JSLHR

Research Note

Shortened Nonword Repetition Task (NWR-S): A Simple, Quick, and Less Expensive Outcome to Identify Children With Combined Specific Language and Reading Impairment

Carlijn M. P. le Clercq,[a] Marc P. van der Schroeff,[a] Judith E. Rispens,[b] Liesbet Ruytjens,[a] André Goedegebure,[a] Gijs van Ingen,[a] and Marie-Christine Franken[a]

Purpose: The purpose of this research note was to validate a simplified version of the Dutch nonword repetition task (NWR; Rispens & Baker, 2012). The NWR was shortened and scoring was transformed to correct/incorrect nonwords, resulting in the shortened NWR (NWR-S).
Method: NWR-S and NWR performance were compared in the previously published data set of Rispens and Baker (2012; $N = 88$), who compared NWR performance in 5 participant groups: specific language impairment (SLI), reading impairment (RI), both SLI and RI, one control group matched on chronological age, and one control group matched on language age.
Results: Analyses of variance showed that children with SLI + RI performed significantly worse than other participant groups in NWR-S, just as in NWR. Logistic regression analyses showed that both tasks can predict an SLI + RI outcome. NWR-S holds a sensitivity of 82.6% and a specificity of 95.4% in identifying children with SLI + RI. The sensitivity of the original NWR is 87.0% with a specificity of 87.7%.
Conclusions: As the original NWR, the NWR-S comprising a subset of 22 nonwords scored with a simplified scoring system can identify children with combined SLI and RI while saving a significant amount of the needed assessment time.
Supplemental Materials: https://doi.org/10.23641/asha.5150116

With nonword repetition (NWR), interactions between the auditory system, phonological representations, articulation, vocabulary, and literacy are tested (Gathercole, 2006). Impaired NWR has been recognized as a behavioral marker of specific language impairment (SLI; Bishop, North, & Donlan, 1996; Casserly & Pisoni, 2013; Conti-Ramsden, Botting, & Faragher, 2001; Dollaghan & Campbell, 1998; Metsala, 1999) with difficulties already surfacing at an early preschool age (McKean, Letts, & Howard, 2013). A number of studies have shown that in school-aged children with SLI who have received literacy instruction, NWR is particularly affected in those who are also poor readers, and to be more specific, impaired NWR performance is predominantly seen in children with both reading impairment (RI) and SLI. When compared with children with only SLI, only RI, and typical development (TD), those children with combined SLI and RI perform significantly worse on NWR (Bishop, McDonald, Bird, & Hayiou-Thomas, 2009; Catts, Adlof, Hogan, & Weismer, 2005; Rispens & Baker, 2012; Rispens & Parigger, 2010). NWR performance is thus an important indication of SLI and/or RI and provides useful information about their language proficiency. Nevertheless, administering an NWR task is time-consuming as, in general, children's repetitions need to be recorded, transcribed, and analyzed at the phoneme-level by a speech-language pathologist or a researcher. As an alternative, a repeated nonword may be judged as correct or incorrect (nonword-level scoring). Although the latter is much quicker to apply, it only expresses that at least one phoneme of the nonword was incorrectly repeated instead of the exact number of correctly or incorrectly repeated phonemes (Graf Estes, Evans, & Else-Quest, 2007). As a result, the measure might be too crude, and differences between groups of children may either be overestimated or underestimated. It is the

[a]Department of Otolaryngology, Head and Neck Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands
[b]Amsterdam Center for Language and Communication, University of Amsterdam, the Netherlands
Correspondence to C. M. P. le Clercq: c.leclercq@erasmusmc.nl

Editor: Sean Redmond
Associate Editor: Lisa Archibald
Received February 12, 2016
Revision received June 29, 2016
Accepted January 9, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0060

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

aim of the current study to establish if this last assumption holds by replicating a study by Rispens and Baker (2012). The NWR as designed and used by Rispens and Baker in their study comprised 40 nonwords, and the responses of the children were scored on the phoneme level. In the current study, our purpose is to validate a shortened NWR (NWR-S) as a simpler and quicker alternative for NWR, using a reanalysis of the data from Rispens and Baker.

