ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Detecting Developmental Language Disorder in Monolingual and Bilingual Children: Comparison of Language-Specific and Crosslinguistic Nonword Repetition Tasks in French and Portuguese

Salomé Schwob[a] and **Katrin Skoruppa**[a]

[a] Institut des Sciences Logopédiques, Maison des Sciences du Langage et de la Communication, Université de Neuchâtel, Switzerland

ARTICLE INFO

Article History:
Received January 11, 2021
Revision received June 26, 2021
Accepted October 25, 2021

Editor-in-Chief: Stephen M. Camarata
Editor: Christos Salis

https://doi.org/10.1044/2021_JSLHR-21-00017

ABSTRACT

Purpose: Over the last decades, many studies have documented the clinical potential of nonword repetition (NWR) tasks for detecting developmental language disorder in mono- (MON) and bilingual (BIL) children by unveiling their difficulties in short-term memory and phonological accuracy. However, the precise nature of the nonwords to be used and the best scoring methods remain under debate. Some authors (e.g., Gutiérrez-Clellen & Simon-Cereijido, 2010) support the use of "language-specific" nonwords designed for a given test language in standardized tests. Other authors (e.g., Chiat, 2015) advocate the use of "crosslinguistic" stimuli, thus allowing assessment independently of the languages spoken by the child.
Method: This research note compares two language-specific tasks (French vs. Portuguese) and a crosslinguistic NWR task in a population of 5- to 7-year-old MON and BIL children. Group comparisons (children with vs. without developmental language disorder; MON vs. BIL children), an error analysis, sensitivity and specificity calculations (assessed according to the recommendations of Plante and Vance, 1994, and Youden, 1950) are reported.
Results: All three tasks significantly differentiate children with and without developmental language disorder with large effect sizes but did not show an effect for bilingualism, which is encouraging for the BIL assessment. As expected, an influence of children's age and length and complexity of the stimuli was also found. The language-specific French task was found to be the most sensitive (max. 88%) and specific (max. 92%); the crosslinguistic task also reached good accuracy percentages for the BIL group (max. 82% sensitivity and 84% specificity).
Conclusion: This research note adds to the evidence that NWR tasks are promising tolls for the identification of MON and BIL children with developmental language disorder.
Supplemental Material: https://doi.org/10.23641/asha.19111628

Language assessment of bilingual (BIL) children is a major challenge in speech and language therapy for a number of reasons: (a) there are few or no standardized tests in many languages (Williams & McLeod, 2012); (b) speech and language therapists often do not speak one of the children's languages and (c) linguistically trained interpreters are hard to find (Boerma & Blom, 2017); and, last but not least, (d) most tests are not normed for BIL children (de Almeida et al., 2017).

One solution is to use tasks that are not or only minimally affected by the child's prior knowledge or experience of a specific language (Coady & Evans, 2008). The task most often mentioned in the literature and included in many clinical tests is nonword repetition (NWR): Children have to repeat word-like syllable sequences that are pronounceable

Correspondence to Salomé Schwob: salome.schwob@unine.ch.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

but do not mean anything in their language, e.g., /klipafu/. Failure on an NWR task is recognized as a marker of developmental language disorder (DLD) in monolingual (MON) and BIL populations (Coady & Evans, 2008; Girbau & Schwartz, 2007; Schwob et al., 2021). The presence and nature of errors in the NWR task are thought to reflect difficulties in short-term verbal memory and phonological accuracy (dos Santos & Ferré, 2018). Nonwords with high phonological complexity are particularly difficult for DLD children whether they are MON or BIL (dos Santos & Ferré, 2018; Estes et al., 2007). Nonword length also appears to be a criterion for difficulty impacting both children with DLD and with a typical language development (TD; e.g., Marini et al., 2017).

