OPEN ACCESS

The Assessment of Phonological Short-term Memory via "Quasi-universal" vs. Language-specific Nonwords as a Clinical Marker for Language Impairments

Eugen Zaretsky — Benjamin P. Lange — Christiane Hey

ABSTRACT:
Background. In recent years, limited phonological short-term memory has become a widely recognized clinical marker for language(-related) impairments both for monolingual and bilingual children. Usually, it is assessed by the repetition of nonwords, sentences, and digit spans. However, in the case of bilingual children, its assessment has been shown to be inaccurate due to the influence of bilinguals' first language phonotactics and their limited command of the second language. **Methods**. Monolingual (n = 712) and bilingual (n = 1,004) German preschoolers (age 4;0–4;11 years) were compared with each other with respect to their German language skills and performance in German-based and "quasi-universal" nonword repetition tasks (that is, items following the phonotactic rules of German vs. many world languages). Associations of both kinds of nonword repetition tasks with children's language impairments were quantified. **Results**. German language skills of the bilinguals were weaker than those of the monolinguals. Whereas the bilinguals scored significantly lower than the monolinguals in German-based nonwords, there were no considerable differences between these subgroups in quasi-universal items. Poor performance in tasks employing both German-based and quasi-universal nonwords was significantly associated with language impairments. In contrast to quasi-universal nonwords, weak performance in German-based items was more strongly associated with limited German language skills than with language impairments. **Conclusions**. Because nonword repetition tasks were designed to identify children with language impairments, and not those with a weak command of German, quasi-universal nonwords appear to be more appropriate for language tests than language-specific items.

KEY WORDS:
nonword repetition, phonological short-term memory, bilingualism, language assessment, German phonotactics

1. INTRODUCTION

Nonword repetition is a complex skill involving both short-term processing in the phonological memory and sometimes also long-term vocabulary, phonotactics, and grammar knowledge (Gathercole, 1995; Munson et al., 2005). Along with tasks involving the repetition of sentences and digit spans, nonword repetition is often utilized to measure phonological short-term memory (PSTM) in numerous language tests, such as the German "Kindersprachscreening" ["Language screening for children", KiSS] (Holler-Zittlau et al., 2011) and "Sprachentwicklungstest für drei- bis fünfjährige Kinder" ["Language development test for three- to five-year-old children", SETK 3–5] (Grimm, 2001). PSTM, quantified by performance in nonword repetition

tasks, constitutes a strong predictor for the first (L1) and second (L2) language acquisition both in typically developing and in impaired children (Adlof & Patten, 2017; Casserly & Pisoni, 2013). The simplicity of the test procedure and the high validity of the test results, especially their predictive validity for language(-related) impairments (which we are framing in this article as language impairments with or without comorbidities such as auditory processing disorder, autism spectrum disorder), led to a very wide use of the respective tests in school enrollment examinations, everyday clinical practice, and beyond. However, repetition tasks with a high linguistic load, that is, tasks that strictly follow regularities of a certain language, can disadvantage bilingual children who are not yet proficient in their L2 if the respective items are based on this L2. It is also unknown whether performance in PSTM tasks with a high or low linguistic load can better predict (or is more strongly associated with) children's language impairments.

Up to now, the large majority of studies have focused on PSTM tasks with a high linguistic load. Many language-related impairments and disorders were shown to be associated with a weak performance in such repetition tasks. For instance, children born preterm perform worse in PSTM tasks than children born full term (Gresch et al., 2018). Late talkers score lower than their typically developing peers (Rujas et al., 2017). The same is valid for children with attention deficit hyperactivity disorder (Orban et al., 2018), intellectual disabilities (van Wingerden et al., 2017), developmental language disorder (Reichenbach et al., 2016), dyslexia (Kuppen & Goswami, 2016), auditory processing disorders (Kiese-Himmel & Nikisch, 2016), and hearing disorders (Sundström et al., 2018). Predispositions for language disorders (e.g., clustering of such disorders in one's family) also appear to be linked to low performance in PSTM tasks (Truong et al., 2016).

Because bilingual children often lag behind in the acquisition of articulation, grammar, and vocabulary in their L2 compared to monolinguals, they often score lower in PSTM tasks than their monolingual peers (e.g., Armon-Lotem & Meir, 2016). However, the same PSTM tasks, irrespective of whether the children are monolingual or bilingual and often with the same cut-off criteria, are used by teachers and pediatricians to differentiate between children with and without language impairments (Li'el et al., 2019). For instance, most school enrollment tests in Germany contain nonword repetition items (e.g., "Sozialpädiatrisches Screening für Schuleingangsuntersuchungen" ["Socio-Pediatric Screening for the School Enrollment Examination", SOPESS]; Petermann et al., 2009). Unexpectedly, bilingual children are sometimes not even mentioned in the respective test manuals (e.g., "Screening des Entwicklungsstandes bei Einschulungsuntersuchungen" ["Developmental Screening for the School Enrollment Examinations", S-ENS]; Döpfner et al., 2005), although in some German states (e.g., Hesse; Ruhland et al., 2022) migrants make up more than a half of all preschool children.

As long as bilinguals' limited competence in L2 remains unconsidered in such repetition tests, PSTM assessment in bilingual children is prone to a lower accuracy in identifying language impairments (Tuller et al., 2018) and to misinterpretations due to the phonotactic influence of the children's L1 (Duncan & Paradis, 2016). For instance, Turkish children acquiring German as L2 tend to omit or simplify consonant

clusters in German-based nonwords in the word final position because such clusters do not occur in this position in the Turkish language (Zaretsky et al., 2013). However, such L2 transfers are usually not taken into account in German repetition tests and are qualified as PSTM errors. Furthermore, nonwords in widely used language tests often resemble existing words or even rhyme with them. Also, they contain existing affixes and thus presuppose certain vocabulary and morphology skills. Therefore, even in the case of nonwords, performance in PSTM tasks depends on the children's competence in the language the nonwords are based on (Jones, 2016). Another popular PSTM task that is widely considered a good clinical marker for language impairments—the repetition of sentences (Archibald & Joanisse, 2009)—also presupposes phonology, vocabulary, and grammar skills in the language the tasks are based on. Due to this high linguistic load, results of bilingual children in repetition tasks often appear deficient in comparison with those of monolinguals (Meir & Armon-Lotem, 2017).

In sum, because PSTM tasks are widely used as a clinical predictor for language(-related) impairments (Bishop et al., 2016), bilinguals' weak performance in such tasks can be interpreted not only in terms of weak L2 skills but also from a medical perspective. That is, bilinguals with too low scores of correct answers on repetition tasks can be classified as language impaired although their poor performance results solely from the lack of the language input in their L2 (Paradis et al., 2013).

Nonwords that are structured to conform to the phonotactics of as many languages as possible could be the answer to the question of how to differentiate between weak L2 skills and language impairments in bilinguals, that is, how to avoid a disadvantaging of bilingual children. "Quasi-universal" nonwords (QUNW) proposed by Chiat (2015) are one of very few published lists of items for PSTM assessment that consider the phonotactic regularities of many world languages (for another list, see Hamman & Ibrahim, 2017). Performance in QUNW (Chiat, 2015) was shown to differ significantly between (Dutch speaking) bilinguals with and without a developmental language disorder (Boerma & Blom, 2017).[1] However, hardly any studies have been published so far by other research groups on these nonwords. Hence, it remains unclear, whether monolingual and bilingual children with and without language impairments (including those with and without comorbidities, more broadly defined than in the article by Boerma and Blom (2017)) as well as children with and without language-related educational needs (limited language skills resulting from insufficient language input) can be differentiated from each other by means of QUNW. It is also unknown whether QUNW or language-specific nonwords are more appropriate for the identification of children with language impairments and educational needs. Also, the results of Boerma and Blom (2017) have not yet been replicated for the German language.

