Language-Specific Non-Words for the Assessment of Working Memory: Dealing With Bilingual Children

Eugen Zaretsky, DPhil,[1] and Benjamin P. Lange, DPhil[2]

Abstract

Bilingual children with a limited command of the second language (L2) often yield unsatisfactory results in L2-based non-word repetition tasks (NWRT) for the assessment of working memory. In this study, monolinguals (MO) and bilinguals (BI) of preschool age acquiring German were compared in regard to their performance on German-based NWRT to choose test items that do not put BI at a disadvantage. Four-year-old children ($N = 876$) were tested with the German language screenings *Sprachscreening für das Vorschulalter* (SSV) and *Kindersprachscreening* (KiSS), including 38 non-words. BI scored significantly lower in NWRT than MO due to a limited command of German as well as a limited language input (e.g., length of kindergarten attendance). After the statistical deletion of children who did not speak German age-appropriately, BI outperformed MO on the SSV, without significant differences on the KiSS. Performance on NWRT depended on the children's command of the German vocabulary and phonotactics. Several relatively "culture-fair" NWRT items were identified.

Working memory is involved in the acquisition of linguistic regularities (Bosma et al., 2017) and has been shown to be a predictor for both L1 (Ogino et al., 2017) and second language (L2) acquisition (Nowbakht, 2019). Deficits in working memory can signal language impairment (Montgomery et al., 2010), as indicated by children with such impairments scoring lower than their typically developing peers on working memory tasks: non-word repetition, digit span, and sentence repetition (Ferré et al., 2015). Over the past 30 years, such tasks have been validated as part of evaluation for dyslexia (Germano et al., 2017), specific language impairment (SLI, Willinger et al., 2017), and, more generally, language impairment (Pham et al., 2018).

However, bilingual children (BI) tend to score significantly lower than their monolingual peers (MO) in the repetition tasks (Asbrock, 2009), which is evident, for instance, from numerous reports on the German school enrolment examination (City Frankfurt/Main, 2017; District administration Goeppingen, 2019). This makes the differentiation between children with limited L2 skills and children with language impairments difficult. If cut-off values for MO are applied to BI, the latter group can be falsely classified as language impaired (Paradis et al., 2013). In this case, children acquiring more than one language can be prescribed unnecessary medical examinations (e.g., examinations for the diagnosis of hearing or auditory processing disorders).

Insufficient performance of typically developing BI on non-word repetition tasks (NWRT) can be traced back to a limited command of L2 and an influence of the first language (L1). For instance, the lack of word-final consonant clusters in Turkish and in any word position in Japanese or the lack of the phoneme /g/ in Arabic make it difficult for immigrants from other, especially non-European, countries, to perform on the same level in German-based non-words as German MO do. At the kindergarten age (3–6 years), some deficits in the acquisition of L2 phonology, such as insufficient vowel discrimination (Darcy & Krueger, 2012), still remain even years after L2 acquisition onset. Therefore, due to the link between phonotactic knowledge and performance on language-specific working memory tasks (Rodrigues & Befi-Lopes, 2013), children with a limited command of the phonotactics of the respective language can be put at a disadvantage by non-words based on their L2 (Parra et al., 2011).

[1]University Hospital of Marburg, Philipps University of Marburg Germany
[2]IU International University of Applied Sciences, Bad Honnef, Germany

Corresponding Author:
Dr. Eugen Zaretsky, Department of Phoniatrics and Pediatric Audiology, University Hospital of Marburg, Philipps University of Marburg Baldingerstr. 1, Marburg 35043, Germany.
Email: zaretsky@staff.uni-marburg.de

Performance on NWRT is predictive, among other things, of children's vocabulary acquisition (Marini et al., 2017), but in its turn depends on the vocabulary size because the long-term memory contributes to holding memory trace, a theoretical means by which memories are physically stored in the brain (Gathercole, 2006). Phonology knowledge develops with the number of acquired words, making phonotactic knowledge dependent upon the child's vocabulary size (Beckman et al., 2007). For instance, children with word learning difficulties such as those with SLI were shown to have a deficit in phonological representations and to score lower in NWRT than their typically developing peers (Befi-Lopes et al., 2010). In children with weak phonological processing, lexical-semantic knowledge supporting pattern completion within the phonological system is often activated to improve the performance on short-term memory tasks, including NWRT (Savill et al., 2019). Thus, children's acquired vocabulary forms and predetermines phonological representations and compensates for the weak phonological processing, when necessary.