Although NWR task scores are arguably little affected by a child's prior lexical and syntactic knowledge, they are still influenced by the phonological specificities of the test language. A child is more likely to repeat nonwords correctly if they share phonological properties of his or her own language (Boerma et al., 2015). Using an NWR created according to the phonological structure of a specific language thus produces controversial results, which can, at least partly, be explained by methodological differences (Boerma & Blom, 2017). Indeed, several studies (for a meta-analysis, see Schwob et al., 2021) report that scores obtained by BIL children in language-specific NWR tasks depend on their exposure of the test language, thus diminishing their discriminative power for DLD. However, since some authors (e.g., de Almeida et al., 2017) still obtain high specificity and sensitivity values, the language-specific NWR task may be regarded as discriminative even for BIL children.

In order to avoid biases related to language exposure and language properties, a "crosslinguistic" (CL) NWR task has been devised by Chiat and colleagues (Chiat, 2015; Chiat & Polisenská, 2016). This task includes nonwords based on frequent syllable structures (consonant–vowel, CV, units such as /ba/) and frequent phonemes of the world's languages (see Chiat, 2015), in order to avoid disadvantaging speakers of specific languages. Boerma et al. (2015), among others, evaluated the clinical value of this task in MON Dutch and BIL children with Dutch as a second language aged 5–6 years, and the results seem promising, since bilingualism did not affect performance on this task and provides percentages of 83% sensitivity and 93% of specificity for BIL children. However, due to its "CVCVCV" design, this task taps mostly into phonological memory and cannot reveal difficulties with specific complex phonological structures (e.g., consonant clusters) that are typical for children with DLD.

In addition to the type of stimuli, other factors are known to influence performance on NWR tasks. The scoring method is often reported to be influential. Interestingly, the number of correct whole words can be more specific and sensitive than a more detailed count of the number of correct phonemes (Boerma et al., 2015). Age is also mentioned as an influencing factor: Younger children have more difficulty repeating nonwords (de Almeida et al., 2017).

Here, we set out to compare the clinical potential of three different NWR tasks (two language-specific ones and one CL) in French- and Portuguese-speaking children (a frequent language combination in Western Switzerland; Federal Statistical Office, 2020) aged 5;0 (years;months) to 7;11. We will also compare different scoring methods (words vs. phonemes), perform an error analysis, and analyze potential effects of further factors relating to the child (age) and the stimuli (syllable length, phonological complexity).

Method

Participants

Sixty-one children aged 5;0 to 7;11 living in the French-speaking part of Switzerland and with no history or suspicion of developmental disorders or socio-affective or hearing impairment participated. According to language exposure data collected via a parental report (based on Tuller, 2015), 31 children were exposed to French at least 85% of awake time at the time of the testing and thus classified as MON. The remaining 30 BIL children had a minimum of 25% exposure to Portuguese at time of the study (and no more than 15% exposure to a third language). Among these BIL children, 19 have French as their dominant language and 11 have Portuguese; 12 are successive BIL and 18 are simultaneous BIL. Thirty-one of the children had a confirmed DLD (17 BIL and 14 MON); they had been diagnosed and followed by a speech and language therapist for at least 3 months to attest to the persistence of difficulties. The remaining 30 children (17 MON and 13 BIL) were reported to have a TD, were never identified as having a DLD, and had never received speech and language therapy.

All children successfully passed an audiometric screening (audibility thresholds ≤ 20 dB at 0.5, 1, 2, and 4 kHz) and a nonverbal intelligence screening using the matrices subtest of the Wide-Range Intelligence Test by Glutting et al. (2000). A parental questionnaire was used to elicit the level of education (scales from 1 to 5: 1 = elementary school, 2 = high school, 3 = apprenticeship, 4 = high school graduation, 5 = high school). An additional 11 children participated but were excluded because they failed the audiometry (n = 3) or the nonverbal intelligence (n = 2) screening or did not fit the exposure criteria (n = 2) or due to a technical problem at the time of testing (n = 4).