1 Although Boerma and Blom (2017) use the term "language impairment" in their study, their definition—no hearing disorders, normal intelligence—comes close to the traditional definition of developmental language disorder (Bishop et al., 2017), that is, impairment without comorbidities.

The present study is the first attempt

- to compare nonword repetition performances of bilingual German preschoolers with those of German monolinguals in terms of a disadvantaging of bilingual children ("culture fairness" of nonword repetition tasks),
- to analyze the capability of language-specific nonwords and QUNW to confirm diagnostic distinctions regarding monolingual and bilingual children's educational needs in acquiring German and language impairments,
- to test QUNW (Chiat, 2015) for another language than Dutch.

The definition of language impairments includes both the developmental language disorder (Bishop et al., 2017) and language impairments with various comorbidities such as hearing disorders, Down syndrome or attention deficit hyperactivity disorder (Neumann et al., 2009). In previous comparable studies (Boerma & Blom, 2017; Hamman & Ibrahim, 2017), only language impairments without comorbidities were addressed. Thus, the study presented here refers to a much larger subgroup of children than previous studies did.

In the present study, children's language impairments were assessed by two gold standards: (a) one of very few validated German language tests that subdivide children into those with educational needs and language impairments (KiSS), (b) questionnaires for parents and kindergarten teachers, with several items on children's and their relatives' language impairments as well as children's participation in speech-language therapies. For children's educational needs in acquiring German, only KiSS results were used as the gold standard because the sensitivity and specificity of KiSS for educational needs can be considered very good and are much higher than those for language impairments (Neumann & Euler, 2010).

According to several studies conducted for other languages than German, PSTM tasks with a low language-specific load do not disadvantage bilingual children as much as those with a high load do (Meir, 2017; Yoo & Kaushanskaya, 2012). Therefore, it was hypothesized that bilingual children would yield better results in QUNW than in German-based tasks. It could not be excluded that bilinguals would even outperform monolinguals in QUNW due to the bilingual executive function advantage (Blom et al., 2014; Sörman et al., 2017). This term refers to a better PSTM performance in bilinguals compared to monolinguals as a result of the double pressure imposed by bilingualism on the executive functions. Also, because of the relatively low language-specific (here: German-specific) load in QUNW, performance in such tasks was expected to be more strongly associated with children's language impairments than with the command of the German language. On the contrary, performance in German-based items was expected to be more strongly associated with German language skills. To sum up, it was hypothesized that a weak performance in QUNW can identify language impairments on the same level as performance in German-based items without disadvantaging children with limited German language skills.

OPEN ACCESS

2. METHOD

The study was approved by the ethics committee of the Marburg University Hospital (approval number 117/16, 06. Sept. 2016).

2.1 PARTICIPANTS

In this prospective study, children aged 4;0–4;11 years were recruited in German kindergartens, without any further inclusion criteria except a declaration of consent signed by parents. At this stage, no exclusion criteria were applied.

Out of 3,315 eligible children, 1,790 were recruited (54%). These children lived both in cities, towns, and in the rural area. Because 39 children out of the 1,790 (2%) showed poor compliance during testing, they were excluded from the study. Further 35 children (2%) could not be tested for organizational reasons such as scheduling conflicts.

The remaining 1,716 children (age 4;0–4;11 years, M = 53.29 months (SD = 3.35); n = 890 boys, 52%, n = 826 girls, 48%) were tested in 328 kindergartens. All children were classified as monolinguals (those speaking the German language only) or bilinguals (children speaking two languages including German) on the basis of KiSS questionnaires for parents. Out of the 1,716 children, 712 (42%) were monolingual German speakers (who were non-immigrants in most cases). All other children (n = 1,004, 59%) were raised bilingually, with the largest subgroups speaking Russian (n = 175, 10%), Turkish (n = 154, 9%), and Arabic (n = 126, 7%) at home. Most of them were immigrants, that is, they and/or their parents moved to Germany from another country. According to a Mann-Whitney U test, monolinguals and bilinguals were of comparable age in months (Z = –0.57, p = .566).

2.2 MATERIALS

All children were tested with the validated language screenings "Kindersprachscreening" (KiSS; Holler-Zittlau et al., 2011) and "Sprachscreening für das Vorschulalter" ["Language screening for the preschool age", SSV]; Grimm, 2003) as well as with a list of non-validated QUNW (Chiat, 2015). Both KiSS and SSV contain German-based nonwords designed for PSTM assessment. Although one list of such nonwords would have sufficed, two lists were employed to assure the replicability of results.

2.2.1 "KINDERSPRACHSCREENING" ["LANGUAGE SCREENING FOR CHILDREN", KISS]

KiSS consists of subtests on speech comprehension, vocabulary, grammar, articulation, and PSTM. The latter is assessed by 20 German-based nonwords (e.g., *Triser*, *Verklasenaft*) and two sentences. The version of KiSS used in this study was an extended version of the most up-to-date KiSS version, KiSS.2 (Holler-Zittlau et al., 2011), in that it contains some additional items in the subtest on nonword repetition and in questionnaires for parents and kindergarten teachers. Additional nonwords were taken from the previous validated KiSS version, KiSS.XL (Neumann & Euler,

2009). Because KiSS.2 contains only four nonwords, it was not directly comparable to SSV with its 18 items and to QUNW by Chiat (2015) with 16 items. Therefore, the extended KiSS version with 20 nonwords from KiSS.XL was utilized.

Cut-off criteria are available for two age groups: 4;0–4;5 and 4;6–4;11 years. KiSS classifies children as those needing (ED) or not needing (NED) additional educational assistance as well as needing (MED) or not needing (NMED) additional medical assistance in acquiring German. The deficient German language competence of ED children results from a lack of German language input or, more generally, from unfavorable sociodemographic conditions for the acquisition of the German language. In terms of the CATALISE study, these are children who had insufficient exposure to the language used by the school or community to be fully fluent in it (Bishop et al., 2017). In the large majority of cases, such children have a migrant background. Their results correspond to those of the lowest 16% of the norming sample, that is, lie about one standard deviation below the average values of (monolingual Germans') total scores of correct answers. The best way to improve German language skills of ED children is participation in German language courses that are usually offered by kindergartens (Holler-Zittlau et al., 2011). No medical examinations or referrals are needed for the participation in such language courses.

On the contrary, MED children would not benefit from language courses alone. Their results correspond to those of the lowest 5% of the norming sample and/or reflect well-known symptoms of language impairments (e.g., the lack of progress in language acquisition in course of more than six months of kindergarten attendance). MED children are referred to pediatricians for medical examinations and usually have to undergo medical therapies. For instance, an immigrant child with very weak German language skills but without any comorbidities would be classified as ED. But in the case of a KiSS result indicative of a hearing impairment the same child would be additionally classified as MED. The MED definition in KiSS includes all language impairments with and without comorbidities. In terms of the CATALISE study, MED children are those with language disorders associated with various comorbidities as well as children with a developmental language disorder, that is, with no known differentiating condition (Bishop et al., 2017). The sensitivity for the MED/NMED classification in KiSS is only 67%, the specificity 79% (cf. ED/NED: sensitivity 98%, specificity 79%; Neumann et al., 2011).

The classifications ED/NED and MED/NMED refer to the German language skills only, but information from questionnaires for parents on bilinguals' development in their non-German L1 is also considered in the MED/NMED classification because language impairments usually find their reflection in both languages (Bishop et al., 2017). To be classified as NED and NMED, children need to achieve certain total scores of correct answers (e.g., three points in speech comprehension for NED). Monolinguals and bilinguals share the same ED/NED cut-off criteria, whereas some MED/NMED criteria differ for these two subgroups. The classification of children as ED/NED and MED/NMED is unique to KiSS. In almost all other German language tests, results are classified as pass/fail only, that is, not language impaired or impaired.

