic accuracy measures to determine whether real-word repetition performance, like nonword repetition performance, might have potential as a clinical marker of SLI in Italian. The last aim was to determine if all these results change as a function of the kind of scoring method used.

Method

Participants

Two groups participated in this experiment (SLI and TD-A). The total sample consisted of 34 monolingual Italian-speaking children between the ages of 3;11 and 5;8. Because gender and SES do not appear to influence children's performance on repetition tasks (Chiat & Roy, 2007; Ellis Weismer et al., 2000), the two groups were not matched on these factors.

Seventeen children with SLI were included in this study. The criteria for this selection were twofold: intervention status and performance on a set of probes testing expressive morphosyntax (Bortolini et al., 2002, 2006). Regarding the first criterion, each child had qualified for language intervention services on the basis of the evaluation done by language therapists in one of two Health Services in the North East of Italy. All 17 children had been diagnosed as SLI with an expressive or receptive-expressive deficit. Each child met the criteria for SLI (Leonard, 1998) and passed a pure-tone hearing screening bilaterally (20 dB HL) at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz (American Speech-Language-Hearing Association, 1997). They ranged in age from 3;11 to 5;8 (mean age 4;9). Children in this group showed typical cognitive abilities on a psychological assessment; performance IQ was evaluated using the Italian version of the Wechsler Preschool and Primary Scale of Intelligence-III (Fancello & Cianchetti, 2008). None of these children had a physical impairment, global developmental delay, neurological dysfunctions, emotional problems, nor had any suffered environmental deprivation. These children had not yet received therapy at the time of testing although all had qualified for enrollment in speech-language services through the Health Service.

The clinicians had judged the children eligible for services on the basis of clinical judgment and the results from one or more of several language tests used in Italy. These tests (described below) were used to identify below age-level receptive and expressive language skills in the areas of vocabulary and grammar. For receptive vocabulary, the tests were the Italian version of the PPVT-R (Stella et al., 2000) or Test di Valutazione del Linguaggio (Cianchetti & Fancello, 2003). The tests used for receptive morphosyntax were Prove di Valutazione della Comprensione Linguistica (Rustioni, 1994) or Test di Comprensione Grammaticale per Bambini (Chilosi & Cipriani, 1995). For expressive vocabulary the Test di Valutazione del Linguaggio (Cianchetti & Fancello, 2003) was used. Finally, expressive morphosyntax was evaluated by therapists through analysis of the children's spontaneous language samples. Spontaneous language samples consisted of at least 100 utterances obtained when the children were looking at a picture book while interacting with the therapist. The children were encouraged to initiate topics and determine the direction of conversation. To maintain conversation, the examiner asked questions or commented. The sample was transcribed using the CHAT transcription format (MacWhinney, 2000). Based on this language sample the therapist estimated expressive morphosyntax as "adequate," "mildly inadequate" or "highly inadequate" based on mean utterance length and the presence of grammatical errors. As a group, the children's MLUs in words averaged 2.75 ($SD = 0.58$; range = 1.74 - 3.61). Based on Italian normative data (Cipriani et al., 1993), children with typical language reach this MLU range at about three years of age. As can be seen from Table 1, clinicians did not employ a uniform set of tests to evaluate the children. Based on these tests, all 17 children exhibited deficits in expressive morphosyntax; for 10 of

these children, deficits in receptive morphosyntax were seen as well. All but one of the children showed an additional weakness in receptive and/or expressive vocabulary. According to the ICD-10 system (World Health Organization, 2005), eight of the 17 children in the group met the criteria for “expressive language disorder” (F80.1) and the remaining nine met the criteria for “receptive language disorder” (F80.2) (the latter term includes combined expressive and receptive deficits).

To promote uniformity in the inclusion process, the second criterion for inclusion in the SLI group was a low score on a task designed to assess children’s use of third person direct object clitic pronouns (the clitic pronouns *lo* “it-MASC SING/him”, *la* “it-FEM SING/her”, *li* “them-MASC PLUR”, and *le* “them-FEM SING”). In this task, for each item, the experimenter described the first of two drawings and prompted the child to describe the second. For example, one item designed to elicit use of the masculine singular clitic *lo* involved the prompt: *Qui la bambina compra il gelato, e qui... (lo mangia)* “Here the girl buys the ice cream, and here... ([she] it eats = she eats it)”. This task has shown good sensitivity and specificity in distinguishing children with SLI from typically developing children in the age range studied here. Sensitivity has ranged from 86.67% to 90.91% and specificity has ranged from 93.33% to 100% (Bortolini et al., 2002, 2006). We employed the cut-off score of 80% correct use of direct object clitic pronouns for this task, corresponding to 2 SDs below the mean. Although one child included in the SLI group approached this cut-off level (with a score of 79%), the group means and standard deviations for the SLI group were extremely similar to those reported in the original Bortolini et al. studies. In the present study, the SLI mean and SD were 29.59% and 25.55%, respectively. In Bortolini et al. (2002) the corresponding mean and SD were 26% and 34.97%; in Bortolini et al. (2006) the mean and standard deviation were 35.27% and 38.29%, respectively. The scores of each of the children in the SLI group on this task are shown in Table 1.

The remaining 17 children served as typically developing age controls (TD-A). They ranged in age from 4;1 to 5;7 (mean age 4;9). Each child in this group was within two months in age of a child in the SLI group; as a result, there was no difference in mean age of the two groups, $t(32) = -.37$, $p = .77$. The TD-A children were recruited from preschools in Padua, Italy. They were not included in the study if they had any language, articulatory, neurological, or psychiatric deficit according to parent or teacher report. All passed a pure-tone hearing screening bilaterally (20 dB HL) at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz.