Children with and without DLD in MON and BIL groups were matched individually on age (median age MON TD: 72 months; MON DLD: 71.5; BIL TD: 87; BIL DLD: 76), H = 5.19, p = .158; nonverbal IQ (median nonverbal IQ MON TD: 50; MON DLD: 52.5; BIL TD: 70;

BIL DLD: 45), $H = 3.71$, $p = .294$; and gender, X^2 (3, $n = 61$) = 1.05, $p = .789$; 37 girls and 24 boys), and these factors did not differ significantly between the four groups. However, we found a significant difference in parental education between the BIL DLD ($Mdn = 2.5$) and the MON TD children ($Mdn = 4.75$), $H = 13.53$, $p = .004$; the other groups did not differ significantly in this respect (median MON DLD: 3; BIL TD: 3).

Recruitment Procedure

Participants were recruited through their schools and speech and language therapists. Families interested in participation received an information sheet, a parental report and an informed consent form, validated by the local ethics commission (Swiss ethics Vaud; project no.: 2017–01881, authorization granted on 07.02.2018). Most of the evaluation sessions took place in the family home or at school. As part of a larger PhD research, participants participated in three sessions of about 1 hr each, which also included other language tasks. The NWR tasks were always done at the beginning of the session and were administered by a native speaker of the test language who was familiar with language assessment; the first author for French and a BIL Portuguese French student in a master's program in speech and language therapy for the Portuguese.

Tasks

Two language-specific NWR tasks were chosen from standardized tests: Exalang 5–8 (Thibault & Helloin, 2010) for French and Provas de Avaliação da Linguagem e da Afasia em Português (Castro et al., 2007) for Portuguese (only completed by the BIL children). The French task consists of 16 two- or three-syllable items and the Portuguese task is composed of 24 one- to three-syllable nonwords. The two tasks have a phonological complexity that can be described on a 3-gradient scale based on dos Santos and Ferré (2018) and Mateus et al. (2005; see Supplemental Material S9 for details of the different tasks): complexity 0 = simple structure like CV: /ladɔ̃/, /fɛ̃t/; complexity 1 = intersyllabic cluster, beginning of nonword by vowel or consonant–semivowel: /bidwano/, /ɐdɔdɐ/; and complexity 2 = intrasyllabic cluster: /drigɔ̃bir/, /trɛf/. The 16 CL nonwords were chosen from the lists provided by Chiat (2015), including items of two to five simple syllables (CV). The items for this latter task were prerecorded by a foreign (Swedish) speaker with final stress, so as to avoid advantaging the MON language group by using French as a test language.

Scoring Procedure

Two scoring methods were applied by counting different units, that is, (a) the total number of incorrect whole words and (b) the number of incorrect phonemes per word. For the latter, we used the procedure of Dollaghan and Campbell (1998). Thus, each phoneme (consonant or vowel) was scored as incorrect with respect to its target phoneme, taking into consideration its position in the word. For the error analysis, we based our analysis on Kapalkovà et al. (2013) and we categorized our errors according to different levels: phoneme level errors (e.g., consonant substitution: /zilu/ for /zibu/; vowel and consonant additions: /zibubu/ for /zibu/), syllable level errors (e.g., vowel omission: /zib/ for /zibu/; cluster reduction: /lover/ for /flovir/), and word level errors (e.g., regressive or progressive assimilation: /bibu/ for /zibu/; lexicalization: in French /caramel/ for /caranel/; creation, when more than four wrong phonemes: /nikalasa/ for /nugitala/; metathesis: /zubi/ for /zibu/).

Interrater Reliability

A second judge categorized 10% of our sample on the CL and French NWR tasks. Interjudge agreement indices were then calculated using the statistical software SPSS (Version 27.0, IBM Corp., 2020). Near-perfect agreement (according to Landis & Koch, 1977) in Cohen's kappa tests was found for whole word scoring (.98 for French NWR task and 100% for CL NWR task). Excellent interrater reliability (according to Koo & Li, 2016) in terms of intraclass correlation coefficient was found for scoring in terms of number of incorrect phonemes (.99 for CL NWR task and 1 for French NWR task). For categorization in terms of error types (for both tasks), ICCs indices also indicate strong to near-perfect agreement (between .66 and 1; Koo & Li, 2016).