Following the recommendations of the CATALISE study (Bishop et al., 2016), multiple sources were combined in the language assessment, including questionnaires

for parents and caregivers. KiSS-based questionnaires for parents and kindergarten teachers include items on the quality and quantity of the language input, the child's and her/his relatives' "language-related medical impairments", the child's sociability, and other sociolinguistic/sociodemographic characteristics. For parents of bilingual children, questionnaires were translated into eleven foreign languages, such as English, Turkish, Russian, and Arabic.

2.2.2 "SPRACHSCREENING FÜR DAS VORSCHULALTER" ["LANGUAGE SCREENING FOR THE PRESCHOOL AGE", SSV]

The second language test, SSV (Grimm, 2003), contains tasks on PSTM only. Just like in KiSS, PSTM is assessed using the repetition of German-based nonwords (e.g., *Nabolira*, *Defsal*) and sentences. The sentences were not the subject of this study. German-based nonwords follow the rules of German phonotactics (e.g., the schwa deletion rule that forbids the vowel [ə] in two adjacent syllables), some of them contain German affixes (e.g., *Dilecktich<u>keit</u>*) and resemble existing German words (e.g., *Toschlander*, cf. sausage brand *Deutschländer*). SSV provides cut-off values for four- to five-year-old children. Only total scores of correct answers (correctly reproduced nonwords) were of interest for this study.

2.2.3 QUASI-UNIVERSAL NONWORDS

Additionally, one of the lists of QUNW suggested by Chiat (2015) was utilized: *Zibu*, *Lita*, *Maki*, *Luni*, *Sipula*, *Bamudi*, *Malitu*, *Lumika*, *Zipalita*, *Mukitala*, *Kasulumi*, *Litisaku*, *Sipumakila*, *Tulikasumu*, *Malusikuba*, and *Litapimuti*. None of the lists has ever been validated for German preschoolers.

In QUNW, phonemes that occur in most world languages and do not contain any consonant clusters are used. To avoid language-specific stress patterns, nonwords should be presented in a segmented form, without a clear stress on any of the syllables ("Say *Si-pu-la*."). Thus, instead of comparing phonotactics of thousands of languages, Chiat (2015) developed nonwords that avoid most language-specific phonotactic features: relatively rare phonemes, all consonant clusters, and stress patterns. No cut-off criteria are available.

The length of both QUNW and German-based nonwords varied between two and five syllables (M = 3.5 syllables in QUNW vs. 3.0 in KiSS and 3.6 in SSV). However, due to the absence of closed syllables, QUNW were, on average, shorter than German-based nonwords (M = 7.1 phonemes in QUNW vs. 7.7 in KiSS and 10.2 in SSV).

2.3 PROCEDURE

Kindergarten teachers informed the parents of children of the appropriate age about the study, distributed and collected declarations of consent as well as questionnaires for parents. An examiner interviewed children individually in their kindergartens in a quiet room. Test sessions lasted 30 minutes on average. The examiner explained to each child that some questions such as "What is it?" or "What does it feel like?" would

be asked on the basis of a large picture (KiSS). In PSTM tasks, children had to repeat nonwords and sentences. All test sessions were audio-recorded so that missing or questionable data could be checked later.

2.4 ANALYTICAL APPROACH

First, descriptive statistics on children's total scores of correct answers on all three PSTM tasks and on percentages of ED and MED children in KiSS were calculated for the whole sample.

Because total scores of all tests (KiSS, SSV, QUNW) of the whole sample, of monolinguals, and of bilinguals were not normally distributed according to Kolmogorov-Smirnov tests ($ps < .001$), non-parametric statistical methods were used, where applicable, in the following calculations. Since not all children could complete the tasks, sample sizes varied slightly in the calculations. Children with incomplete results, that is, those without total scores of correct answers in some of the tasks, were not excluded to avoid bias in the data. Dichotomized KiSS results (ED/NED, MED/NMED) were available for all children.

German language competence was analyzed for monolingual compared to bilingual children because it was expected that performance in German-based nonwords would partially depend on German language skills. Dichotomized KiSS results (ED/NED, MED/NMED) were cross-tabulated with the classification "monolinguals/bilinguals". A Chi-squared test was calculated for each of two cross-tables. Significant Chi-squared results would mean that the proportions of ED or MED children were significantly higher in one of the subgroups (probably in bilinguals compared to monolinguals).

Next, results of bilinguals and monolinguals in nonword repetition tasks (KiSS, SSV, QUNW) were compared by Mann-Whitney *U* tests. It was expected that, in case of significantly weaker German language skills in the bilingual subgroup, monolinguals would outperform bilinguals in total scores of correct answers on German-based nonwords, but not on QUNW because performance in the latter probably does not depend much on German language skills.

The effect size in all Mann-Whitney *U* tests was estimated with the probability of superiority index (Grissom & Kim, 2012). It measures the probability with which a randomly selected score of one group is larger (or smaller) than a randomly selected score of the other group. If there is no difference between two groups, the index is $\hat{p} = .5$. The more $\hat{p}$ deviates from .5 towards 1 or 0, the larger the effect size.

Next, differences between ED and NED as well as between MED and NMED children in the repetition of German-based nonwords and QUNW were analyzed by Mann-Whitney *U* tests, with total scores of correct answers on all three PSTM tests as dependent variables. These calculations were conducted separately for monolinguals and bilinguals. A considerable difference between total scores of correct answers on nonwords depending on the ED/NED classification would mean that a poor performance in PSTM tasks is associated with weak German language skills, although such tasks were designed for the detection of language impairments. Following the same rationale, both monolingual and bilingual MED children, that is, children with

language impairments, were expected to score lower in all three repetition tasks than NMED children.

After univariate tests for differences in PSTM test scores depending on KiSS classifications, multivariate tests for associations, namely three classification trees with total scores of correct answers on PSTM tasks as dependent variables (KiSS, SSV, QUNW), were utilized to analyze the hierarchy of interactions of the independent variables ED/NED, MED/NMED, and "monolinguals/bilinguals". In contrast to regressions and comparable methods, classification trees demonstrate hierarchical structures of factors and their interdependence (Bühl, 2018). Classification trees (here, the method "Exhaustive Chi-squared Automatic Interaction Detector" was used) split the respective sample into subgroups depending on the most significant differentiator among the input variables (e.g., ED/NED). For each subgroup of the respective differentiator (e.g., ED and NED), other independent variables are tried out as further differentiators. Thus, hierarchies of factors are built that can be described as tree branches. All three PSTM tasks were expected to be more closely associated with the MED/NMED classification than with the classifications ED/NED and "monolinguals/bilinguals". This means that the MED/NMED classification was expected to have the highest ranking in (or to be the highest branches of) the classification trees. Also, QUNW were expected to be less strongly associated with the ED/NED and "monolinguals/bilinguals" classifications than German-based ones because performance in QUNW probably does not depend much on the German language skills.

The KiSS classification MED/NMED as a gold standard for German-based nonwords and QUNW might be deficient due to the low sensitivity of KiSS for MED children (see above). As an alternative, questionnaires for kindergarten teachers were tried out as a gold standard, namely dichotomous items on children's known "language-related medical impairments" and participation in speech-language therapies. Differences in total scores of correct answers in three repetition tasks (KiSS, SSV, QUNW) depending on respective questionnaire items were analyzed by Mann-Whitney *U* tests, separately for monolinguals and bilinguals, which resulted in a total of 12 calculations (2 questionnaire items × 2 subgroups: monolinguals vs. bilinguals × 3 PSTM total scores). It was expected that QUNW would demonstrate more significant differences between children with and without language impairments than German-based nonwords because children's performance in QUNW was probably less distorted by the influence of German language skills and, thus, QUNW might have measured PSTM more directly.