Method

Participants

The participants in this study were a subsample of those reported by Rispens and Baker (2012). In this study, five groups of children participated: children with SLI and RI, children with only SLI, children with only RI, and two groups of TD children. From the chronological age-matching TD (CA TD) group that participated in the Rispens and Baker study, 25 were selected for this study. Only children with CA TD who were within the same age range as the children with SLI + RI were included in order to optimize matching. The study population used consisted of 88 participants divided into five study groups, including six boys and four girls with SLI (mean age 8;1 [years;months] years), seven boys and seven girls with RI (mean age 8;3 years), 17 boys and six girls with SLI + RI (mean age 8;0 years), 12 boys and 13 girls with TD matched on chronological age (CA TD; mean age 7;9 years), and nine boys and seven girls with TD matched on vocabulary age (raw scores on the Dutch Peabody Picture Vocabulary Test; Schlichting, 2005) of the children with SLI, serving as a language age–matched control group (LA TD; mean age 5;8 years); see Table 1 for the descriptive characteristics of the children. The 14 children with RI were recruited through special needs teachers, who tutored these children specifically for their poor reading and spelling skills. They had TD oral language skills according to their teacher and parents, and they had not been referred to speech and language therapists for concerns about their language development. SLI was diagnosed when a child performed at least 1.5 *SD* below the mean in at least two language domains measured with Dutch standardized language tests or when a child performed more than 2 *SD* below the mean on a Dutch standardized general language test. These diagnostic tests were carried out by a multidisciplinary team from the special needs schools, and an SLI diagnosis was required in order for a child to be accepted into the special needs schools. Children who had evidence or a history of speech output problems had been excluded from participation in the study as such problems interfere with the reliability of NWR scoring. All children had normal hearing and normal or corrected-to-normal vision. Only those children who were raised in a family with at least one parent who was a native speaker of Dutch and in which Dutch had been spoken from birth onward were included. The special needs schools were located in the north and the middle of the Netherlands. Reading performance was tested in all children with SLI by two standardized reading tasks, namely the real-word reading task (Brus & Voeten, 1973) and a pseudoword reading task (van den Bos, Spelberg, Scheepstra, & de Vries, 1994). RI was diagnosed when performance was at least 1 standard deviation below the mean score on both tasks. The group with SLI was divided in two groups on the basis of whether or not they had additional reading problems evidenced by their scores on the two reading tasks. With regard to the matching between the impaired and TD groups, the CA TD group was closely matched on chronological age to the children with SLI + RI (post hoc Games-Howell, $p = .992$). The LA TD group was closely matched on language age on the basis of the Dutch Peabody Picture Vocabulary Test scores (post hoc Games-Howell, $p = .486$; see Rispens & Baker, 2012, for further details).

Table 1. Participant characteristics.

Group	*N*	Gender		Age		
		Boys	Girls	*M*	*SD*	Range
SLI + RI	23	17	6	8;0	0;2	7;0–8;10
SLI	10	6	4	8;1	0;2	7;2–8;10
RI	14	7	7	8;3	0;7	7;1–9;7
CA TD	25	12	13	7;9	0;4	7;1–8;9
LA TD	16	9	7	5;8	0;3	5;4–6;4

Note. Ages are expressed in years;months. SLI = specific language impairment; RI = reading impairment; CA TD = chronological age–matched typically developing control group; LA TD = language age–matched typically developing control group.

Stimuli

The NWR task reported in Rispens and Baker (2012) consists of 40 nonwords, comprising nonwords of two, three, four, or five syllables of either high or low phonotactic probability. These phonotactic probabilities were calculated using a tool to calculate Dutch phonotactic frequencies (Adriaans, 2006) on the basis of the corpus of spoken Dutch (Oostdijk, 2000); see the Appendix for an overview of all nonwords and their phonotactic probabilities. Nonwords were presented using a sound system linked to a laptop computer in two pseudorandomized blocks of 20 nonwords. Conducting the task took approximately 3 min without taking the pause in between the two blocks into account. The responses of the participants were recorded and transcribed, and the percentages of phonemes repeated correctly per nonword were calculated, requiring at least another 30 to 45 min per child, depending on the amount of errors.

Stimuli in NWR-S

From the original NWR task, 22 nonwords were selected for our NWR-S (see the Appendix for an overview of these nonwords and their phonotactic probabilites). All two-syllable nonwords were omitted because Rispens and Baker's (2012) study showed that all 8-year-olds could

correctly repeat these nonwords. Of the three- to five-syllable nonwords, four nonwords of low phonotactic probability and four nonwords of high phonotactic probability of each length were selected. There was one exception in the three-syllable nonwords, which included only two nonwords of high phonotactic probability to compensate for omitting the (most simple) two-syllable nonwords.