Statistical Analyses

In order to determine whether there were significant differences between the different groups of children, univariate and multivariate linear models and repeated measures analysis of variance (with the Greenhouse–Geisser correction) were conducted using SPSS Version 27 (IBM Corp., 2020), after transformation into ranks according to the method proposed by Conover and Iman (1981) because our date did not follow a normal distribution. For small number of missing data on parental education of MON children ($n = 10$), we performed multiple imputations (Donders et al., 2006) to generate these observations and control for them in our statistical model. We reported effect sizes for the outcomes of clinical and linguistic status, and length and complexity factors by using the partial eta squared (η^2). We follow Cohen (1988) to consider small ($\eta^2 = .01$), medium ($\eta^2 = .06$), and large ($\eta^2 = .14$) effects. We also reported effect sizes for error analysis using Cramer's V ($\leq .2$ = small effect; between .2 and .6 = moderate; $\geq .6$ = large).

Differences in terms of error categories were analyzed using Fisher exact tests. We also generated receiver operating characteristic (ROC) curves (Dunn, 2014) for the different

tasks and different scoring method, from which we derived sensitivity and specificity (which were interpreted according to Plante & Vance, 1994, with values of 80% and above, the test is considered accurate). Finally, Youden (1950) were also calculated in order to determine the optimal threshold to distinguish between participants with and without DLD.

Results

Effect of Disorder

Considering the number of incorrect words produced in the different groups (see Figure 1), we note a large effect of clinical status: children with DLD perform significantly worse than children without DLD in all three tasks when counting the number of incorrect words: French NWR: $F(1, 54) = 65.85$, $p < .001$; Portuguese NWR: $F(1, 25) = 15.5$, $p = .010$; and CL NWR: $F(1, 54) = 18.61$, $p < .001$. We find similar result taking into account the number of incorrect phonemes also with large effect size (see Supplemental Material S1). We note that performance significantly increases with age on the three NWR tasks: CL NWR: $F(1, 54) = 5.78$, $p = .020$; French NWR: $F(1, 54) = 15.49$, $p < .001$; and Portuguese NWR: $F(1, 25) = 18.61$ $p < .001$. Older children (TD or DLD) produce fewer errors (words or phonemes). Scatter plots displaying individual data can be found in Supplemental Materials S2, S3, and S4.

Effect of Bilingualism

Looking at the number of incorrect words produced by MON and BIL children, we find no significant difference between the two groups in the French and CL tasks: French NWR: $F(1, 54) = 3.34$, $p = .073$; CL NWR: $F(1, 54) = 0.18$, $p = .673$. Similar results are observable by taking into consideration the number of incorrect phonemes (see Supplemental Material S1). We also found no interaction between language status and the presence of a disorder in our tasks: French NWR: $F(1, 54) = 0.00$, $p = .986$; CL NWR: $F(1, 54) = 1.14$, $p = .290$. The effect sizes associated with these nonsignificant results on bilingualism are small.