Because questionnaire items on "language-related medical impairments" may refer to impairments both with and without comorbidities, several more specific calculations on differences in the performance in PSTM tasks depending on children's impairments were carried out. KiSS questionnaires for kindergarten teachers and parents contained information on some already diagnosed medical impairments and children's familial predisposition for language disorders: whether the child has a language delay (first words after the second birthday or later), language disorders in the family etc. Differences in the three PSTM scores (KiSS, SSV, QUNW) depending on these dichotomous medical variables were, again, quantified by Mann-Whitney

U tests. Some of the variables had to be excluded due to low sample sizes in one of the subgroups ($n < 67$ under condition of the effect size $d = 0.5$, error probability $\alpha = .05$, power .80, two-tailed): children's permanent hearing disorders, auditory processing disorders, early or risk birth, and frequent visits to pediatricians. It was hypothesized that QUNW might yield better results in the differentiation between children with and without these impairments than German-based items because the results were probably less distorted by German language skills.

3. RESULTS

On average, children repeated 14.02 (*SD* = 3.92) KiSS nonwords, 9.47 (*SD* = 4.01) SSV nonwords, and 10.96 (*SD* = 2.89) QUNW correctly. Descriptive statistics on dichotomized KiSS.2 results can be found in Table 1. Monolinguals were less often classified as ED than bilinguals according to KiSS ($\chi^2(1, 1{,}716) = 506.33$, $p < .001$). In the MED/NMED classification, the difference between monolinguals and bilinguals did not reach statistical significance ($\chi^2(1, 1{,}716) = 3.04$, $p = .081$).

Monolinguals yielded significantly higher total scores of correctly repeated German-based nonwords in KiSS and SSV than bilinguals (see Table 2). Bilinguals yielded slightly better results in QUNW.

In Table 3, the performance of ED and MED children in the repetition of nonwords was compared to the performance of NED and NMED children, respectively. In all three language tests, total scores of both German-based nonwords and QUNW differed significantly depending on these two classifications (ED/NED, MED/NMED). According to the effect sizes, total scores of German-based items differed more prominently depending on the ED/NED and MED/NMED classifications than QUNW scores, but only in the case of bilinguals. In the case of monolinguals, effect size values were comparable.

	NED & NMED	**ED**	**NED**	**MED**	**NMED**	**ED/MED**	**MED/ED**
All children	834/1,716 (49%)	845/1,716 (49%)	781/1,716 (51%)	290/1,716 (17%)	1,426/1,716 (83%)	253/845 (30%)	253/290 (87%)
Monolinguals	562/712 (79%)	121/712 (17%)	591/712 (83%)	107/712 (15%)	605/712 (85%)	78/121 (65%)	78/107 (73%)
Bilinguals	272/1,004 (27%)	724/1,004 (72%)	280/1,004 (28%)	183/1,004 (18%)	821/1,004 (82%)	175/724 (24%)	175/183 (96%)

TABLE 1: Descriptive statistics on the dichotomized results of the language test "Kindersprachscreening" (KiSS) regarding percentages of children needing or not needing additional educational (ED/NED) or medical (MED/NMED) assistance in acquiring German.

Note. NED & NMED = typically developing children (incl. children not needing additional educational assistance), ED/MED = ED children who were also classified as MED (subgroup of ED), MED/ED = MED children who were also classified as ED (subgroup of MED)

	German-based nonwords: KiSS					German-based nonwords: SSV					Quasi-universal nonwords				
	Z	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*
Monolinguals vs. bilinguals: *U* test	–7.34	< .001	.40	697	995	–4.84	< .001	.43	693	981	–2.16	.031	.47	699	995
Monolinguals vs. bilinguals: *M* (*SD*)	14.92 (3.32) vs. 13.39 (4.18)					10.08 (3.63) vs. 9.04 (4.20)					10.89 (2.64) vs. 11.03 (3.05)				

TABLE 2: Mann-Whitney *U* tests: differences in the total scores of correctly repeated nonwords between monolingual and bilingual children.
Note. KiSS = "Kindersprachscreening", SSV = "Sprachscreening für das Vorschulalter", $\hat{p}$ = probability of superiority index, *n1* = sample size for monolinguals, *n2* = sample size for bilinguals

	German-based nonwords: KiSS					German-based nonwords: SSV					Quasi-universal nonwords				
Monolinguals	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*
ED vs. NED: *U* test	–7.64	< .001	.27	107	590	–6.07	< .001	.32	106	587	–6.17	< .001	.32	110	589
ED vs. NED: *M* (*SD*)	12.09 (4.30) vs. 15.43 (2.82)					7.90 (3.95) vs. 10.47 (3.43)					9.20 (3.23) vs. 11.18 (2.40)				
MED vs. NMED: *U* test	–8.16	< .001	.25	102	595	–5.55	< .001	.33	100	593	–6.26	< .001	.31	104	595
MED vs. NMED: *M* (*SD*)	12.09 (4.30) vs. 15.43 (2.82)					7.90 (3.95) vs. 10.47 (3.43)					9.20 (3.23) vs. 11.18 (2.39)				
Bilinguals	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*
ED vs. NED: *U* test	–12.28	< .001	.25	715	280	–10.69	< .001	.28	701	280	–6.60	< .001	.37	715	280
ED vs. NED: *M* (*SD*)	12.41 (4.18) vs. 15.87 (2.99)					8.13 (4.11) vs. 11.33 (3.49)					10.62 (3.14) vs. 12.06 (2.53)				
MED vs. NMED: *U* test	–8.93	< .001	.29	181	814	–9.32	< .001	.28	174	807	–6.74	< .001	.34	181	814
MED vs. NMED: *M* (*SD*)	10.49 (4.92) vs. 14.03 (3.71)					6.23 (4.22) vs. 9.65 (3.94)					9.45 (3.57) vs. 11.38 (2.80)				

TABLE 3: Mann-Whitney *U* tests: differences in the total scores of correctly repeated nonwords between children needing or not needing additional educational (ED/NED) or medical (MED/NMED) assistance in acquiring German according to the language test "Kindersprachscreening".
Note. KiSS = "Kindersprachscreening", SSV = "Sprachscreening für das Vorschulalter", $\hat{p}$ = probability of superiority index, *n1* = sample size for "fail" (ED or MED), *n2* = sample size for "pass" (NED or NMED)

Classification trees were utilized to analyze hierarchies of associations and interactions between total scores of correct answers on three PSTM tasks and ED/NED, MED/NMED as well as the "monolinguals/bilinguals" classifications (see Table 4, only highest branches). Children's total scores of correct answers in German-based nonwords were most closely associated with their classification as ED/NED, total scores of correct answers in QUNW with the classification MED/NMED. Lower branches of the classification trees (not depicted in Table 4) also revealed considerable differences between German-based nonwords and QUNW. Both in KiSS ($F(1, 868) = 4.43$, $p = .036$) and SSV nonwords ($F(1, 865) = 11.61$, $p = .001$), the next node for the NED children was the "monolinguals/bilinguals" classification, with monolinguals ($M = 15.43$, $SD = 2.82$ in KiSS, 10.47, $SD = 3.43$ in SSV) yielding significantly worse results than bilinguals ($M = 15.87$, $SD = 2.99$ in KiSS, 11.33, $SD = 3.49$ in SSV). The next node for the ED subgroup was the MED/NMED classification both in KiSS ($F(1, 820) = 73.27$, $p < .001$) and SSV ($F(1, 805) = 67.50$, $p < .001$) nonwords. ED children who were classified as NMED outperformed ED children who were classified as MED (KiSS: $M = 13.17$, $SD = 3.69$ vs. 10.53, $SD = 4.72$; SSV: $M = 8.83$, $SD = 3.87$ vs. 6.33, $SD = 4.07$). Thus, classification trees of both German-based tasks (KiSS, SSV) revealed the same structure. In the case of QUNW, the highest node MED/NMED was followed by the node ED/NED for NMED children, $F(1, 1{,}407) = 11.63$, $p = .001$. NMED children who were classified as NED outperformed NMED children who were classified as ED ($M = 11.49$, $SD = 2.47$ vs. 11.00, $SD = 2.86$). The percentage of correctly predicted PSTM task results can be considered satisfactory in all three classification tree models: 88% for KiSS, 86% for SSV, and 92% for QUNW.