The direct object clitic pronoun task was also administered to all children with typical development (Bortolini et al., 2002, 2006), using the same cut-off score of 80%. All TD-A children scored above this level ($M = 92\%$, $SD = 6$, range = 82-100) and, not surprisingly, were significantly more accurate than the children with SLI on this measure, $t(32) = 9.38$, $p < .001$, $d = 10$. The TD-A children’s mean percentage correct was similar to those found by Bortolini et al. who reported a mean of 96.40% and a SD of 7.77% in their 2002 study and a mean of 92.18% and a SD of 6.40 in their 2006 investigation.

All research procedures were conducted according to the guidelines of the University of Padua for the protection of human participants. Parental consent was obtained for each child before inclusion in the study.

Materials and Procedure

The repetition tasks were based on those used by Dispaldro et al. (2009, 2011) and included a list of real words and a list of nonwords. The original tasks each had eight two-syllable and eight three-syllable words. For the revised tasks, eight four-syllable words were added to each, for a total of 24 real words and 24 nonwords (see Appendix A).

The real-word repetition task included words that were assumed to be known by preschool children, based on norms reported in Barca et al. (2002). These words had simple and complex syllabic structure (CV and CVC or CCV), consistent with the articulatory and phonotactic characteristics of Italian. They had primary stress on the penultimate syllable, the most frequent stress pattern in Italian.

Items in the nonword repetition task were constructed to be similar in phonemic structure to those in the real-word task. Each nonword was created from a real word by using the same initial phoneme and, wherever possible, replacing the remaining phonemes with ones similar in sonority and/or manner of articulation. The two sets of stimuli were matched for length in phonemes and in syllables, initial phoneme, syllabic structure, low phonological neighborhood size, and low phonotactic probability.

We created stimuli which would be easy to repeat for young children, inasmuch as the phonemic and syllabic characteristics of the nonwords are usually mastered by Italian-speaking children as young as three (Zmarich & Bonifacio, 2004).

The children were tested individually in a quiet room, with only the examiner present. The tasks were presented to each child in random order. All children's responses were recorded on a computer using a Sony ECM CZ-10 microphone and Audacity software. Responses were transcribed after the experimental session.

Scoring and Reliability

The children's responses were transcribed using the International Phonetic Alphabet (IPA) system by the first author, a native speaker of Italian. Two different scoring methods were used.

Scoring Method 1: Percentage of phonemes correct with allowances for developmental phonological errors—A “correct production” of a word has conventionally been defined as producing the phonemes in the same order as the target with no omissions or substitutions (additions and distortions are allowed) (Dollaghan & Campbell, 1998). However, when using the repetition task with preschoolers, many researchers have adapted these scoring conventions to take into account articulation ability (Chiat & Roy, 2004; Conti-Ramsden, 2003; Deevy et al., 2010; Gray, 2003; Thal et al., 2005). For example, developmental errors have been treated as correct or as unscorable. Minimizing the impact of developmental phonological errors on scores allows us to assess more accurately the contributions of PSTM to repetition performance. As a result, we can be more confident that group differences are not driven by differences in articulation ability and that diagnostic accuracy results are not complicated by individual differences in this ability. Accordingly, we examined the responses of the children and scored as correct any substitution that reflected a developmental phonological error in Italian, following Bortolini (1995); other substitutions and omissions were scored as errors. Additions or distortions of phonemes were scored as correct. This scoring procedure represented a highly conservative approach, because it narrowed the differences between the two groups, owing to the fact that the TD-A children had relatively few developmental errors (and these were limited to difficulties with [r]). Following the scoring of all responses, the number of correctly repeated phonemes was counted and the percentage calculated for each child. This method was used with the children in both groups.

The audiorecordings of five children in each group (30% of participants) were randomly selected for transcription by a second trained listener. Phoneme-by-phoneme percentage of agreement for judgments of correctness was 95% (range from 90% to 100%).

Scoring Method 2: Percentage of whole-words correct, with no allowances for developmental phonological errors—Since this study aims to obtain results that could be useful for clinical and school-based practice, we also used a second scoring method that would be more viable in these environments. Both the real-word and the nonword repetition tasks were scored using the whole-word method. That is, each repeated item was evaluated as a whole and scored as either entirely correct or as an incorrect production of the target (Chiat & Roy, 2007; Dispaldro et al., 2009; Gathercole et al., 1994). Whole-word scoring was used because it is simpler and more practical than scoring phoneme by phoneme. In an additional effort to simplify scoring, no adjustments were made for developmental phonological errors using this method. Each word was scored as correct if all consonant and vowel segments were repeated correctly and there were no additions or substitutions. Under this scoring method, no special knowledge of phonetic transcription or developmental phonology would be required in order to score the task.

The audiorecordings of five children in each group (30% of the participants) were randomly selected and re-transcribed independently by a second trained listener. Phoneme-by-phoneme percentage of agreement for judgments of correctness was 99% (range from 99% to 100%).

Results

As noted earlier, this study had four aims. First, we sought to determine whether repetition performance would be better for real words than for nonwords. We also asked whether children in the SLI group would perform at lower levels than children in the TD-A group. The third and fourth aims were planned assuming that the first two expectations would be confirmed. The third aim was to use diagnostic accuracy measures to determine whether real-word repetition performance, like nonword performance, might have potential as a clinical marker of SLI in Italian. Finally, we wished to determine if all these results would change as a function of the kind of scoring method used. We approached these aims by examining group (SLI vs. TD-A) and word type (real-word, nonword) differences using the first scoring method, and repeated these comparisons employing the second scoring method. We then examine diagnostic accuracy, again using each scoring method in turn.

Group and Word Type Comparisons

Scoring Method 1: Percentage of phonemes correct with allowances for developmental phonological errors—The mean percentages of phonemes correctly repeated (with developmental phonological errors treated as correct) and the corresponding standard deviations are shown in Table 2. Means are reported for all words and nonwords, as well as for two-syllable, three-syllable, and four-syllable words and nonwords.

The correlation between real-word and nonword repetition was calculated. We found that the two tasks were highly correlated for the SLI group ($r = 0.83$, $p < .001$); however, no correlation was found for the TD-A group ($r = 0.29$, $p = .262$).