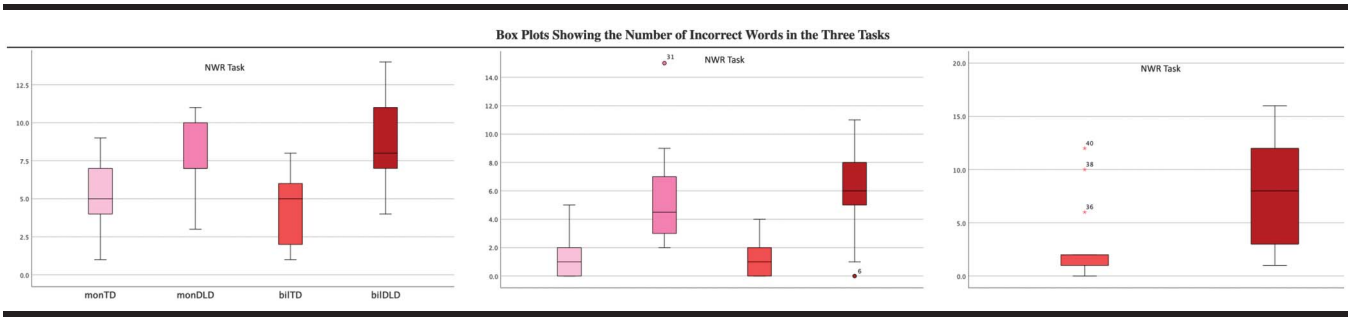
Effect of Stimuli Characteristics

We found significant effects of syllabic length for the three tasks with large effect sizes: CL NWR: $F(2.5, 152.9) = 26.23$, $p < .001$; French NWR: $F(1, 59) = 22.79$, $p < .001$; and Portuguese NWR: $F(1.7, 49.1) = 4.59$, $p = .018$, and phonological complexity for the language-specific tasks with also large effect sizes: French NWR: $F(1.7, 100.4) = 25.97$, $p < .001$; Portuguese NWR: $F(1.7, 47.8) = 22.34$, $p < .001$, in our results (see Supplemental Material S5). We note that long and complex words are more difficult for all children. However, we also find significant interactions between disorder and length in the CL and French tasks, when we consider the number of incorrect phonemes; CL NWR: $F(1.28, 75.74) = 9.41$, $p = .001$; French NWR: $F(1, 59) = 8.98$, $p = .004$. Children with a DLD make more errors than TD children on long words than on words with fewer syllables. No interaction between length and disorder was found for the Portuguese NWR task, $F(1.8, 52.1) = 1.84$, $p = .171$. Finally, we see significant interactions between disorder and phonological complexity in the language-specific tasks: French NWR: $F(1.7, 100.4) = 15.75$, $p < .001$; Portuguese NWR: $F(1.7, 47.8) = 3.82$, $p = .035$. In the French task: DLD children make more mistakes than TD children on complex words. While in the Portuguese task, the difference between DLD and TD children is larger for simple words than for complex words (the simple ones are also the most frequent in this task).

Error Analysis

As noted before, children with DLD produce more errors than children with TD. We now turn to analyzing

Figure 1. Box plots showing the number of incorrect words in the three tasks. CL = Crosslinguistic task; TD = children with a typical language development; DLD = children with a developmental language disorder; mon = monolingual children; bil = bilingual children; NWR = nonword repetition. The repetition tasks are taken from the Exalang (Thibault & Helloin, 2010) and Provas de Avaliação da Linguagem e da Afasia em Português (Castro et al., 2007) test batteries.



whether the distribution of errors differs between the two groups (see Supplemental Material S6). For the NWR task in French, we find significant differences in the proportion of errors between TD and DLD children at the phoneme ($p < .001$), syllable ($p = .040$), and word ($p < .001$) levels with moderate and large effect sizes. For the phoneme level, we find significant differences in the proportion of errors for consonant ($p < .001$) and vowel substitutions ($p = .007$) and for consonant ($p = .031$) and vowel additions ($p = .007$). For the syllable level, we find significant differences in the proportion of errors of vowel omissions ($p = .038$) and cluster reduction ($p = .023$). For the word level, we note differences only for assimilation errors ($p = .009$). We note statistical differences in Portuguese NWR for only two levels (phoneme: $p = .001$; word: $p = .047$) with moderate and large effect sizes. We note differences in consonant ($p = .026$) and vowel substitutions ($p = .010$), for vowel additions ($p = .002$) and for lexicalizations ($p = .025$). In the CL task, although there are many errors, we observe no significant differences by level (phoneme: $p = .224$; syllable: $p = .461$; word: $p = .116$). However, the effect sizes found for the different levels are moderate.