	German-based nonwords: KiSS		**German-based nonwords: SSV**		**Quasi-universal nonwords**	
Mean	14.02 (*SD* = 3.92)		9.47 (*SD* = 4.01)		10.96 (*SD* = 2.89)	
n (%)	1,692 (100)		1,674 (100)		1,694 (100)	
	ED/NED		ED/NED		MED/NMED	
	$F(1, 1{,}690) = 337.29$, $p < .001$		$F(1, 1{,}672) = 205.44$, $p < .001$		$F(1, 1{,}692) = 117.36$, $p < .001$	
	ED	NED	ED	NED	MED	NMED
Mean (*SD*)	12.37 (4.20)	15.57 (2.88)	8.10 (4.09)	10.75 (3.47)	9.33 (3.43)	11.29 (2.65)
n (%)	822 (49)	870 (51)	807 (48)	867 (52)	285 (17)	1,409 (83)

TABLE 4: Classification trees with total scores of correct answers in nonword repetition tasks as dependent variables and dichotomous "Kindersprachscreening" (KiSS) results as well as "monolinguals/bilinguals" classification as independent variables (only the highest branches).
Note. SSV = "Sprachscreening für das Vorschulalter", MED/NMED = children needing or not needing additional medical assistance in acquiring German, ED/NED = children needing or not needing additional educational assistance in acquiring German

In Table 5, Mann-Whitney *U* tests for differences in PSTM scores depending on children's known "language-related medical impairments" as well as on their participation in speech-language therapies are presented. These variables served as an alternative gold standard for language impairments. Total scores of QUNW differed more

significantly between children with and without "language-related medical impairments" than German-based nonwords did. In the case of participation in speech-language therapies, no difference was found.

	German-based nonwords: KiSS					German-based nonwords: SSV					Quasi-universal nonwords				
Monolinguals	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*
Language impairments: *U* test	−3.29	.001	.37	55	579	−3.40	.001	.36	53	577	−4.02	< .001	.34	57	579
"Yes" vs. "No": *M* (*SD*)	13.13 (4.24) vs. 15.10 (3.22)					8.40 (4.20) vs. 10.27 (3.57)					9.42 (3.11) vs. 11.05 (2.59)				
Speech-language therapy: *U* test	−4.67	< .001	.29	43	598	−3.40	.001	.34	41	595	−3.99	< .001	.32	45	598
"Yes" vs. "No": *M* (*SD*)	12.09 (4.30) vs. 15.09 (3.19)					8.07 (4.13) vs. 10.16 (3.57)					9.24 (3.18) vs. 11.01 (2.60)				
Bilinguals	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*	*Z*	*p*	$\hat{p}$	*n1*	*n2*
Language impairments: *U* test	−2.34	.019	.41	65	844	−2.29	.022	.041	64	833	−2.85	.004	.40	66	843
"Yes" vs. "No": *M* (*SD*)	12.28 (4.45) vs. 13.68 (3.96)					8.11 (4.06) vs. 9.27 (4.13)					9.94 (3.48) vs. 11.22 (2.87)				
Speech-language therapy: *U* test	−5.25	< .001	.32	75	848	−5.34	< .001	.31	74	835	−5.41	< .001	.31	76	847
"Yes" vs. "No": *M* (*SD*)	10.57 (4.87) vs. 13.64 (4.00)					6.37 (4.44) vs. 9.29 (4.08)					8.76 (3.86) vs. 11.21 (2.90)				

TABLE 5: Mann-Whitney *U* tests: differences in total scores of three nonword repetition tasks and questionnaire items depending on "children's language-related medical impairments" as well as on their participation in speech-language therapies.
Note. KiSS = "Kindersprachscreening", SSV = "Sprachscreening für das Vorschulalter", $\hat{p}$ = probability of superiority index, *n1* = sample size for "Yes", *n2* = sample size for "No"

Calculations on differences in PSTM scores depending on questionnaire items regarding children's language impairments as well as on a familial predisposition for language disorders yielded statistically significant results in most cases. Children with a language delay in their L1 yielded weaker PSTM results in all three tests—KiSS (Z = −9.46, p < .001, $\hat{p}$ = .32, n = 1,526), SSV (Z = −8.29, p < .001, $\hat{p}$ = .34, n = 1,510), and QUNW (Z = −7.79, p < .001, $\hat{p}$ = .35, n = 1,528)—than children with an age-appropriate (German) language acquisition. The same is valid for children with head injuries and/or operations compared to the other children in all three tests except that the result for KiSS was marginally significant (KiSS: Z = −1.91, p = .057, $\hat{p}$ = .45, n = 1,549; SSV: Z = −2.67, p = .008, $\hat{p}$ = .42, n = 1,532; QUNW: Z = −3.20, p = .001, $\hat{p}$ = .41, n =1,551). Children who underwent non-language therapies (any kind of therapy excluding speech and language therapy) also yielded lower results in all three PSTM tasks—KiSS (Z = −2.92, p = .003, $\hat{p}$ = .40, n = 1,532), SSV (Z = −3.08, p = .002, $\hat{p}$ = .39, n = 1,515), and QUNW (Z = −2.55, p = .011, $\hat{p}$ = .41, n = 1,534)—than children who did not undergo such therapies. Children with

cases of dyslexia among relatives scored lower in two out of three PSTM tasks (SSV: $Z = -2.27$, $p = .023$, $\hat{p} = .45$, $n = 1{,}513$; QUNW: $Z = -2.70$, $p = .007$, $\hat{p} = .44$, $n = 1{,}530$). In KiSS, the result did not reach statistical significance ($Z = -1.59$, $p = .113$, $\hat{p} = .46$, $n = 1{,}528$).

Language disorders in the family found their reflection in all three repetition tasks (KiSS: $Z = -2.69$, $p = .007$, $\hat{p} = .41$, $n = 1{,}535$; SSV: $Z = -2.85$, $p = .004$, $\hat{p} = .41$, $n = 1{,}519$; QUNW: $Z = -2.14$, $p = .032$, $\hat{p} = .43$, $n = 1{,}537$). Children from such families scored significantly lower in nonword repetition tasks than children from unaffected families. The same is valid for children with cases of a language (L1) delay among relatives (KiSS: $Z = -5.28$, $p < .001$, $\hat{p} = .34$, $n = 1{,}514$; SSV: $Z = -4.53$, $p < .001$, $\hat{p} = .36$, $n = 1{,}498$; QUNW: $Z = -3.45$, $p = .001$, $\hat{p} = .40$, $n = 1{,}516$). Children whose relatives underwent speech-language therapies scored lower in two PSTM tasks than children from unaffected families, with one result being marginally significant (KiSS: $Z = -2.13$, $p = .033$, $\hat{p} = .45$, $n = 1{,}530$; SSV: $Z = -1.80$, $p = .073$, $\hat{p} = .46$, $n = 1{,}514$; QUNW: $Z = -2.53$, $p = .011$, $\hat{p} = .44$, $n = 1{,}532$). Thus, for three variables, German-based nonwords demonstrated more prominent effect sizes than QUNW, for other three variables less prominent effect sizes, with differences that can be described as marginal. For two variables, the results were not statistically significant in all three PSTM tasks ($ps > .05$): children's regular medicine intake and frequent otitis media.