Due to this link between working memory and vocabulary skills, children repeat non-words with a high lexical load more easily than items with a low load (Gathercole, 1995). For the same reason, BI score lower on items based on their L2 than MO on items based on L1, especially on tasks with a high semantic load (Yoo & Kaushanskaya, 2012). Not only do many non-words follow the typical syllable structure and accent patterns of the target language; some of them are also deliberately designed to contain language-specific morphemes or to rhyme with existing words. Such items therefore presuppose linguistic knowledge that many BI have not yet acquired, thereby putting them at risk for a low NWRT score and contributing to misidentification of language disorders.

Hence, in case of BI, crosslinguistic phonological and lexical transfers from their L1 are to be expected in L2-based NWRT (de Abreu et al., 2013). Therefore, the assessment of working memory by language-specific NWRT in BI is often criticized (e.g., Grimm, 2016). However, this critique did not change the daily clinical practice due to the lack of validated NWRT that would account for the phonotactic regularities across languages. As a possible alternative to the extremely challenging development and validation of such "universal" tests, relatively "culture-fair" non-words from existing language-specific tests can be selected, that is, non-words that do not yield significantly different results for MO and BI.

A large number of variables related to the language input, including attendance of kindergartens, associations, and study groups, age of the onset of language acquisition or language use at home, has been shown to be associated with children's language competence (Zaretsky et al., 2020). Therefore, a link between these variables and children's performance on language-specific NWRT can be expected in terms of a higher quality and quantity of the language input associated with higher scores of correct results in NWRT.

This study aimed (a) to examine whether BI of preschool age acquiring German are put at a disadvantage by commonly used German-based non-words, (b) to correlate children's performance on NWRT with the command of the German phonotactics and vocabulary, but also quality and quantity of the language input, and (c) to select items that are less influenced by German language skills and, thus, less biased. In a search for the most appropriate German-based non-words, BI and MO were compared in regard to their performance on two NWRT under the two following conditions: (a) all children irrespective of their German language skills and (b) only children with age-appropriate German language skills. For (a), it was hypothesized that some German-based non-words would put BI at a disadvantage due to their limited command of the German language. On the contrary, for (b), no significant differences between BI and MO were expected.

Method

Test Subjects and Test Battery

In this retrospective study, language test results of 876 children (479/54.7% male, 397/45.3% female; age 4 years 0 months to 4 years 11 months, age mean and median 51 months) from 154 German kindergartens were utilized. Originally, these data were collected for the purpose of validation of various language tests. More than half of the children were monolingual Germans ($n = 520/59.4\%$), all other acquired one more language ($n = 356/40.6\%$), most often Turkish ($n = 65/7.4\%$ of the sample), Arabic ($n = 39/4.5\%$), and Russian ($n = 36/4.1\%$). There was no statistically significant age difference between BI and MO according to a Mann–Whitney U test.

All children were tested with the language screening "Sprachscreening für das Vorschulalter" (SSV; Grimm, 2003). SSV includes the following 18 German-based non-words: _Billop_, _Kalifeng_, _Defsal_, _Ronterklabe_, _Toschlander_, _Entiergent_, _Gattwutz_, _Glösterkeit_, _Dilecktichkeit_, _Krapselistong_, _Nebatsubst_, _Seregropist_, _Skatagurp_, _Waltikosander_, _Pristobierichkeit_, _Kabusaniker_, _Ippazeumerink_, _Vominlapertust_.

Additionally, all children were tested with the second version of the language test "Kindersprachscreening" (KiSS; Holler-Zittlau et al., 2011), with subtests on speech comprehension, vocabulary, phonology, grammar, and working memory. KiSS categorizes children as (a) speaking German age-appropriately, (b) needing additional educational support (ED), and (c) needing additional medical support (MED) in acquiring German. In terms of the international CATALISE study, ED children are those who had

insufficient exposure to the language used by the school or community to be fully fluent in it and who do not have a language disorder (Bishop et al., 2017). MED children are those with language impairments (Bishop et al., 2016). Thus, ED children need more language input (e.g., a German language reinforcement course offered by many kindergartens), whereas MED children need some kind of medical assistance that would contribute to the development of their L1 or L2 competence, such as hearing aids in case of a hearing disorder. Deficits in working memory are characteristic of MED children (see Introduction), but not ED ones.