Discriminative Value

Finally, we examined the discriminative value of the tasks. After generating ROC curves (Dunn, 2014) for the different tasks and different scoring method (see Supplemental Material S7), we calculated Youden's indices for all possible thresholds. We note that only the French task (with scoring in number of incorrect words) provides sensitivity and specificity scores (see Table 1) that have good discriminant accuracy (Plante & Vance, 1994) with 88% of sensitivity and 92% of specificity.

We then separated the MON and BIL groups in the two tasks that were used in both and calculated the discriminative values according to linguistic status (see Table 1 and Supplemental Materials S7 and S8). Considering Youden (1950), we note that the two tasks tend to be more discriminating for the BIL population than for the MON population in our research. We also observe that the CL task obtains sensitivity (82%) and specificity (76%) percentages approaching the Plante and Vance (1994) standards, and the diagnostic accuracy of this task improves for the group of BIL children with a count of incorrect phonemes (82% of sensitivity and 84% of specificity).

General Discussion and Conclusions

This research note assessed the diagnostic accuracy of NWR tasks in identifying developmental disorders in BIL and MON school-age children speaking French and Portuguese. In line with other recent studies (e.g., Kapalková et al., 2013), the first analyses show that the three tasks distinguish the group of children with DLD from children with TD on a group level. No differences were, however, found between the MON and BIL groups, which is encouraging for the use of these tasks in BIL language assessment (Antoniejevic-Elliott et al., 2019). These results can partly be explained by the phonologies of French and Portuguese, which are very similar (including potential sources of difficulty for children, such as the presence of nasal vowels and consonant clusters, Kihm, 2016). However, in the French NWR task, we observe a tendency: MON children seem to perform slightly better than BIL children, which indicates that we should still be cautious in using language-specific NWR task even if the two languages shared by BIL children are close.

Table 1. Discriminating values of tasks according to monolingual and bilingual children with a typical language development and children with a developmental language disorder.

Tasks	Scoring method	Population	Sensitivity	Specificity	Threshold	Youden index
French NWR	words	all	88%	92%	2.5	0.80
French NWR	phonemes	all	77%	100%	5.5	0.77
CL NWR	words	all	82%	76%	6.5	0.59
CL NWR	phonemes	all	77%	73%	11.5	0.50
French NWR	words	MON	92%	76%	2.5	0.69
French NWR	words	BIL	88%	92%	2.5	0.80
French NWR	phonemes	MON	100%	70%	2.5	0.70
French NWR	phonemes	BIL	88%	100%	5.0	0.88
CL NWR	words	MON	85%	64%	6.5	0.50
CL NWR	words	BIL	82%	76%	6.5	0.59
CL NWR	phonemes	MON	85%	58%	8.5	0.44
CL NWR	phonemes	BIL	82%	84%	11.5	0.67
Portuguese NWR	words	BIL	82%	76%	2.5	0.59
Portuguese NWR	phonemes	BIL	82%	76%	44.5	0.59

Note. $N = 61$ ($n = 31$ children with a developmental language disorder and $n = 30$ children with a typical language development). The repetition tasks are taken from the Exalang (Thibault & Helloin, 2010) and Provas de Avaliação da Linguagem e da Afasia em Português (Castro et al., 2007) test batteries. NWR = nonword repetition task; CL = crosslinguistic task; MON = monolingual children; BIL = bilingual children.