4. DISCUSSION

Due to the well-known association between weak performance in PSTM tasks and language impairments, items on the repetition of nonwords are included in many language tests. In this study, widely used German-based nonwords were compared with QUNW suggested by Chiat (2015). It was hypothesized that the use of QUNW could reduce differences between monolingual and bilingual children in PSTM task results while identifying children with language impairments as reliably as German-based nonwords do. Both expectations were confirmed, although results differed to a certain degree depending on the chosen gold standard. In contrast to German-based nonwords, QUNW demonstrated comparable levels of difficulty for monolinguals and bilinguals. Also, total scores of QUNW differentiated between (a) children with and without language impairments and (b) those undergoing and not undergoing speech-language therapies as well as total scores of German-based items did. Thus, for the first time, the use of QUNW proposed by Chiat (2015) was shown to be able to detect language impairments (including those with and without comorbidities) both in monolingual and bilingual children. Also, for the first time, the results of Boerma and Blom (2017) were replicated for German and with a broader definition of language impairments. Furthermore, QUNW were shown to be more culture-fair than German-based ones. Thus, in the case of German-based nonwords, almost equal proportions of monolingual and bilingual children with language impairments did not result in a comparable performance in repetition tasks. On the contrary, repetition scores of QUNW did not contradict language test results on children's language impairments and did not differ much in the subgroups of monolingual and bilingual children.

Both in published results of school enrollment examinations and in some previous studies, German monolinguals were shown to outperform bilinguals in most German-based nonword repetition tasks (e.g., in the school enrollment test S-ENS; cf. Heudorf, 2017). However, some exceptions do exist. For instance, no statistically significant differences between monolinguals' and bilinguals' performance were found in Italian-based nonwords (Vender et al., 2019). Also, no such differences were found in the Mottier test (Mottier, 1951; cf. Wild & Fleck, 2013). In the "Heidelberg Auditive Screening for the School Enrollment Examination" (HASE; Brunner & Schöler, 2002), bilinguals sometimes even outperformed monolingual German children (cf. Landratsamt Göppingen, 2019). Such contradictions in the PSTM test results might be attributed to the linguistic load of PSTM tasks. Indeed, Vender et al. (2019) explain comparable results of monolinguals and bilinguals in Italian-based nonwords by the simple phonotactic structure of this language. Whereas S-ENS uses German-specific consonant clusters in nonwords, the Mottier test avoids them completely and HASE almost completely. The expectation that language skills interfere with the interpretation of PSTM tasks was also confirmed by the results of the study presented here. In the German-based nonwords, monolinguals outperformed bilinguals in total scores of correct answers. In QUNW, on the contrary, bilinguals slightly outperformed monolinguals. Although in QUNW the difference was not far from being only marginally significant, it might deliver further evidence for the hypothesis on the bilingual executive function advantage (Blom et al., 2014; Sörman et al., 2017).

Bilingual children's poor performance in the repetition of German-based nonwords found its explanation in their comparatively limited German language skills. Just like in most other studies on four-year-old German children, including those on KiSS (Tomasik et al., 2020) and SSV (Grimm, 2017), bilinguals in the study presented here yielded significantly weaker German language test results than monolinguals. They were significantly more often classified as needing additional educational assistance in acquiring German (ED). However, in QUNW, monolinguals lost their advantage. Their advanced command of German phonotactics, morphology, and vocabulary could not contribute to their nonword repetition performance because QUNW did not contain any German-specific stress patterns, phonemes or affixes.

Language impairments do not depend on children's ethnic or monolingual/bilingual background (Neumann et al., 2009). Therefore, no statistically significant differences in the proportions of monolingual and bilingual children with language impairments were found. Because nonword repetition tasks were designed to detect language impairments, equal proportions of children with such impairments among monolinguals and bilinguals were supposed to result in equal or comparable results of repetition tasks. The fact that bilinguals were outperformed by monolinguals in German-based nonwords underlines the necessity of the development of more culture-fair PSTM tasks, as has been demanded for decades in the literature on language and working memory assessment (e.g., Grimm, 2016). On the contrary, in QUNW, no discrepancy between proportions of children with language impairments and performance in repetition tasks was found.

The difference between German-based nonwords and QUNW became even more prominent in the classification trees. Poor performance in German-based nonwords was associated, first and foremost, with weak German language skills. On the contrary, poor performance in QUNW was associated with language impairments. German-based items were also linked to children's bilingual background, that is, again, to children who were often classified as needing educational assistance in acquiring German (in KiSS), whereas in the tree branches with QUNW educational needs did not appear at all. Thus, QUNW came closer to the original designation of repetition tasks, namely to the detection of children with language impairments. Although a weak performance in QUNW was not associated with educational needs as clearly as performance in German-based items was (see also the results of Mann-Whitney *U* tests), this cannot be considered their weakness because educational needs in KiSS are to be detected by means of other subtests.

Because KiSS is worse in detecting children with language impairments than children with educational needs (Tomasik et al., 2020), two gold standards for language impairments were tried out. Some of the language impairments not detected in KiSS were probably mentioned in the questionnaires for kindergarten teachers. Although no sensitivity and specificity values for the respective questionnaire items are available, they are known to be closely associated with KiSS results (Zaretsky et al., 2014). According to the first gold standard—dichotomized KiSS outcome—the difference between children with and without language impairments (MED and NMED) was more prominent, in terms of effect sizes, in the results of German-based nonwords than in the results of QUNW in three out of four comparisons. According to the second gold standard—questionnaire items on language impairments and participation in speech-language therapies—the difference between these two kinds of nonwords was almost non-existent. This discrepancy can be explained by the facts that (a) KiSS leaves about one third of children with language impairments undetected, (b) questionnaire items reflect all diagnosed children's language impairments but probably overlook some undiagnosed ones. Thus, both gold standards miss some of the children with language impairments, which might have resulted in discrepant findings.

A family history of language disorders or dyslexia is a well-known risk factor for children's language impairments (Bishop et al., 2017). In accordance with previous studies (Kalnak et al., 2014; Schulze et al., 2018), various language impairments, such as children's language delay, dyslexia among relatives, familial language disorders, and familial language delay, were significantly associated with poor PSTM performance. Again, no considerable differences were found between German-based nonwords and QUNW.

Among possible limitations of the study, it should be emphasized that statistically significant differences between subgroups of children (e.g., MED vs. NMED) in the repetition of nonwords cannot be interpreted in terms of the quality criteria of nonwords for the identification of children with language-related medical issues. In other words, significant differences between total scores of correct answers of children with and without language-related impairments do not presuppose that the nonwords constitute a reliable tool for the diagnostics of such impairments. All

attempts to "translate" such differences into acceptable sensitivity and specificity values for the identification of language-related impairments failed (e.g., Zaretsky & Hey, 2022), which shows that nonwords cannot be used as an independent predictor or diagnostic tool and should be combined with other diagnostic methods. Other possible limitations of the study are the recruitment bias (only children whose parents signed an informed consent participated in the study) and weaknesses of the gold standards for language-related impairments (limited sensitivity of the MED/NMED classification, subjectivity of the questionnaire items).

In recent years, nonwords have been shown to be a promising marker for language impairments not only for monolingual, but also for bilingual children (Bonifacci et al., 2020; Guasti et al., 2021; Guiberson & Rodríguez, 2020). In the present study, further evidence was found that the use of nonword repetition tasks, especially those based on QUNW, can differentiate between both monolingual and bilingual children with educational needs and language impairments (in terms of total scores of correct answers). However, questions remain. A more reliable gold standard than those in the present study—ideally, diagnosed language impairments—should be used for the assessment of these criteria. Taking into account that bilinguals with minimal L2 skills cannot be tested with widely used language tests such as KiSS and SSV, QUNW might also appear promising in regard to their predictive validity for the performance of these bilinguals in the school enrollment examination or written language acquisition in primary school. It is to be expected that, analogous to children with developmental language disorder (de Bree et al., 2010), children with a limited performance in such repetition tasks at the age of four would lag behind in L2 acquisition at the age of six or later due to their diagnosed or not yet diagnosed language impairments.