A mixed-design ANOVA was carried out on the percentage of phonemes correct for each list, with word type (real word, nonword) and word length (two syllables, three syllables, four syllables) as within-subjects factors, and group (SLI, TD-A) as a between-subjects factor. This analysis showed a significant effect of group, $F(1, 32) = 32.27$, $p < .001$, $\eta^2 = .50$. Considering both real words and nonwords together, the children with SLI were less accurate than the TD-A children in percentage of words/nonwords correctly repeated (82% vs. 98%, $d = 1.38$). A significant effect of word type was also found, $F(1, 32) = 20.84$, $p < .001$, $\eta^2 = .39$, which showed that children repeated nonwords (88%) less accurately than real words (92%, $d = .34$). The interaction between word type and group was also

significant, $F(1, 82) = 4.67$, $p = .04$, $\eta^2 = .13$; the difference between children with SLI and the TD-A group was greater for nonwords (SLI: 79%, TD-A: 97%, $d = 1.62$) than for real words (SLI: 85%, TD-A: 99%, $d = 1.05$).

A significant effect for word length was also observed, $F(2, 64) = 24.66$, $p < .001$, $\eta^2 = .44$; Bonferroni post-hoc testing at the 0.01 level revealed that two-syllable words/nonwords (94%) were easier to repeat than three-syllable words/nonwords (90%) which, in turn, were easier to repeat than four-syllable words/nonwords (86%). There was a significant interaction between length and group such that increasing length hampered the performance of the children with SLI more than that of the TD-A children, $F(2, 64) = 14.22$, $p < .001$, $\eta^2 = .31$. In children with SLI, a paired sample t -test ($p < .01$) showed that two-syllable words/nonwords (89%) were easier to repeat than three- (81%) ($d = .84$) and four-syllable words/nonwords (76%) ($d = 1.35$). A difference was also found between three- and four-syllable words/nonwords ($d = .35$). In the TD-A group, a marginal difference was present between two- (99%) and four-syllable-words/nonwords (97%) ($p = .015$, $d = .64$) and three- (98%) and four-syllable words/nonwords ($p = .018$, $d = .62$). No difference was present between two- and three-syllable words/nonwords. A comparison between groups ($p < .01$) revealed that the difference between the children with SLI and the TD-A children was greater for four-syllable words/nonwords ($d = 1.47$) than for two- ($d = 1.07$) and three-syllable words/nonwords ($d = 1.30$). Finally, the interaction between word length and word type was not significant, $F(2, 64) = 2.88$, $p = .064$, $\eta^2 = .08$.

We believed that our method of making allowances for developmental phonological errors represented a conservative approach, as it was likely to reduce the differences between the SLI and TD-A groups. However, given that the differences between the SLI and TD-A groups proved to be relatively small for two- and three-syllable words/nonwords in the preceding analysis, it seemed informative to re-compute our analysis making no allowances for developmental phonological errors. We reasoned that if the effect sizes for group for these shorter lengths were found to be larger than in the previous analysis we could suspect that these developmental phonological errors were the cause of the greater differences between groups. Indeed, we again found a significant difference between the SLI and TD-A groups for all lengths, namely, for two- (SLI: 84%; TD-A: 98%, $d = 1.35$), three- (SLI: 78%; TD-A: 98%, $d = 1.35$), and four-syllable words/nonwords (SLI: 73%; TD-A: 96%, $d = 1.58$). The difference between groups was greater in particular for two-syllable words/nonwords when developmental phonological errors were treated as incorrect than when they were treated as correct.

Scoring Method 2: Percentage of whole-words correct, with no allowances for developmental phonological errors—

The mean percentages of words correctly repeated and the corresponding standard deviations are shown in Table 2. Means are reported for all words and nonwords, as well as for two-syllable, three-syllable, and four-syllable words/nonwords. The correlation between real-word and nonword repetition was calculated. We found that the two tasks were highly correlated for the SLI group ($r = .89$, $p < .001$); however, no correlation was found for the TD-A group ($r = .28$, $p = .272$).

A mixed-design ANOVA was carried out on the percentage of words correct for each list, with word type (real word, nonword) and word length (two syllables, three syllables, four syllables) as within-subjects factors, and group (SLI, TD-A) as a between-subjects factor. This analysis showed a significant effect of group ($F(1, 32) = 100.64$, $p < .001$, $\eta^2 = .76$); children with SLI (39%) were less accurate than the TD-A children (88%) ($d = 2.57$). A significant effect of word type was also found ($F(1, 32) = 56.45$, $p < .001$, $\eta^2 = .64$) which showed that children repeated nonwords (57%) less accurately than real words (70%, $d = .$

43). The interaction between word type and group was not significant, $F(1, 82) = .47$, $p = .50$, $\eta^2 = .01$.

A significant effect for word length was also observed ($F(2, 64) = 19.93$, $p < .001$, $\eta^2 = .38$). Bonferroni post-hoc testing at the 0.01 level revealed that two-syllable words/nonwords (71%) were easier to repeat than three-syllable words/nonwords (64%) which, in turn, were easier to repeat than four-syllable words/nonwords (56%). There was a significant interaction between length and group such that increasing length adversely affected the performance of the children with SLI more than that of the TD-A children ($F(2, 64) = 4.47$, $p = .015$, $\eta^2 = .12$). In children with SLI, a paired sample t -test ($p < .01$) showed that two-syllable words/nonwords (50%) were easier to repeat than three- (38%) ($d = .57$) and four-syllable words/nonwords (28%, $d = 1.04$). A difference was also found between three- and four syllable-words/nonwords ($d = .43$). In the TD-A group, a difference was present only between two- (92%) and four-syllable-words/nonwords (84%, $d = 1.07$). A comparison between groups ($p < .01$) revealed that the differences between the children with SLI and the TD-A children were larger for four-syllable words/nonwords ($d = 2.92$) than for two- ($d = 1.92$) and three-syllable words/nonwords ($d = 2.18$). Finally, the interaction between word length and word type was not significant, $F(2, 64) = 2.89$, $p = .06$, $\eta^2 = .08$.