Scoring Method in NWR-S

The current study aimed to develop a task that can be conducted in real time. We transformed the original method (percentage of phonemes correct, phoneme-level scores) into a quicker, simplified method (percentage of nonwords correct, nonword-level scores). The repeated nonwords are judged by the examiner directly after each repetition of the child with a button-press device with one button pressed for a correctly repeated nonword and another button pressed for an incorrect response. The next nonword is offered after each real-time judgment. Consistent articulation errors are permitted, for instance, due to a lisp or dialect, as the phonemes remain the same.

Procedure and Data Analysis

All nonwords present in the NWR task were previously scored using the old (phoneme-level) method, forming the data set used in the study of Rispens and Baker (2012). For each nonword in the data set, the phoneme-level score was transformed into a nonword-level score (correctly or incorrectly repeated nonword) in addition to the original phoneme-level scores (NWR-S). Pearson's correlation analyses were performed on the different modifications from NWR into NWR-S to study the association and to quantify the strength of the relationship between the two tasks.

The main analyses were in order to replicate the performance results from Rispens and Baker's (2012) NWR. An analysis of variance (ANOVA) followed by post hoc Games-Howell analyses, due to unequal variances and unequal sample sizes, was used to assess the discriminative power of NWR-S for the five separate participant groups. Logistic regression analyses were performed on the predictive value of NWR and NWR-S on the children with SLI + RI, only SLI, and only RI. To assess the diagnostic accuracy of NWR-S compared with NWR, the logistic regression analyses were used to conduct receiver operating characteristics in order to evaluate the sensitivity, specificity, and likelihood ratios of both tasks. A sensitivity and specificity of at least 80% were regarded as acceptable, and values above 90% were considered to be good. The positive and negative likelihood ratios (LR+ and LR−) indicate the odds that a score in the affected and the unaffected range, respectively, came from a child with SLI + RI. An LR+ of ≥ 10.00 and an LR− of ≤ 0.10 were considered desirable (Sackett, Haynes, Guyatt, & Tugwell, 1991). All analyses were performed using SPSS Statistics 21 (IBM, 2012).

Results

Correlation of NWR and NWR-S

The selected 22 nonwords were first compared with the original 40 nonwords using the original phoneme-level scores to assess the impact of shortening the task. As visualized in Figure 1a, Pearson's correlations for 40 and 22 nonwords were strong ($r = .983$, $p < .001$). When using the nonword-level scores, Pearson's correlations for 40 and 22 nonwords were strong as well ($r = .931$, $p < .001$; Figure 1b). Last, because the NWR was adjusted on both the number of nonwords and the scoring method, the combination of these adjustments in NWR-S were compared with the original NWR outcomes. This resulted in a correlation coefficient of .861, $p < .001$ (see Figure 1c). Sub analyses per study group are available in Supplemental Materials S1, S2, S3, S4, and S5.