To conclude, the use of QUNW proposed by Chiat (2015) can improve PSTM task performance of bilinguals in comparison with monolinguals. German-based and quasi-universal PSTM items yielded comparable results in the prediction of questionnaire items on language impairments as well as on the participation in speech-language therapies. QUNW showed weaker results than German-based items in the detection of children with educational needs in acquiring German. However, such items were developed not for the assessment of German language skills but, rather, for the direct assessment of PSTM irrespective of the L1 and L2 skills. In contrast to language-specific nonwords, QUNW can be used to assess PSTM even in newcomers from other countries with minimal or no L2 command because such nonwords are hardly associated with language-specific input. Integration of nonword repetition items such as those proposed by Chiat (2015) into language tests and school enrollment examinations might improve test quality criteria, especially in regard to the detection of bilingual children with language impairments.

REFERENCES

Adlof, S. M., & Patten, H. (2017). Nonword repetition and vocabulary knowledge as predictors of children's phonological and semantic word learning. *Journal of Speech, Language, and Hearing Research*, *60*(3), 682–693. https://doi.org/10.1044/2016_JSLHR-L-15-0441

Archibald, L. M. D., & Joanisse, M. F. (2009). 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*, *52*(4), 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099)

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: Evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, *51*(6), 715–731. https://doi.org/10.1111/1460-6984.12242

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PlosONE*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 consortium (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Blom, E., Kuentay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

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

Bonifacci, P., Atti, E., Casamenti, M., Piani, B., Porrelli, R., & Mari, R. (2020). Which measures better discriminate language minority bilingual children with and without Developmental Language Disorder? A study testing a combined protocol of first and second language assessment. *Journal of Speech, Language, and Hearing Research*, *63*(6), 1898–1915. https://doi.org/10.1044/2020_JSLHR-19-00100

Brunner, M., & Schöler, H. (2002). *HASE — Heidelberger Auditives Screening in der Einschulungsuntersuchung* [HASE — Heidelberg Auditive Screening for the School Enrollment Examination]. WESTRA.

Bühl, A. (2018). *SPSS: Einführung in die moderne Datenanalyse ab SPSS 25* [SPSS: Introduction to the modern data analysis for SPSS 25 or later]. Pearson Studium.

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*(3), 460–470. https://doi.org/10.1097/MAO.0b013e3182868340

Chiat, S. (2015). Non-word repetition. In S. Armon-Lotem, J. de Long & N. Meir (Eds.), *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment* (pp. 125–150). Multilingual Matters.

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*(1), 36–44. https://doi.org/10.1002/dys.395

Döpfner, M., Dietmair, I., Mersmann, H., Simon, K., & Trost-Brinkhues, G. (2005). *S-ENS. Screening des Entwicklungsstandes bei Einschulungsuntersuchungen* [S-ENS. Developmental Screening for the School Enrollment Examinations]. Hogrefe.

Duncan, T. S., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2

vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language, and Hearing Research*, *59*(1), 39–48. https://doi.org/10.1044/2015_jslhr-l-14-0020

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83–94. https://doi.org/10.3758/BF03210559

Gresch, L. D., Marchman, V. A., Loi, E. C., Fernald, A., & Feldman, H. M. (2018). Nonword repetition and language outcomes in young children born preterm. *Journal of Speech, Language, and Hearing Research*, *61*(5), 1203–1215. https://doi.org/10.1044/2018_JSLHR-L-17-0217

Grimm, A. (2016). Quatschwörter nachsprechen — gleiche Anforderungen für alle Kinder? [Repetition of nonwords — equal requirements for all children?]. *Diskurs Kindheits- und Jugendforschung*, *11*(1), 113–117. https://doi.org/10.3224/diskurs.v11i1.22252

Grimm, H. (2001). *SETK 3–5. Sprachentwicklungstest für drei- bis fünfjährige Kinder* [SETK 3–5. Language Development Test for Three- to Five-Year-Old Children]. Hogrefe.

Grimm, H. (2003). *SSV — Sprachscreening für das Vorschulalter* [SSV — Language Screening for the Preschool Age]. Hogrefe.

Grimm, H. (2017). *SSV. Sprachscreening für das Vorschulalter. Kurzform des SETK 3–5* [SSV. Language Screening for the Preschool Age. Short version of SETK 3–5]. (2nd ed.). Hogrefe.

Grissom, R. J., & Kim, J. J. (2012). *Effect Sizes for Research: Univariate and Multivariate Applications.* (2nd ed.). Routledge/Taylor & Francis Group.

Guasti, M. T., White, M. J., Bianco, G., Arosio, F., Camilleri, B., & Hasson, N. (2021). Two clinical markers for DLD in monolingual Italian speakers: What can they tell us about second language learners with DLD? *Clinical Linguistics & Phonetics*, *35*(9), 829–846. https://doi.org/10.1080/02699206.2020.1830303

Guiberson, M. M., & Rodríguez, B. L. (2020). Working memory and linguistic performance of dual language learners with and without developmental language disorders. *American Journal of Speech-Language Pathology*, *29*(3), 1301–1306. https://doi.org/10.1044/2019_AJSLP-19-00109

Hamman, C., & Ibrahim, L. A. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, *2*, 16. https://doi.org/10.3389/fcomm.2017.00016

Heudorf, U. (2017). *Kindergesundheit in Frankfurt am Main. Daten des Gesundheitsamtes 2002–2014. Ergänzungsbericht 2015–2016* [Children's health in Frankfurt/Main. Data of the public health department 2002–2014. Additional report on years 2015–2016]. Magistrate, Public Health Department of Frankfurt/Main, Germany.

Holler-Zittlau, I., Euler, H. A., & Neumann, K. (2011). Kindersprachscreening (KiSS) — das hessische Verfahren zur Sprachstandserfassung [Kindersprachscreening (KiSS) — the Hessian language test]. *Sprachheilarbeit*, *5*(6), 263–268.

Jones, G. (2016). The influence of children's exposure to language from two to six years: The case of nonword repetition. *Cognition*, *153*, 79–88. https://doi.org/10.1016/j.cognition.2016.04.017

Kalnak, N., Peyrard-Janvid, M., Forssberg, H., & Sahlén, B. (2014). Nonword repetition — A clinical marker for specific language impairment in Swedish associated with parents' language-related problems. *PLoS One*, *9*(2), e89544. https://doi.org/10.1371/journal.pone.0089544

Kiese-Himmel, C., & Nickisch, A. (2016). Korrelieren phonologische KZG-Leistungen mit sprachbasierten Dimensionen bei AVWS-Kindern? [Does phonological short-term memory performance correlate with language-based dimensions in children with APD?]. *Laryngo-Rhino-Otologie*, *95*(1), 24–28. https://doi.org/10.1055/s-0035-1555916

Kuppen, S. E. A., & Goswami, U. (2016). Developmental trajectories for children with dyslexia and low IQ poor readers.

Developmental Psychology, *52*(5), 717–734. https://doi.org/10.1037/a0040207

Landratsamt Göppingen. (2019). *Einschulungsuntersuchung 2017/2018 im Landkreis Göppingen* [School enrollment examinations 2017–2018 in the administrative district Goeppingen]. District administration office Goeppingen (Germany), public health department. Available at https://www.gesundheitsamt-bw.de/fileadmin/LGA/_DocumentLibraries/SiteCollectionDocuments/01_Themen/GesundheitsdatenBerichte/Kommunale_GBE/GBE_ESU_LKGoeppingen_2018.pdf

Li'el, N., Williams, C., & Kane, R. (2019). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. *International Journal of Speech-Language Pathology*, *21*(6). https://doi.org/10.1080/17549507.2018.1513073

Meir, N. (2017). Effects of Specific Language Impairment (SLI) and bilingualism on verbal short-term memory. *Linguistic Approaches to Bilingualism*, *7*(3–4), 301–330. https://doi.org/10.1075/lab.15033.mei

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology*, *8*, 1442. https://doi.org/10.3389/fpsyg.2017.01442

Mottier, G. (1951). Über Untersuchungen der Sprache lesegestörter Kinder [On studies regarding the language of children with reading disorders]. *Folia Phoniatrica*, *3*, 170–177. https://doi.org/10.1159/000262507

Munson, B., Kurtz, B. A., & Windsor, J. (2005). 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*, *48*(5), 1033–1047. https://doi.org/10.1044/1092-4388(2005/072)

Neumann, K., & Euler, H. A. (2009). *Einführung einer flächendeckenden Sprachstandserfassung in Hessen. Forschungsbericht* [Introduction of a state-wide language screening in Hesse. Research report]. Goethe University of Frankfurt/Main, Germany.