Assessment of Diagnostic Accuracy

Given that children with SLI performed at a lower level than typically developing children both on real words and nonwords, we proceeded to test how well these measures could discriminate between the two groups. Sensitivity and specificity were calculated using a logistic regression analysis. Nonword and real word accuracy was first examined separately, and then in combination. Sensitivity and specificity values of at least 80% were regarded as acceptable, whereas values at least 90% were regarded as good (Plante & Vance, 1994). The response operating characteristics (ROC) curve was also examined to identify the best cut-off point for the test. The area under the ROC represents the probability that a randomly selected TD-A child will receive a higher test score than a randomly selected child with SLI. Likelihood ratios (Dollaghan, 2007; Klee, 2008) were calculated only in the case in which 100% sensitivity or specificity values was not obtained. A positive likelihood ratio (LR+) was calculated using the formula: sensitivity/(1-specificity). This ratio indicates the odds that a score in the “affected” range came from a child with SLI rather than a TD-A child. A negative likelihood ratio (LR-) was calculated with the formula: (1-sensitivity)/specificity; this ratio indicates the odds that a score in the “unaffected” range came from a child with SLI. Following Sackett et al. (1991), a LR+ of 10.0 and a LR- of .10 were considered desirable. A summary of the diagnostic accuracy results is given in Table 3.

Scoring Method 1: Percentage of phonemes correct with allowances for developmental phonological errors—

As can be seen, both nonwords and real words showed good sensitivity and specificity. The LR+ shows that, for both tasks, children with SLI were about 16 times more likely than TD-A children to show low performance on the repetition tasks; the LR- showed that children with SLI were 0.06 times as likely as TD-A children to show high performance on the repetition tasks. In both tasks, the child with SLI that performed above the cut-off was Child 1 (see Table 1). When the two measures were combined in logistic regression, specificity improved further, while sensitivity remained substantially the same; likelihood ratios could not be calculated, due to the 100% specificity value obtained.

Scoring Method 2: Percentage of whole-words correct, with no allowances for developmental phonological errors—

Nonwords and real words showed excellent sensitivity and specificity both when they were considered separately and when considered

together. Likelihood ratios could not be calculated, due to the 100% sensitivity and specificity values obtained.

Discussion

The purpose of this study was to explore the utility of real-word and nonword repetition as clinical markers of SLI in Italian-speaking preschool children. Nonword repetition has been shown to distinguish children with SLI from TD-A children in Italian with acceptable accuracy (Bortolini et al., 2006); however, it is important to show that these results can be replicated, given that similar studies of English have produced mixed results. We considered real-word repetition as a potential clinical marker on the basis of previous findings that it was a stronger predictor of grammatical ability in typically developing Italian-speaking children than nonword repetition (Dispaldro et al., 2011). In addition, other investigators have found different relationships between real-word and nonword repetition performance in clinical groups as compared to TD children (Roy & Chiat, 2007). In evaluating the clinical utility of these measures, we took care to carry out the scoring in two ways – the first scoring method was more suitable to addressing questions of theory, the second method could be easily carried out in the context of clinical practice.

Group and Word Type Comparisons

In general, we found that children with SLI repeated words/nonwords less accurately than TD-A children. As noted by others (e.g., Gathercole & Baddeley, 1990), limitations in PSTM might have been responsible for the poorer performance of the SLI group. However, we cannot claim that the differences between groups were attributable entirely to PSTM. In fact, although the present study confirms that the differences between typical and atypical children became greater with increasing syllable length, the children with SLI appeared to be less accurate than their same-age peers even at the two-syllable level. In our opinion, limitations in discrimination and/or encoding might also have contributed to the group differences (e.g., Gathercole, 2006). Although we cannot rule out the possibility that production limitations contributed to these differences between groups (e.g., Shriberg et al., 2009), we did attempt to remove this potential confound by scoring developmental phonological errors as correct in Scoring Method 1. A comparison between groups revealed that the difference in accuracy between children with SLI and TD-A children on two-syllable words was larger when no allowance was made for developmental errors ($d = 1.35$) than when these developmental errors were scored as acceptable ($d = 1.07$) (effect sizes at the three- and four-syllable level did not differ). This suggests that at least some of the difference between groups that might have been attributable to production difficulties in the children with SLI was removed under Scoring Method 1.

Under Scoring Method 1, we found a significant difference between nonword repetition and real-word repetition for the children with SLI but not for the TD-A children. For the children with SLI, nonword repetition accuracy fell significantly below their real-word accuracy; for the TD-A children, word type did not affect performance. While the two tasks share several skills, such as speech perception, motor planning, PSTM and long-term linguistic knowledge (e.g., wordlikeness and phonotactic frequency), these two tasks differ in the central skill required. PSTM is primarily responsible for nonword repetition performance, whereas real-word repetition relies on the stored phonological and semantic knowledge of previously learned words (e.g., Dispaldro et al., 2009). Our study confirmed this difference for children with SLI, who showed better performance when they could draw on stored lexical knowledge in addition to PSTM than when they had to rely on PSTM resources alone (Casalini et al., 2007; Chiat & Roy 2007; Miniscalco & Gillberg, 2009; Sahlén et al. 1999).

However, there is reason to suspect that the children with SLI did not benefit as much as the TD-A children from stored lexical knowledge in the real-word repetition task. Recall that for the children with SLI, there was a significant correlation between real-word and nonword repetition performance ($r = .83$, $p < .001$ for Scoring Method 1, $r = .89$, $p < .001$ for Scoring Method 2), whereas no such relationship was seen for the TD-A children. This raises the possibility that, for the children with SLI, knowledge of the real words was rather superficial, reducing the benefits that might have accrued from lexical storage and forcing the children to rely more on PSTM than was the case for the TD-A children. For the TD-A children, greater knowledge of the real words might have rendered the real-word repetition task quite different in nature from the nonword task. This may be one reason why we found no relationship between the TD-A children's performance on these two tasks.