Replication of NWR Results Using NWR-S: Discrimination of Clinical Groups

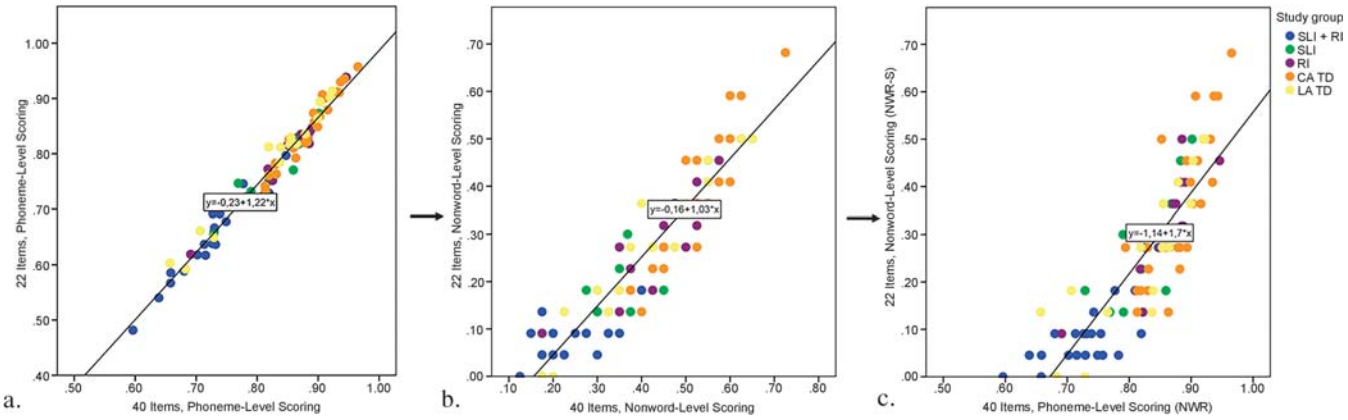
Rispens and Baker (2012) reported that children with SLI + RI had significantly impaired NWR performance compared with their peers (SLI, RI, CA TD, and LA TD). An ANOVA was carried out to compare the original NWR performance between the participant groups, $F(4, 83) = 19.14$, $p < .001$. Post hoc Games-Howell analyses, due to unequal sample sizes, showed significantly lower scores for children with SLI + RI ($M = 73.0\%$, $SD = 5.9$; see Figure 2) compared with the other study groups ($p = .002$ compared with children with only SLI, $p < .001$ compared with children with only RI and with CA TD, and $p = .009$ compared with children with LA TD). There were no significant differences between the SLI, RI, CA TD, and LA TD participant groups (all ps $> .10$; see Table 2).

An ANOVA on performance on NWR-S showed similar significant differences between groups, $F(4, 83) = 13.92$, $p < .001$. Post hoc Games-Howell analyses showed significantly lower scores for children with SLI + RI ($M = 8.5\%$, $SD = 6.2$; see Figure 2) compared with the other study groups ($p = .007$ compared with children with only SLI, $p < .001$ compared with children with RI and CA TD, and $p = .005$ compared with children with LA TD). As with the results of the NWR, there were no significant differences between the SLI, RI, CA TD, and LA TD participant groups (all ps $> .40$; see Table 2).

Replication of NWR Results Using NWR-S: Diagnostic Accuracy

Children with SLI + RI performed significantly worse than the other included study groups on both NWR and NWR-S (see Figure 2). Logistic regression analyses confirmed that both a poorer NWR and a poorer NWR-S performance can predict an SLI + RI outcome (accounting for 54.3% of the variation, $\beta = -24.5$, 95% CI [−34.8, −14.2] for NWR and NWR-S accounting for 62.2% of the variation, $\beta = -21.3$, 95% CI [−31.3, −11.2]).

Figure 1. Correlation plots for the transformation from nonword repetition (NWR) into shortened nonword repetition (NWR-S; $N = 88$). (a) Shortening the original NWR from 40 to 22 nonwords while using the original phoneme-level scoring (Pearson's $r = .983$, $p < .001$). (b) Comparing 22 to 40 nonwords when using nonword-level scores (Pearson's $r = .931$, $p < .001$). (c) Comparing 22 nonwords using the nonword-level scoring method (NWR-S) with the original 40 nonwords scored on the phoneme-level (NWR; Pearson's $r = .861$, $p < .001$). SLI = specific language impairment; RI = reading impairment; CA TD = chronological age–matched typically developing control group; LA TD = language age–matched typically developing control group.



Receiver operating characteristics curves were calculated to assess if NWR and NWR-S have comparable diagnostic accuracy. The area under the receiver operating curve, sensitivity, and specificity were compared between both tasks. NWR and NWR-S showed similar diagnostic strength in identifying SLI + RI on the basis of the obtained task results. The area under the curve was .906 (95% CI [.844, .969]) for NWR and .929 (95% CI [.870, .988]) for NWR-S and thus comparable ($\chi^2 = 0.84$, $p = .361$; see Figure 3). The corresponding sensitivity of NWR was

Figure 2. Nonword repetition (NWR) and shortened nonword repetition (NWR-S) performance for children with SLI + RI versus the children without the double deficit, average scores. NWR: SLI + RI = 73.0%*, peers = 85.1%; NWR-S: SLI + RI = 8.5%*, peers = 30.9%. *Significantly lower than peers without the double deficit ($p < .001$). SLI = specific language impairment; RI = reading impairment. Error bars: 95% CI.