In KiSS, the same cut-off values are applied to the classification of BI and MO as ED. However, KiSS does differentiate between cut-off values for BI and MO in the MED classification, with lower values for BI. According to KiSS criteria, 240 children (27.4%) from the sample were classified as ED and 110 (12.6%) as MED. There was an overlap between these two groups because most of MED children were also classified as ED ($n = 83/9.5\%$). All other children ($n = 609/69.5\%$) spoke German age-appropriately.

Because KiSS.2 contains only four non-words (*Triser*, *Misküranok*, *Nabolira*, *Verklasenaft*), direct comparisons with SSV (18 items) were hardly possible. For such comparisons, a subsample of children who were tested with an extended set of the following 20 non-words was utilized (Neumann & Euler, 2009): four items mentioned above and *Gune*, *Lamposta*, *Kaulen*, *Meling*, *Flome*, *Siraf*, *Schabu*, *Kneefode*, *Feilister*, *Posalin*, *Verwumpeln*, *Gestrabelt*, *Schrimane*, *Kirumander*, *Wasubritzen*, *Berofterheit*. This set of items was a part of a previous KiSS version and later was reduced to four items. A total of 201 children were tested with this extended KiSS version (101/50.2% male; 100/49.8% female; 100/50.2% BI, 101/49.8% MO; age 4 years 0 months to 4 years 11 months; mean and median age 51 months), without significant age differences between BI and MO. These children did not constitute a separate sample but, due to identical KiSS subtests (except NWRT), were a part of the sample described above.

In addition to the language screening and NWRT, KiSS contains questionnaires for parents and kindergarten teachers, with items that can be categorized as follows: (a) demographic variables such as birth date, (b) quality and quantity of the language input in German and (an)other language(s) (e.g., age when the child began to acquire German), (c) children's and parents' language-related medical issues (e.g., language disorders in the family), and (d) participation in language courses and therapies.

Statistical Analysis

As it was hypothesized that weaker German language skills of BI would find their reflection in the weaker performance on both NWRT, first, German language skills of BI and MO were compared. The classification BI/MO was cross-tabulated with the KiSS classifications as ED and MED, including a chi-square calculation. Performance of BI and MO on KiSS subtests on vocabulary and phonology—two linguistic domains that were shown to be associated with the results of NWRT—were compared by Mann–Whitney U tests.

To analyze the association of the German language skills with the performance on NWRT, two classification trees were calculated (method "Exhaustive CHAID"), with total scores of correct answers in SSV and KiSS NWRT as dependent variables and (a) children's age in months, (b) classification BI/MO, (c) KiSS vocabulary score, and (d) KiSS phonology score as independent variables. For the same purpose, Spearman correlations between both NWRT and KiSS subtests on vocabulary and phonology were calculated.

NWRT from SSV and KiSS were analyzed in respect to possible disadvantages for BI compared with MO. Total scores of correct answers in both NWRT were compared in two Mann–Whitney U tests for BI and MO. Next, rates of correct and wrong answers in each non-word were compared for BI and MO in cross-tables with a chi-square calculation. Also, several selected consonant clusters in non-words (e.g., *Seregropist*) were analyzed by the same statistical method in regard to the omission of consonants in BI and MO subgroups. Such omissions were expected to occur more frequently in the BI subgroup due to a limited command of the German phonotactics and influence of other languages.

Associations between total scores of correct answers in the NWRT from SSV and children's sociodemographic characteristics were assessed by correlations. For these calculations, some KiSS questionnaire items on the quality and quantity of the L1 and L2 input were utilized (see Table 1). For ordinal variables (e.g., how often the child plays with German-speaking children), Spearman correlations were calculated. For dichotomous variables (e.g., whether the child attended a nursery school), point-biserial correlations were used. It was hypothesized that a link between performance on NWRT and sociodemographic variables related to the L1 and L2 input would be stronger in case of BI than in case of MO. The sample sizes for KiSS were too low, so that KiSS had to be excluded from these correlational analyses.