Unlike the children with SLI, the TD-A children approached ceiling levels under both scoring methods, and this might have contributed to the non-significant relationship between real-word and nonword repetition performance that we found for this group. However, finding such high performance levels for Italian-speaking TD-A children of this age is not unusual. Using a different set of nonwords than the one used in the present study, Bortolini et al. (2006) also found high scores on the part of their Italian-speaking TD-A participants – children who were similar in age to the participants in the present study. Yet, in the Bortolini et al. study, as in our study, group differences were quite clear. For the scoring method dealing with percentage of phonemes correct, the TD-A and SLI groups in the Bortolini et al. investigation showed mean values of 97% and 84%, and in the present study the corresponding values were 97% and 79%. For the scoring method requiring the entire word to be correct, the mean percentages of nonwords produced correctly in the Bortolini et al. study were 80% and 40% and in the present study these values were 83% and 32%, respectively.

We consider two possible reasons for the generally high percentages in the Italian data, especially on the part of the TD-A children. First, Italian words tend to be longer than words in languages such as English. For example, in the Italian database Lessico di Frequenza dell'Italiano Parlato "Lexical Frequency of Spoken Italian" (LIP, De Mauro et al., 1993; Mancini & Voghera, 1994), one-syllable words are very rare (0.96%), two-syllable words and five-syllable words occur with some frequency (14.83% and 13.01%, respectively), but three- and four-syllable words form the main part of the corpus (34.57% and 31.74% respectively). A greater familiarity with three- and four-syllable words might enable Italian-speaking children to gain greater command of longer linguistic material, which might translate into relatively good nonword repetition as well. Casalini et al. (2007), using a different two- and three-syllable nonword task, showed a ceiling effect in both typical preschool and first grade Italian children.

The second reason for ceiling effects in the TD group may have been the simplicity of our targets. Our stimuli were designed to be easy to repeat for young children. Nonwords were composed of phonemic and syllabic characteristics which are mastered by children as young as three years of age (Zmarich & Bonifacio, 2004). Moreover, the majority of real words are known by three- and four-year-old children. For example, according to the Italian MacArthur database (Caselli et al., 2007) 'nebbia' [fog], 'tamburo' [drum] and 'torre' [tower] are acquired at 33 months, 29 months, and 28 months, respectively. Thus, it was expected that both the real words and nonwords would be easy for TD children to repeat. In fact, the two- and three-syllable real and nonwords used here were tested in two previous Italian studies (Dispaldro et al., 2009, 2011); these studies found ceiling effects in four-year-old typically developing children. However, if the repetition task proved easy for TD children either because of a language input advantage or because the words and syllable structures were relatively familiar, it is clear that these advantages were not sufficient to

close the gap between Italian-speaking children with SLI and those who are developing language normally.

Comparison of Scoring Methods

In this study we found that the magnitude of group differences was greater under Scoring Method 2 (whole-word scoring) ($d = 2.57$) than under Scoring Method 1 (percentage of phonemes correct scoring) ($d = 1.38$). These results seem to show that scoring method could contribute to effect size. The difference in magnitude was primarily due to the fact that children with SLI had much lower scores under whole-word scoring; the scores of TD children did not change dramatically (Table 2). The two scoring methods also differed in that under Scoring Method 1, but not under Scoring Method 2, allowances were made for developmental phonological errors. It seems likely that this difference contributed to the larger performance gap between the groups. As noted earlier, when we modified Scoring Method 1 by not allowing for developmental phonological errors, we found a larger effect size for the comparison between the SLI and TD-A groups at the two-syllable level than we found with the original Scoring Method 1. Under Scoring Method 2, a lone developmental error on a two-syllable word/nonword would prevent that item from being counted as correct.

Diagnostic Accuracy

We found that both real-word and nonword repetition measures showed good sensitivity and specificity in distinguishing preschool-aged children with SLI from their TD same-age peers. However, before discussing the implications of these findings, we raise several caveats. First, the sensitivity and specificity values obtained should be considered in light of the gold standard used – that is, the basis on which the children were categorized as language impaired or typically developing in the first place. In the present study, children were categorized as language impaired because they qualified for intervention, and they scored poorly on a direct object clitic pronoun production task. All and only the children in the SLI group qualified for language intervention services. Although language intervention status has been used as a gold standard in previous studies (e.g., Ellis Weismer et al., 2000) and has considerable face validity, clinical judgment was used in combination with available test instruments when making these decisions. It seems likely that the specific criteria used to recommend treatment could vary across professionals, and sensitivity and specificity values could change with such variation. In order to define our groups in a more constrained way, we used an additional test as part of our gold standard. All children were also administered a direct object clitic pronoun task which was shown in earlier studies to have a good level of sensitivity and specificity (Bortolini et al., 2002; 2006). Our findings for the clitic pronoun measure were in line with these previous findings for Italian-speaking children within the same age range. Direct object clitics tend to be a weakness in SLI groups not only in Italian, but also in other Romance languages such as Spanish (Bedore & Leonard, 2001) and French (Jakubowicz et al., 1998). However, given the fact that this measure is largely a measure of grammatical ability, its use as a gold standard may have led to greater representation of children with grammatical impairments and, possibly, under-representation of children with SLI whose more serious problems did not revolve around grammar.

The second caveat is that sensitivity and specificity values are higher when, as in the experiment here, the number of children with language impairment is equal to the number of children with no impairment (Dollaghan, 2007). In addition, our SLI group constituted a clinically referred sample, as all children in this group had been brought to the attention of professionals with concerns about the children's language development. Therefore, we do not claim that the measures used here will yield comparable values in an epidemiological

study that focuses on the general population of preschoolers, where the prevalence of SLI is closer to 7%.