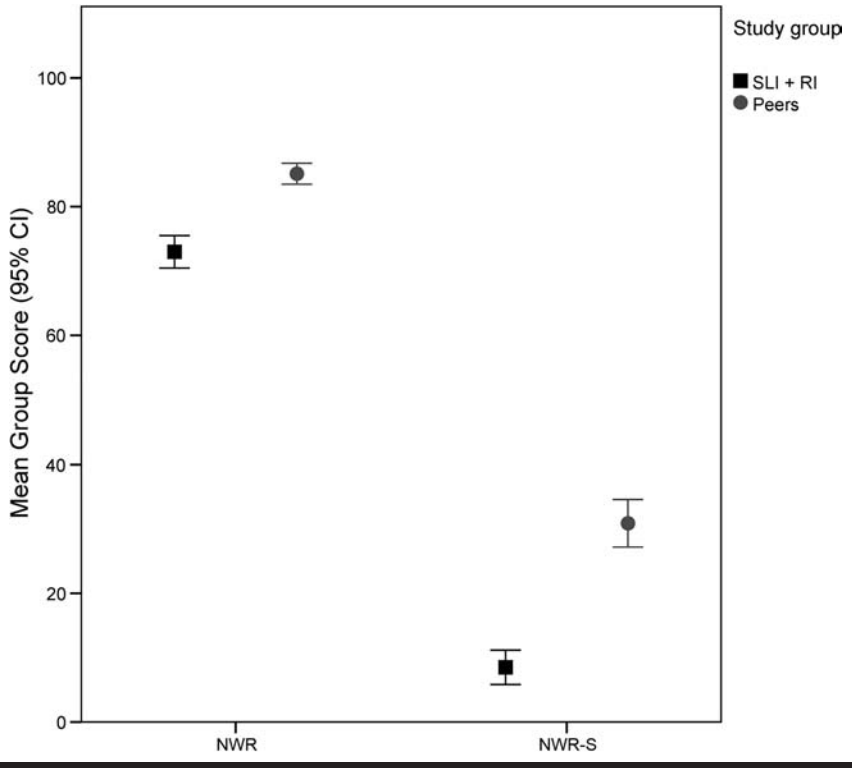


Table 2. Nonword repetition (NWR) and shortened nonword repetition (NWR-S) results for the five study groups.

Task	Group	Group result		*p* values post hoc comparisons			
		M	*SD*	SLI + RI	SLI	RI	CA TD
NWR	SLI + RI	73.0	5.9				
	SLI	82.8	5.6	.002			
	RI	85.3	5.9	< .001	.823		
	CA TD	87.9	4.7	< .001	.131	.621	
	LA TD	82.1	8.6	.009	.999	.753	.134
NWR-S	SLI + RI	8.5	6.2				
	SLI	27.6	12.9	.007			
	RI	30.8	12.0	< .001	.967		
	CA TD	35.3	15.8	< .001	.575	.861	
	LA TD	26.1	16.1	.005	.999	.889	.401

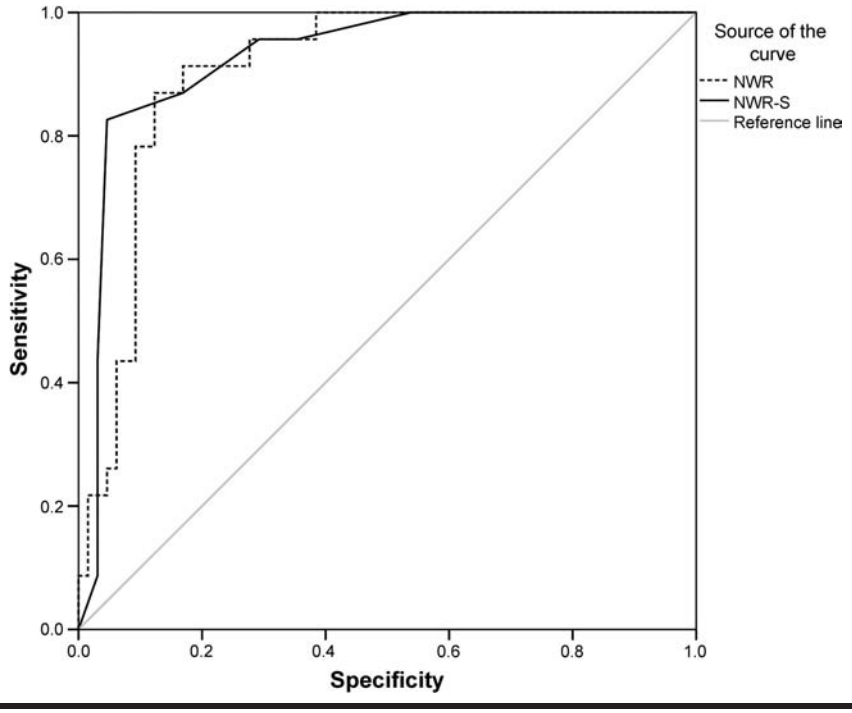
Note. All group results are reported as the percentage correctly repeated. Post hoc comparisons were performed using Games-Howell testing due to unequal sample sizes of the groups. Em dashes indicate data not [obtained/reported/available]. SLI = specific language impairment; RI = reading impairment; CA TD = chronological age–matched typically developing control group; LA TD = language age–matched typically developing control group.