Next, performance of BI and MO on NWRT under the condition of comparable German language skills was assessed. All children who were classified as ED and MED were excluded. BI and MO who spoke German age-appropriately did not differ in respect to their age in months. Mann–Whitney U test was utilized to verify whether significant differences between BI ($n = 172$) and MO ($n = 417$) in the total scores of correct answers in the KiSS subtest on the vocabulary and phonology still existed after the exclusion of ED and MED children. Next, total scores of

Table 1. Spearman (ρ) and Point-Biserial (r_{pb}) Correlations: Associations Between the Total Score of Correct Answers in SSV Non-Word Repetition Task and Items on the Language Input From KiSS Questionnaires for Kindergarten Teachers (QK) and Parents (QP).

Questionnaire items	ρ or r_{pb}	n
QK: Length of kindergarten attendance: 0–14 vs. 15+ months	$r_{pb} = .17^{***}$	261
QK: Attendance of a nursery school in the first 2 years of life (yes/no)	$r_{pb} = -.18^{**}$	248
QK: Regular or irregular kindergarten attendance	$r_{pb} = -.16^{*}$	172
QK: Kindergarten attendance for half a day or full day	$r_{pb} = .14$	183
QK: The child speaks out when playing (never, seldom, sometimes, often, always)	$\rho = .23^{**}$	184
QK: There is at least one more child speaking the same non-German language in the kindergarten group (yes/no)	$r_{pb} = .00$	180
QK: If yes, how often the child plays with this child (never, …, always)	$\rho = .11$	151
QK: Age when the child began to acquire German (0–4 years)	$\rho = -.24^{**}$	150
QK: The child plays with German-speaking children in the kindergarten (never, …, always)	$\rho = .21^{**}$	204
QP: The child plays with German-speaking children after the kindergarten time (yes/no)	$r_{pb} = -.20^{**}$	265
QP: Attendance of associations and study groups (yes/no)	$r_{pb} = -.24^{**}$	170

Note. SSV = Sprachscreening für das Vorschulalter; KiSS = Kindersprachscreening; QK = Questionnaires for Kindergarten Teachers; QP = Questionnaires for Parents.
$^{*}p < .05.$ $^{**}p < .01.$ $^{***}p < .001.$

correct answers in both NWRT were compared by two Mann–Whitney U tests for BI and MO. BI were expected to catch up with MO. Possible disadvantages for BI in each of the SSV and KiSS non-words were assessed by cross-tables between rates of correct and wrong answers in NWRT and the BI/MO classification. Also, calculations of cross-tables between the BI/MO classification and variables on the omission of consonants in consonant clusters were repeated.

Because of a very heterogeneous constellation of the BI subgroup, with dozens of ethnic and linguistic backgrounds, a considerable stratification regarding performance on NWRT could be assumed. These differences were exemplified here by a comparison of BI speaking European ($n = 172$) and non-European ($n = 156$) languages. For 28 BI, no reliable information on their L1 was available. Therefore, they were not included in any of these subgroups. Both "Europeans" and "non-Europeans" were on average 51 months old, without significant difference in age. It was hypothesized that the "non-Europeans" would yield lower total scores of correct answers in NWRT due to a larger typological difference between German and their L1 than in case of the "Europeans." Again, after the comparison of the percentages of ED and MED children in these two subgroups in cross-tables, Mann–Whitney U tests were calculated to assess differences between "Europeans" and non-Europeans" in the total scores of correct answers in the SSV and KiSS NWRT. After the exclusion of all ED and MED children, Mann–Whitney U tests were repeated. Also, the omission of consonants in consonant clusters was compared for "Europeans" and "non-Europeans" in cross-tables before and after the exclusion of ED and MED children. Because both subgroups were BI, it was not expected that "non-Europeans" would outperform "Europeans" after the exclusion of children with limited German language skills.

For all mean values (M), standard deviations ($\pm$) were given. The effect size in all Mann–Whitney U tests was quantified with the probability of superiority index ($\hat{p}$; Grissom & Kim, 2012). Values close to .5 are considered as low, values close to .0 as high.