Our finding of good sensitivity and specificity with the nonword repetition measure indicates that, in Italian, as in several other languages, this type of ability is a useful means of separating language impaired from typically developing groups (Bortolini et al., 2006; Deevy, et al., 2010; Gray 2003; Redmond et al., 2011). This finding was not inevitable, given the differences in diagnostic accuracy found for nonword repetition tasks within and across languages (e.g., Archibald & Joanisse, 2009; Conti-Ramsden, 2003; Conti-Ramsden & Hesketh, 2003; Conti-Ramsden et al., 2001; Oetting & Cleveland, 2006; Oetting et al., 2008; Stokes et al., 2006).

Regarding real-word repetition, the fact that this task exhibited good sensitivity and specificity suggests that it taps an ability which is highly relevant to the language impairment - typical language development distinction. This may be because this measure places sufficient demands on PSTM to do some of the same “diagnostic” work accomplished by nonword repetition. For example, we observed that length effects were seen in real-word repetition as well as in nonword repetition. Alternatively, or in combination with PSTM effects, our real-word repetition measure might tap lexical skills, which constitute a weakness in the children with SLI, albeit not one closely related to their limitations in grammar. While PSTM likely contributes to decrements in performance with length, it is possible that weakness in lexical knowledge also contributed to these results. Notably, 13 of the 17 children with SLI in this study had expressive and/or receptive vocabulary scores one or more standard deviations below the mean.

It is important to stress that for both real-word repetition and nonword repetition, similar findings were observed using two different scoring methods. However, when we used a whole-word method we found excellent sensitivity and specificity. In addition to its excellent diagnostic accuracy, this scoring method is preferable because it is the simpler one to employ in a clinical setting; it does not require close phonetic transcription or determination of individual children’s phonological errors.

In summary, the findings of the present study suggest that a real-word repetition task can be a useful tool for identifying children with and without SLI who were classified according to both intervention status and performance on a test of expressive grammar (a direct object clitic pronoun task). However, it seems clear to us that a measure of real-word repetition cannot render a measure of nonword repetition unnecessary for the purpose of identifying children at risk for SLI. As we showed, nonword repetition seemed especially weak in the children with SLI, quite possibly due to its greater reliance on PSTM. Finally, our results demonstrated that diagnostic accuracy did not vary according to the scoring method used. Scoring Method 1 seemed more sensitive to PSTM deficits given that developmental phonological errors did not enter into the scoring. However, given its ease of scoring, Scoring Method 2 might hold the advantage when the goal is the initial identification of children at risk for language impairment, with more detailed diagnostic testing to be conducted at a subsequent point.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the children who participated in the study and their parents who gave their consent. We thank Francesca Scali for her help in data collection and Maura Messina for her reliability scoring. Finally, we thank the schools and the Health Services for their cooperation.

References

- Archibald LMD, Joanisse MF. On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*. 2009; 52:899–914.
- Baddeley, A. Working memory. Oxford University Press; Oxford: 1986.
- Barca L, Burani C, Arduino LS. Word naming times and psycholinguistic norms for Italian nouns. *Behaviour Research Methods, Instruments & Computers*. 2002; 34:424–434.
- Bedore L, Leonard L. Grammatical morphology deficits in Spanish-speaking children with SLI. *Journal of Speech, Language, and Hearing Research*. 2001; 44:905–924.
- Bishop DVM, North T, Donlan C. Genetic basis of specific language impairment: Evidence from a twin study. *Developmental Medicine and Child Neurology*. 1995; 37:56–71. [PubMed: 7828787]
- Bortolini, U. Prove per la Valutazione Fonologica del Linguaggio Infantile. Edit Master; Padova: 1995.
- Bortolini U, Arfé B, Caselli MC, Degasperi L, Deevy P, Leonard LB. Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition. *International Journal of Language and Communication Disorders*. 2006; 41:695–712. [PubMed: 17079223]
- Bortolini U, Caselli MC, Deevy P, Leonard LB. Specific language impairment in Italian: The first steps in the search for a clinical marker. *International Journal of Language and Communication Disorders*. 2002; 37:77–93. [PubMed: 12012618]
- Campbell T, Dollaghan C, Needleman H, Janosky J. Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research*. 1997; 40:519–525.
- Casalini C, Brizzolara D, Chilosi A, Cipriani P, Marcolini S, Pecini C, Roncoli S, Burani C. Non-word repetition in children with specific language impairment: A deficit in phonological working memory or in long-term verbal knowledge? *Cortex*. 2007; 43:769–776. [PubMed: 17710828]
- Caselli, MC.; Stefanini, S.; Pasqualetti, P. Parole e frasi nel “Primo vocabolario del bambino”. Nuovi dati normativi fra i 18 ed i 36 mesi e forma breve del questionario. Franco Angeli; Milano: 2007.
- Chiat S, Roy P. The preschool repetition test: An evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language, and Hearing Research*. 2007; 50:429–443.
- Cipriani, P.; Chilosi, AM.; Bottari, P.; Pfanner, L. UniPress; Padova: 1993. L’acquisizione della morfosintassi in italiano.
- Chilosi, AM.; Cipriani, P. Test di Comprensione Grammaticale per Bambini. Del Cerro; Tirrenia: 1995.
- Cianchetti, C.; Fancello, GS. Test di Valutazione del Linguaggio. Erikson; Trento: 2003.
- Coady J, Evans JL, Kluender KR. Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language and Communication Disorders*. 2010; 45:494–509. [PubMed: 19821795]
- Conti-Ramsden G. Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*. 2003; 46:1029–1037.
- Conti-Ramsden G, Bottin N, Faragher B. Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*. 2001; 42:741–748. [PubMed: 11583246]
- Conti-Ramsden G, Hesketh A. Risk markers for SLI: A study of young language-learning children. *International Journal of Language and Communication Disorders*. 2003; 38:251–263. [PubMed: 12851078]