We also looked at other factors that may influence NWR task scores. We observe that age, length, and phonological complexity are influencing factors for all three tasks, a result consistent with previous studies (e.g., de Almeida et al., 2017). The use of complex materials, that is, stimuli that vary in length and complexity, seems to be a good way to dissociate children with and without DLD. While the CL task does not test language-specific complex elements of phonological structures, we observe that it is in this task where TD and DLD children make the most errors, probably because it contains the longest items (up to five syllables) and involves phonological memory. It is important to note that the use of a prerecorded voice of a Swedish speaker for the CL NWR task certainly made the task more complex for the different groups of children (MON and BIL TDs and DLDs), which may also explain the presence of many errors. However, in this task, the number of errors produced is large, but the proportions are not significantly different between TD and DLD children. In the two language-specific tasks, whose items vary in length and complexity, we note significantly different proportions depending on the clinical status of the children with moderate and large effect sizes. In particular, we observed more frequent errors (especially vowels additions and substitutions) in children with DLD in these two tasks.

Finally, we looked at the discriminative value of the tasks. We tested the use of two types of scoring methods: the number of incorrect words and the number of incorrect phonemes since the methods of analysis seem to have an impact on diagnostic quality (e.g., Boerma et al., 2015). The best option seems to be to use the number of incorrect words since its yields better Youden (1950), but only the language-specific French task provides a good discriminatory accuracy (Plante & Vance, 1994). In fact, this task comes from a test whose psychometric qualities were controlled (Thibault & Helloin, 2010). The Portuguese task, on the other hand, seems less sensitive and specific. Interestingly, the best threshold found in this study does not correspond to the norm values, which yield even worse Youden's indices (see Schwob & Skoruppa, 2021), highlighting the weak discriminative power of this task that has been normed with only a few individuals (de Almeida et al., 2017). Finally, the CL task also seems promising, particularly for BILs, especially since it is usable for most languages (Antoniejevic-Elliott et al., 2019). Moreover, this task is particularly interesting for languages that do not have standardized tools available on the market. Since the items vary considerably in length, but not in complexity, this task appears to assess verbal short-term memory more than phonological representation. We note, however, that at the age of our participants, the specific phonological complexities of the Portuguese and French language have an influence on the children's scores and should thus be taken into consideration.

In addition, when we are interested in length effects, a count by the number of incorrect phonemes was more sensitive because it allowed us to identify different strategies according to the groups of children: children with a DLD struggle to repeat long nonwords and produce the beginning or do not repeat at all; conversely, children with TD tend to try to produce the nonword (even if it is long) with some phonemes errors. These differences are not noticeable when we count the number of incorrect words, as both responses yield a "0" score.

In this research note, we have focused only on NWR tasks, although it is widely acknowledged that the variety of assessment tasks also contributed to improving diagnostic power (Boerma & Blom, 2017). Thus, it would be interesting to replicate this study with larger groups of participants and we should also be interested in combining NWR tasks with other tools that have been identified as promising (e.g., dynamic assessment: Hasson & Botting, 2010; narrative tasks: Boerma & Blom, 2017). However, as a first step, we were able to show in this research note that the use of NWR tasks to assess a MON or BIL French–Portuguese child is appropriated and thus merits its place in the oral language assessment.

Acknowledgments

The authors would like to thank all participants and their families for taking part in this research, as well as colleagues, students, and members of institutions who allowed them to contact their families. They would also like to thank their Portuguese-speaking student assistants for their contribution to this study. Parts of a prior version of this research note were translated from French by the free version of the software DeepL.

References

Antonijevic-Elliott, S., Lyons, R., O'Malley, M. P., Meir, N., Haman, E., Banasik, N., Carroll, C., McMenamin, R., Rodden, M., & Fitzmaurice, Y. (2019). Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics, 34*(4), 293–311. https://doi.org/10.1080/02699206.2019.1637458

Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible. *Journal of Communication Disorders, 66*(1), 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A Quasi-Universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language, and Hearing Research, 58*(6), 1747–1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058

Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R., & Coltheart, M. (2007). *Provas de avaliação da linguagem e da*

afasia em Português [Tasks for the assessment of language processing and aphasia in Portuguese, PALPA-P]. CEGOC.