87.0% with a specificity of 87.7% (optimal cutoff value 78.6%, LR+ of 7.07, LR− of 0.15). The sensitivity of NWR-S was 82.6% with a specificity of 95.4% (optimal cutoff value 11.4%, LR+ 17.96, LR− 0.18).

Discussion

This study showed that the outcomes of NWR-S consisting of 22 nonwords with a simplified nonword-level scoring system are similarly distributed as the 40 original nonwords with a phoneme-level scoring system from NWR as reported in Rispens & Baker (2012). This original NWR task scored on phoneme level is a time-consuming procedure and is not suitable for screening purposes in daily routine or large-scale research, such as population-based studies. Therefore, we adjusted the NWR into a shorter and simplified version, resulting in the NWR-S. The 22 nonwords used in the NWR-S were selected on the basis of phonotactic probability in comparison with the full set of nonwords in the original NWR task. All nonwords of two syllables were omitted. It was reasonable that older children are further along in their language development, and Rispens and Baker's (2012) study had shown that two-syllable nonwords are easily repeated by our target population of 8-year-olds. To challenge these older children, we retained the more difficult nonwords containing three to five syllables. Shortening of clinical tests has been applied successfully in the past. Radeborg, Barthelom, Sjoberg, and Sahlen (2006) showed that there was no difference between the full and a shortened version of a NWR task, and the shorter version was recommended. Likewise, a shortened version of a language subtest of a developmental assessment scale (Wong, Leung, Siu, & Lam, 2012) and of a naming test

Figure 3. Receiving operator characteristics (ROC) to assess the diagnostic accuracy of nonword repetition (NWR) and shortened nonword repetition (NWR-S) in distinguishing children with combined specific language and reading impairment (SLI + RI) from peers without the double deficit. The area under the ROC curve was .906 for NWR with a sensitivity of 87.0% and a specificity of 87.7%. The area under the ROC curve for NWR-S was .929 with a sensitivity of 82.6% and a specificity of 95.4%. Diagonal segments are produced by ties.



(Mack, Freed, Williams, & Henderson, 1992) were demonstrated to show comparable measurement properties to the original longer versions.

Our results show that there were strong and significant similarities between the longer NWR with phoneme-level scoring and the NWR-S using scores on the nonword level. We therefore conclude that scoring on the nonword level offers a real-time alternative for the original offline scoring in percentage phonemes correct. Real-time scoring requests immediate judgment before the next nonword is offered. This implicates immediate evaluation of the given responses with an instantaneous available test result. Even though nonword-level scoring has been shown to be less sensitive than scoring per phoneme, scoring nonwords as either correct or incorrect has been found to be an efficient and effective method for distinguishing between children with SLI and those with a typical language development (Graf Estes et al., 2007; Gray, 2003; Roy & Chiat, 2004; Topbaş, Kaçar-Kütükçü, & Kopkalli-Yavuz, 2014). This is because children with language impairment suffer more from stricter scoring methods than children with TD, causing larger group differences (Dispaldro, Leonard, & Deevy, 2013). Therefore, Dispaldro et al. (2013) even stated that because of the simpler scoring method in a clinical setting and sufficient diagnostic accuracy, the use of nonword-level scoring is preferred. It especially holds the advantage when the aim is to initially identify children at risk for language impairment with more detailed diagnostic testing to be conducted at a subsequent point. This is consistent with the results in the present study. NWR-S may compromise on precision and nuance of the obtained task scores compared with NWR. As expected, performance was poorer when using scoring on the nonword level than on the phoneme level. However, the analyses demonstrated no performance differences in group comparisons and an equal sensitivity and specificity.