Neumann, K., & Euler, H. A. (2010). *Einführung einer flächendeckenden Sprachstandserfassung in Hessen. Forschungsbericht* [Introduction of a state-wide language screening in Hesse. Research report]. Goethe University of Frankfurt/Main, Germany.

Neumann, K., Keilmann, A., Rosenfeld, J., Schönweiler, R., Zaretsky, Y., & Kiese-Himmel, C. (2009). Sprachentwicklungsstörungen bei Kindern. Leitlinien der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie (gekürzte Fassung) [Guidelines of the German Society of Phoniatrics and Pediatric Audiology on developmental speech and language disorders of children (short version)]. *Kindheit und Entwicklung*, *18*(4), 222–231. https://doi.org/10.1026/0942-5403.18.4.222

Neumann, K., Zaretsky, Y., & Euler, H. A. (2011). *Einführung einer flächendeckenden Sprachstandserfassung in Hessen. Forschungsbericht* [Introduction of a state-wide language screening program in Hesse. Research report]. Goethe University of Frankfurt/Main, Germany.

Orban, S. A., Rapport, M. D., Friedman, L. M., Eckrich, S. J., & Kofler, M. J. (2018). Inattentive behavior in boys with ADHD during classroom instruction: The mediating role of working memory processes. *Journal of Abnormal Child Psychology*, *46*(4), 713–727. https://doi.org/10.1007/s10802-017-0338-x

Paradis, J., Schneider, P., & Sorenson Duncan, T. (2013). Discriminating children with language impairment among English language learners from diverse first language backgrounds. *Journal of Speech, Language and Hearing Research*, *56*(3), 971–981. https://doi.org/10.1044/1092-4388(2012/12-0050)

Petermann, F., Daseking, M., Oldenhage, M., & Simon, K. (2009). *Sozialpädiatrisches Screening für Schuleingangsuntersuchungen (SOPESS)* [Socio-Pediatric Screening for the School Enrollment Examination (SOPESS)]. LIGA. NRW.

Reichenbach, K., Bastian, L., Rohrbach, S., Gross, M., & Sarrar, L. (2016). Cognitive functions in preschool children with specific language impairment. *International Journal of Pediatric Otorhinolaryngology*, *86*, 22–26. https://doi.org/10.1016/j.ijporl.2016.04.011

Ruhland, I., Fuhr-Becker, G., & Avdic, A. (2022). *Hessischer Integrationsmonitor 2022* [Hessian integration monitor 2022]. Hessian Ministry for Social Affairs and Integration.

Rujas, I., Mariscal, S., Casla, M., Lázaro, M., & Murillo, E. (2017). Word and nonword repetition abilities in Spanish language: Longitudinal evidence from typically developing and late talking children. *The Spanish Journal of Psychology*, *20*, E72. https://doi.org/10.1017/sjp.2017.69

Schulze, K., Vargha-Khadem, F., & Mishkin, M. (2018). Phonological working memory and FOXP2. *Neuropsychologia*, *108*, 147–152. https://doi.org/10.1016/j.neuropsychologia.2017.11.027

Sörman, D. E., Josefsson, M., Marsh, J. E., Hansson, P., & Ljungberg, J. K. (2017). Longitudinal effects of bilingualism on dual-tasking. *PLoS One*, *12*(12), e0189299. https://doi.org/10.1371/journal.pone.0189299

Sundström, S., Löfkvist, U., Lyxell, B., & Samuelsson, C. (2018). Prosodic and segmental aspects of nonword repetition in four- to six-year-old children who are deaf and hard of hearing compared to controls with normal hearing. *Clinical Linguistics & Phonetics*, *32*(10), 950–971. https://doi.org/10.1080/02699206.2018.1469671

Tomasik, M. J., Lindner, N., Weiland, M., Schmidt, M., Moser, U., & Kieslich, M. (2020). Landesweite Normierung des Hessischen Kindersprachscreenings [State-wide norming of the Hessian language screening for children]. *Diagnostica*, *66*(3), 178–189. https://doi.org/10.1026/0012-1924/a000246

Truong, D. T., Shriberg, L. D., Smith, S. D., Chapman, K. L., Scheer-Cohen, A. R., DeMille, M. M. C., Adams, A. K., Nato, A. Q., Wijsman, E. M., Eicher, J. D., & Gruen, J. R. (2016). Multipoint genome-wide linkage scan for nonword repetition in a multigenerational family further supports chromosome 13q as a locus for verbal trait disorders. *Human Genetics*, *135*(12), 1329–1341. https://doi.org/10.1007/s00439-016-1717-z

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., dos Santos, C., Ibrahim, L. A., & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders*, *53*(4), 888–904. https://doi.org/10.1111/1460-6984.12397

van Wingerden, E., Segers, E., van Balkom, H., & Verhoeven, L. (2017). Foundations of reading comprehension in children with intellectual disabilities. *Research in Developmental Disabilities*, *60*, 211–222. https://doi.org/10.1016/j.ridd.2016.10.015

Vender, M., Delfitto, D., & Melloni, C. (2019). How do bilingual dyslexic and typically developing children perform in nonword repetition? Evidence from a study on Italian L2 children. *Bilingualism: Language and Cognition*, *23*(4), 884–896. https://doi.org/10.1017/S1366728919000828

Wild, N., & Fleck, C. (2013). Neunormierung des Mottier-Tests für 5- bis 17-jährige Kinder mit Deutsch als Erst- oder Zweitsprache [New norms for the Mottier test for children aged 5–17 years with German as the first or second language]. *Praxis Sprache*, *3*, 152–157.

Yoo, J., & Kaushanskaya, M. (2012). Phonological memory in bilinguals and monolinguals. *Memory & Cognition*, *40*(8), 1314–1330. https://doi.org/10.3758/s13421-012-0237-x

Zaretsky, E., Euler, H. A., Neumann, K., & Lange, B. P. (2014). Sociolinguistic predictors of language deficits in pre-school children with and without immigrant background. In *ACLL2014 – The Asian Conference on Language Learning* (pp. 42–53). The International Academic Forum.

Zaretsky, E., & Hey, C. (2022). Deutsch-basierte und (quasi-)universelle Kunstwörter als Prädiktoren für sprachbezogene medizinische Störungsbilder bei Vorschulkindern [German-based and (quasi-universal) nonwords as predictors for language-related medical issues

of preschoolers]. In M. Spreer, M. Wahl & H. Beek (Eds.), *Sprachentwicklung im Dialog. Digitalität — Kommunikation — Partizipation* (pp. 198–204). Schulz-Kirchner Verlag.

Zaretsky, E., Neumann, K., Euler, H. A., & Lange, B. P. (2013). Pluralerwerb im Deutschen bei russisch- und türkischsprachigen Kindern im Vergleich mit anderen Migranten und monolingualen Muttersprachlern [Acquisition of the German plural by Turkish and Russian children in comparison with other immigrants and monolingual native speakers]. *Zeitschrift für Slawistik*, *58*(1), 43–71. https://doi.org/10.1524/slaw.2013.0003

Eugen Zaretsky | Department of Phoniatrics and Paediatric Audiology, Marburg University Hospital
<eugen.zaretsky@staff.uni-marburg.de>

Benjamin P. Lange | Department of Social Sciences, IU International University of Applied Sciences
<b.lange@iubh-dualesstudium.de>

Christiane Hey | Department of Phoniatrics and Paediatric Audiology, Marburg University Hospital
<christiane.hey@med.uni-marburg.de>