- Deevy P, Wisman Weil L, Leonard LB, Goffman L. Extending the use of the NRT to preschool-age children with and without specific language impairment. *Language, Speech, and Hearing Services in Schools*. 2010; 41:277–288.
- Dispaldro M, Benelli B, Marcolini S, Stella G. Real-word repetition as a predictor of grammatical competence in Italian children. *International Journal of Language and Communication Disorders*. 2009; 44:941–961. [PubMed: 19107657]
- Dispaldro M, Deevy P, Altoé G, Benelli B, Leonard LB. A cross-linguistic study of real-word and non-word repetition as predictors of grammatical competence in children with typical language development. *International Journal of Language and Communication Disorders*. 2011; 46:564–578. [PubMed: 21899673]
- De Mauro, T.; Mancini, F.; Vedovelli, M.; Voghera, M. *Lessico di frequenza dell'Italiano parlato*. Etas Libri; 1993.
- Dollaghan, CA. *The handbook for evidence-based practice in communication disorders*. Paul H. Brookes; Baltimore: 2007.
- Dollaghan C, Campbell TF. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1136–1146.
- Dollaghan C, Biber M, Campbell TF. Lexical influence on non-word repetition. *Applied Psycholinguistics*. 1995; 16:211–222.
- Dollaghan C, Biber M, Campbell TF. Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research*. 1993; 57:421–436.
- Edwards J, Beckman ME, Munson B. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in non-word repetition. *Journal of Speech, Language, and Hearing Research*. 2004; 47:421–436.
- Ellis Weismer S, Tomblin JB, Zhang X, Buckwalter P, Chynoweth JG, Jones M. Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*. 2000; 43:865–878.
- Fancello, G.; Gianchetti, C. *Wechsler Preschool and Primary Scale of Intelligence – III*. O.S; Firenze: 2008.
- Gathercole SE. Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*. 1995; 23:83–94.
- Gathercole SE. Non-word repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*. 2006; 27:513–544.
- Gathercole SE, Baddeley AD. Phonological memory deficit in language-disordered children: Is there a causal connection? *Journal of Memory and Language*. 1990; 29:336–360.
- Gathercole SE, Willis C, Emslie H, Baddeley AD. The children's test of nonword repetition: A test of phonological working memory. *Memory*. 1994; 2:103–127. [PubMed: 7584287]
- Girbau D, Schwartz RG. Phonological working memory in Spanish–English bilingual children with and without specific language impairment. *Journal of Communication Disorders*. 2008; 41:124–145. [PubMed: 17765916]
- Graf Estes K, Evans J, Else-Quest N. Differences in nonword repetition of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*. 2007; 50:177–195.
- Gray S. Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*. 2003; 36:129–151. [PubMed: 12609578]
- Gray S. Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language and Hearing Research*. 2004; 47:1117–1132.
- Jakubowicz C, Nash L, Rigaut C, Gérard C-L. Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition*. 1998; 7:113–160.
- Kamhi A, Catts H. Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders*. 1986; 51:337–347. [PubMed: 3773490]
- Kamhi A, Catts H, Mauer D, Apel K, Gentry B. Phonological and spatial processing abilities in language- and reading-impaired children. *Journal of Speech and Hearing Disorders*. 1988; 55:140–148. [PubMed: 2299830]

- Klee T. Consideration for appraising diagnostic accuracy studies in communication disorders. *Evidence-Based Communication Assessment and Intervention*. 2008; 2:34–45.
- Leonard, LB. Children with specific language impairment. MIT Press; Cambridge, MA: 1998.
- MacWhinney, B. The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Lawrence Erlbaum Associates; Mahwah, NJ: 2000.
- Mancini, F.; Voghera, M. Lunghezza, tipi di sillabe e accento in Italiano. In: De Mauro, T., editor. *Come parlano gli Italiani*. La Nuova Italia; 1994.
- Meisels SJ. Developmental screening in early childhood: The interaction of research and social policy. *American Review of Public Health*. 1988; 9:527–550.
- Metsala J, Chisholm GM. The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics*. 2010; 31:489–506.
- Miniscalco C, Gillberg C. Non-word repetition in young school-age children with language impairment and/or neuropsychiatric disorder. *Research in Developmental Disability*. 2009; 30:1145–1154.
- Munson B, Kurtz B, Windsor J. The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2005; 48:1033–1047.
- Oetting JB, Cleveland LH. The clinical utility of nonword repetition for children living in the rural south of the US. *Clinical Linguistics & Phonetics*. 2006; 20:553–561. [PubMed: 17056486]
- Oetting JB, Cleveland LH, Cope RF. Empirically derived combinations of tools and clinical cutoffs: An illustrative case with a sample of culturally/linguistically diverse children. *Language, Speech, and Hearing Services in Schools*. 2008; 39:44–53.
- Plante E, Vance R. Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*. 1994; 25:15–24.
- Redmond SM, Thompson HL, Goldstein S. Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research*. 2011; 54:99–117.
- Reilly S, Wake M, Ukoumunne OC, Bavin E, Prior M, Cini E, Conway L, Eadie P, Bretherton L. Predicting language outcome at 4 years of age: Findings from early language in Victoria study. *Pediatrics*. 2010; 126:1530–1537.
- Roy P, Chiat S. A prosodically controlled word and nonword repetition task for 2- to 4-year olds: Evidence from typically-developing children. *Journal of Speech, Language, and Hearing Research*. 2004; 4:223–234.
- Rustioni, D. Prove di Valutazione della Comprensione Linguistica. Giunti O. S.; Firenze: 1994.
- Sackett, D.; Haynes, R.; Guyatt, G.; Tugwell, P. *Clinical epidemiology*. Little, Brown; Boston: 1991.
- Sahlén B, Reuterskiöld-Wagner C, Nettelbladt U, Radeborg K. Non-word repetition in children with language impairment - Pitfalls and possibilities. *International Journal of Language and Communication Disorders*. 1999; 34:337–352. [PubMed: 10884905]
- Shriberg LD, Lohmeier HL, Campbell TF, Dollaghan CA, Green JR, Moore CA. A nonword repetition task for speakers with misarticulations: The syllable repetition task (SRT). *Journal of Speech, Language, and Hearing Research*. 2009; 52:1189–1212.
- Stella, G.; Pizzioli, C.; Tressoldi, PE. Peabody – Test di Vocabolario Recettivo - P.P.V.T.-R. Peabody Picture Vocabulary Test – Revised. Omega Edizioni; 2000.
- Stokes SF, Wong AMY, Fletcher P, Leonard LB. Nonword repetition and sentence repetition as clinical markers of SLI: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*. 2006; 49:219–236.
- Thal D, Miller S, Carlson J, Vega M. Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research*. 2005; 48:1481–1495.
- Tomblin JB, Records N, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*. 1997; 40:1245–1260.