Both the NWR and the NWR-S tasks discriminate between children with SLI + RI and their peers but not between SLI- and RI-only children and children with TD. This is different from what is more often reported in the literature (Bishop et al., 1996; Conti-Ramsden et al., 2001; de Bree, Wijnen, & Gerrits, 2010). Rispens and Baker (2012) hypothesized in their study that this could be explained by the development of orthographic skills, which influences phonological processing and this NWR performance. Another possibility of this difference with the literature is the used definition of reading problems. In the study of Rispens and Baker, a threshold of 1 *SD* below the mean was used for impaired reading performance. This can be considered as a lenient criterion for the definition of RI.

The current study was conducted using the data set from the previously published study from Rispens and Baker (2012). NWR was administered once to construct the data set, after which the different scores were derived. This could have inflated the results. The various analyses were carried out to confirm that different aspects of NWR and NWR-S are comparable, and children score in the same clinical category on NWR-S as on NWR. However, we acknowledge that proof of NWR-S would have been stronger with independent measurements.

The present study aimed to design a task that provides easier administration and scoring of an NWR task, which is necessary for large-scale research. The shorter and simpler NWR-S comprises 22 nonwords and a nonword-level scoring system in real time. In particular, this real-time scoring method makes NWR-S a much quicker alternative to NWR. Future studies are needed to confirm the currently presented validity of NWR-S in identifying children with SLI and/or RI in clinical practice.

References

Adriaans, F. (2006). PhonotacTools (Test version) [Computer software]. Utrecht, the Netherlands: Utrecht Institute of Linguistics OTS.

Bishop, D. V., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children who read words accurately despite language impairment: Who are they and how do they do it? *Child Development, 80,* 593–605.

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37,* 391–403.

Brus, B. T., & Voeten, M. J. M. (1973). *Een-Minuut test. Vorm A en B* [One-minute test. Form A and B]. Nijmegen, the Netherlands: Berkhout Testmateriaal.

Casserly, E. D., & Pisoni, D. B. (2013). Nonword repetition as a predictor of long-term speech and language skills in children with cochlear implants. *Otology & Neurotology, 34,* 460–470.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48,* 1378–1396.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 741–748.

de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition and literacy in Dutch children at-risk of dyslexia and children with SLI: Results of the follow-up study. *Dyslexia, 16,* 36–44.

Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: A study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 56,* 323–336.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513–543.

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50,* 177–195.

Gray, S. (2003). 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, 36,* 129–151.

IBM. (2012). SPSS Statistics for Windows (Version 21.0) [Computer software]. Armonk, NY: Author.

Mack, W. J., Freed, D. M., Williams, B. W., & Henderson, V. W. (1992). Boston Naming Test: Shortened versions for use in Alzheimer's disease. *Journal of Gerontology, 47,* P154–P158.

McKean, C., Letts, C., & Howard, D. (2013). Developmental change is key to understanding primary language impairment: The case of phonotactic probability and nonword repetition. *Journal of Speech, Language, and Hearing Research, 56,* 1579–1594.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91,* 3–19.

Oostdijk, N. (2000, May–June). *The spoken Dutch corpus: Overview and first evaluation*. In M. Gravilidou, G. Caravannis, S. Markantonatou, S. Piperidis, & G. Stainhaouer (Eds.), *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC)* (Vol. 2, pp. 887–894). Athens, Greece: LREC.

Radeborg, K., Barthelom, E., Sjoberg, M., & Sahlen, B. (2006). A Swedish non-word repetition test for preschool children. *Scandinavian Journal of Psychology, 47,* 187–192.

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research, 55,* 683–694.

Rispens, J., & Parigger, E. (2010). Non-word repetition in Dutch-speaking children with specific language impairment with and without reading problems. *The British Journal of Developmental Psychology, 28*(Pt. 1), 177–188.

Roy, P., & Chiat, S. (2004). 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, 47,* 223–234.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. (1991). *Clinical epidemiology: A basic science for clinical medicine*. Boston, MA: Little, Brown.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test–III NL.* Amsterdam, the Netherlands: Harcourt Test Publishers.