Chiat, S. (2015). Non-word repetition. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 95–122). Multilingual Matters. https://doi.org/10.21832/9781783093137-008

Chiat, S., & Polisenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research, 59*(5), 1179–1189. https://doi:10.1044/2016_JSLHR-L-15-0293

Coady, J., & Evans, J. (2008). Uses and interpretations of nonword repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 43*(1), 1–40. https://doi.org/10.1080/13682820601116485

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Conover, W., & Iman, R. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*(3), 124–129. https://doi.org/10.1080/00031305.1981.10479327

de Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., Zebib, R., & Barthez, M.-A. (2017). Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism, 7*(3–4), 331–358. https://doi.org/10.1075/lab.15019.alm

Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Donders, A., Van der Heijden, G., Stijnen, T., & Moons, K. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59*(10), 1087–1091. https://doi.org/10.1016/j.jclinepi.2006.01.014

dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition, 25*(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Dunn, G. (2014). Statistics in psychiatry. In M. Lovric (Ed.), *International encyclopedia of statistical science* (pp. 1136–1138). Arnold.

Estes, K. G., Evans, J., & Else-Quest, N. (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*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Federal Statistical Office. (2020). *Most frequent main languages*. https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.assetdetail.11947533.html

Girbau, D., & Schwartz, R. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders, 42*(1), 59–75. https://doi.org/10.1080/13682820600783210

Glutting, J., Adams, W., & Sheslow, D. (2000). *Wide Range Intelligence Test*. Wide Range.

Gutiérrez-Clellen, V., & Simon-Cereijido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English speaking children: Does the language of assessment matter. *Learning Disabilities Research & Practice, 25*(10), 48–58. https://doi.org/10.1111/j.1540-5826.2009.00300.x

Hasson, N., & Botting, N. (2010). Dynamic assessment of children with language impairments: A pilot study. *Child Language Teaching and Therapy, 26*(3), 249–272. https://doi.org/10.1177/0265659009349982

IBM Corp. (2020). *IBM SPSS Statistics for Macintosh (Version 27.0)* [Computer software].

Kapalková, S., Polisenská, K., & Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: Novel scoring methods. *International Journal of Language & Communication Disorders, 48*(1), 78–89. https://doi.org/10.1111/j.1460-6984.2012.00189.x

Kihm, A. (2016). *Fiche langue le portugais européen*. CNRS website. https://lgidf.cnrs.fr/sites/lgidf.cnrs.fr/files/images/PORTUGAIS_24_06_19_A4.pdf

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi:10.1016/j.jcm.2016.02.012

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Marini, A., Ruffino, M., Sali, M. E., & Molteni, M. (2017). The role of phonological working memory and environmental factors in lexical development in Italian-speaking late talkers: A one-year follow-up study. *Journal of Speech, Language, and Hearing Research, 60*(12), 3462–3473. https://doi.org/10.1044/2017_JSLHR-L-15-0415

Mateus, M. M., Falé, I., & Freitas, M. (2005). *Fonética e fonologia do Português* [Portuguese phonetics and phonology]. Universidades Aberta.

Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P., & Skoruppa, K. (2021). Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 64*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Schwob, S., & Skoruppa, K. (2021). Identifying developmental language disorders in French-Portuguese bilingual children: Resources available to speech-language pathologists. *Canadian Journal of Speech-Language Pathology and Audiology, 45*(4), 251–266. https://cjslpa.ca/files/2021_CJSLPA_Vol_45/No_4/CJSLPA_Vol_45_No_4_2021_1230.pdf

Thibault, M. P., & Helloin, M. C. (2010). *Exalang 5–8*. Happyneuron.

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 301–331). Multilingual Matters. https://doi.org/10.21832/9781783093137-013

Williams, C., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology, 14*(3), 292–305. https://doi.org/10.3109/17549507.2011.636071

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3