- Washington JA, Craig HK. A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology*. 2004; 13:329–340. [PubMed: 15719899]
- World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 10th Revision. Geneva: 2005.
- Zmarich, C.; Bonifacio, S. Gli inventari fonetici dai 18 ai 27 mesi d'età: Uno studio longitudinale. In: Albano Leoni, F.; Cutugno, F.; Pettorino, M.; Savy, R., editors. *Il parlato Italiano*. Auria Editore; Napoli: 2004. *Atti del Convegno Nazionale*

Table 1

Expressive and Receptive Language of Participants with SLI: Summary Scores

Child	Age ^a	PIQ ^b	Expressive Language			Receptive Language		
			lexicon	Grammar	MLU ^f	lexicon	grammar	
			TVL ^c	EM ^d	DOCP ^e	PPVT ^g	TVL ^h	TCGB ⁱ PVCL ^j
1	5;5	106	1°	M	79%	104	50°	
2	4;6	115	50°	H	0%		95°	75°
3	5;0	91	1°	H	13%	70		10°
4	5;8	85	1°	M	50%	72		25°
5	4;1	96	1°	H	31%		15°	10°
6	4;9	90	1°	H	40%		45°	10°
7	3;11	109	5°	H	0%	80	25°	
8	4;3	105	25°	H	64%	83		HM
9	5;8	85	25°	H	0%	3.31	45°	50°
10	4;10	96	5°	M	27%	3.39	84	25°
11	4;4	106	1°	H	20%	2.97	5°	10°
12	4;5	85	5°	H	46%	2.97	35°	LM
13	5;1	85	1°	H	20%	2.63	25°	75°
14	5;3	98	1°	H	69%	3.50	25°	10°
15	4;8	85	1°	H	0%	2.28	1°	G
16	4;4	96	35°	M	8%	2.86	25°	25°
17	5;4	100	25°	M	36%	3.62	25°	25°

^a AGE is expressed in years:months.^b PIQ (Performance IQ, WPPSI-III) has a mean of 100 and a SD of 15^c TVL (Test di Valutazione del Linguaggio): Expressive vocabulary is expressed in 11 centile points (1°, 5°, 15°, 25°, 35°, 45°, 55°, 65°, 75°, 85°, 95°).^d EM: Expressive Morphosyntax was evaluated by clinicians through spontaneous language samples as adequate (A), mildly inadequate (M) or highly inadequate (H).^e DOCP (Third Person Direct Clitic Pronoun Use): the values are expressed in percentages of correct production. Performance of 2 SD below the mean was used as one of the inclusion criteria for the SLI group.^f MLU is expressed in words.

g_{PPVT} : has a mean of 100 and a SD of 15

h_{TVL} (Test di Valutazione del Linguaggio): Receptive vocabulary is expressed in 11 centile points (1°, 5°, 15°, 25°, 35°, 45°, 55°, 65°, 75°, 85°, 95°).

i_{TCGB} (Test di Comprensione Grammaticale per Bambini): results are expressed in 5 centile points (10°, 25°, 50°, 75°, 95°).

j_{PVCL} (Prove di Valutazione della Comprensione Linguistica): results are expressed in 6 classes (1 = insufficient, P = poor, LM = low middle, M = middle, HM = high middle, G = good, VG = very good).

Mean Percentage (and Standard Deviations) at each Word Length for Real Words (RW) and Nonwords (NW) for Each Scoring Method.

Table 2

		2 syllables		3 syllables		4 syllables		Total	
		RW	NW	RW	NW	RW	NW	RW	NW
Scoring	SLI	90 (11)	87 (8)	84 (15)	78 (15)	80 (15)	72 (14)	85 (13)	79 (11)
Method									
1	TD	99 (1)	98 (3)	99 (2)	98 (3)	98 (3)	95 (4)	99 (1)	97 (2)
Scoring	SLI	56 (26)	45 (21)	44 (27)	30 (24)	36 (22)	19 (20)	46 (21)	32 (18)
Method									
2	TD	95 (6)	89 (13)	94 (9)	85 (16)	94 (10)	74 (17)	94 (5)	83 (11)

Scoring method 1: Percentage of phonemes correct with allowances for developmental phonological errors.
Scoring method 2: Percentage of words correctly produced.

Percentages and Number of Children Correctly Classified by Nonwords (NW) and Real Words (RW) for each Scoring Method.

Table 3

Scoring	Measure	SLI Group (Sensitivity)	TD-A Group (Specificity)	Cut off	Area under ROC curve ^a	LR ^b ₊	LR ₋
Method 1	Nonword	94.1% (16/17)	94.1% (16/17)	93%	.991	15.94	0.06
	Real Word	94.1% (16/17)	94.1% (16/17)	96.5%	.991	15.94	0.06
	NW & RW	94.1% (16/17)	100% (17/17)	93.75%	.995		
Method 2	Nonword	100% (17/17)	100% (17/17)	65%	1.000		
	Real Word	100% (17/17)	100% (17/17)	79%	1.000		
	NW & RW	100% (17/17)	100% (17/17)	72%	1.000		

^aROC curve = response operating characteristics curve

^bLR₊ = positive likelihood ratio; LR₋ = negative likelihood ratio