Topbaş, S., Kaçar-Kütükçü, D., & Kopkalli-Yavuz, H. (2014). Performance of children on the Turkish Nonword Repetition Test: Effect of word similarity, word length, and scoring. *Clinical Linguistics & Phonetics, 28,* 602–616.

van den Bos, K. P., Spelberg, H. C. L., Scheepstra, A. J. M., & de Vries, J. R. (1994). *De Klepel. Een test voor de leesvaardigheid van pseudo-woorden* [The Klepel. A test for the ability to read pseudo-words]. Nijmegen, the Netherlands: Berkhout Testmateriaal.

Wong, A. M., Leung, C., Siu, E. K., & Lam, C. C. (2012). Validating the language domain subtest in a developmental assessment scale for preschool children. *Research in Developmental Disabilities, 33,* 1633–1641.

Appendix

Items and their phonetic transcriptions of the original 40-item nonword repetition (NWR) task and the selected items of the 22-item shortened nonword repetition (NWR-S) task.

40-item NWR task			22-item NWR-S task		
Orthography	**IPA**	**Phonotactic probability**	**Orthography**	**IPA**	**Phonotactic probability**
Two-syllable items					
weugof	wø:xɔf	Low			
kuimup	kœymYp	Low			
luubuf	lybYf	Low			
joefeum	jufø:m	Low			
feusut	føsYt	Low			
hiewam	hiwɑm	High			
raanom	ranɔm	High			
geeres	xerɛs	High			
woosel	wosɛl	High			
daanes	danɛs	High			
Three-syllable items					
woezuunim	wuzynlm	Low	woezuunim	wuzynlm	Low
muihuuguf	mœyhyxYf	Low	muihuuguf	mœyhyxYf	Low
soeguipem	suxœypɛm	Low	soeguipem	suxœypɛm	Low
nuigeusup	nœyxø:sYp	Low	nuigeusup	nœyxø:sYp	Low
veujoetup	vø:jutYp	Low	loowaamas	lowamɑs	High
kaaroodin	karodɪn	High	taanoolon	tanolɔn	High
voopeeket	vopekɛt	High			
loowaamas	lowamɑs	High			
taanoolon	tanolɔn	High			
deevoenos	devunɔs	High			
Four-syllable items					
meufuusuinef	mø:fysœynɛf	Low	meufuusuinef	mø:fysœynɛf	Low
suijiegoonif	søyjixonɪf	Low	suijiegoonif	søyjixonɪf	Low
juuvuigoowuf	jyvœyxowYf	Low	juuvuigoowuf	jyvœyxowYf	Low
guiweusoegir	xœywø:suxɪr	Low	guiweusoegir	xœywø:suxɪr	Low
fuiseuwoesut	fœysø:wusYt		liejootaanig	lijotanɪx	High
liejootaanig	lijotanɪx	High	peewaatoopes	pewatopɛs	High
peewaatoopes	pewatopɛs	High	saaviebeemer	savibemɛr	High
liekoovoepar	likovupɑr	High	kooviewaalan	koviwalɑn	High
saaviebeemer	savibemɛr	High			
kooviewaalan	koviwalɑn	High			
Five-syllable items					
baamerienooves	bamɛrinovɛs	High	baamerienooves	bamɛrinovɛs	High
geerutievaanot	xerYtivanɔt	High	tieloniedaanag	tilɔnidanɑx	High
tieloniedaanag	tilɔnidanɑx	High	wookaloemoodon	wokalumodɔn	High
wookaloemoodon	wokalumodɔn	High	beemonievoekes	bemɔnivukɛs	High
beemonievoekes	bemɔnivukɛs	High	fuugiwuinoefep	fyxɪwœynufɛp	Low
fuugiwuinoefep	fyxɪwœynufɛp	Low	soegonuifeusir	suxɔnœyfø:sɪr	Low
soegonuifeusir	suxɔnœyfø:sɪr	Low	nuijigeufuusut	nœyjɪxø:fysYt	Low
geumuwoekuubir	xø:mYwukybɪr	Low	jeunimeusuifir	jø:nɪmø:sœyfɪr	Low
nuijigeufuusut	nœyjɪxø:fysYt	Low			
jeunimeusuifir	jø:nɪmø:sœyfɪr	Low			

Note. IPA = International Phonetic Alphabet.