Results

Performance of All MO and BI on KiSS and NWRT

In KiSS, BI were classified significantly more often both as ED, $\chi^2(1) = 135.5$, $p < .001$; 173/48.6% vs. 67/12.9%, and MED, $\chi^2(1) = 11.5$, $p = .001$; 61/17.1% vs. 49/9.4%, than MO. Both in the KiSS subtest on vocabulary ($Z = -14.2$, $p < .001$, $\hat{p} = .20$; $M = 4.2 \pm 2.6$ vs. 7.0 ± 2.6) and phonology ($Z = -5.0$, $p < .001$, $\hat{p} = .40$; $M = 8.0 \pm 2.1$ vs. 8.7 ± 1.8), BI scored significantly lower than MO.

Next, the hierarchy of factors of influence for SSV and KiSS NWRT results was assessed by two classification trees. For the total score of correct answers in non-words from SSV, the children's phonology skills in German, quantified by KiSS, were found to be the most important factor, $F(3, 872) = 58.5$, $p < .001$. Children with total scores < 5 repeated, on average, 5.4 ± 3.4 non-words correctly, those with total scores 5-7 repeated 7.6 ± 3.6 non-words, those with 8–9 correct answers repeated 9.4 ± 3.5 non-words, and children with more than 9 correct answers repeated 11.0 ± 3.6 non-words. For the latter two subgroups, the next important factor was the KiSS subtest on vocabulary, $F(1, 416) = 12.5$, $p < .001$ and $F(1, 275) = 22.2$, $p < .001$, with higher vocabulary scores being associated with better NWRT results. Other factors did not appear in the classification tree.

For KiSS, only one factor was found to be associated with the total score of correctly repeated non-words, $F(1, 199) = 31.7$, $p < .001$: Children with a KiSS vocabulary score < 5 yielded, on average, 12.7 ± 4.2 correct answers in this NWRT, and children with a vocabulary score $5+$ yielded 15.7 ± 3.2 correct answers.

The same findings—a high association of KiSS phonology score with the performance on SSV NWRT and a high association of KiSS vocabulary score with the performance on KiSS NWRT—were confirmed by Spearman correlations (ps $< .001$). The total score of correctly repeated SSV non-words correlated with a KiSS vocabulary score lower than with the KiSS phonology score ($\rho = .31$ vs. $.42$), whereas a total score of correctly repeated KiSS non-words yielded a higher correlation coefficient for the vocabulary than for the phonology subtests ($\rho = .40$ vs. $.26$).

A link between the performance on NWRT and sociodemographic characteristics of the children was assessed separately for BI and MO. Several questionnaire items were correlated with the total score in SSV NWRT. In case of MO, only kindergarten attendance for half a day vs. full day yielded a significant result (a higher NWRT score in the latter case): $r_{pb} = .15$, $p = .026$, $n = 218$. Results of BI are given in Table 1. In all correlations, a larger quantity and/or quality of the German language input was associated with a higher NWRT score.

BI repeated significantly less non-words correctly than MO both in SSV ($Z = -3.2$, $p = .001$, $\hat{p} = .44$; $M = 8.8 \pm 4.0$ vs. 9.7 ± 3.7) and KiSS ($Z = -2.6$, $p = .009$, $\hat{p} = .39$; $M = 13.4 \pm 4.4$ vs. 15.1 ± 3.3). In SSV, BI yielded lower results for the items *Kalifeng*, $\chi^2(1) = 12.6$, $p < .001$, *Toschlander*, $\chi^2(1) = 4.9$, $p = .028$, *Gattwutz*, $\chi^2(1) = 8.9$, $p = .003$, *Glösterkeit*, $\chi^2(1) = 17.4$, $p < .001$, *Dilecktichkeit*, $\chi^2(1) = 9.1$, $p = .003$, and *Waltikosander*, $\chi^2(1) = 13.5$, $p < .001$. In KiSS, the same was valid for the items *Meling*, $\chi^2(1) = 8.2$, $p = .004$, *Flome*, $\chi^2(1) = 10.9$, $p = .001$, *Kneefode*, $\chi^2(1) = 9.4$, $p = .002$, *Berofterheit*, $\chi^2(1) = 6.9$, $p = .008$, and *Verklasenaft*, $\chi^2(1) = 4.9$, $p = .028$. No examples of better results of BI compared with MO were found.

Omissions of consonants in consonant clusters (marked bold below) occurred more often in the answers of BI, compared with MO, in the SSV items *Entierge**nt***, $\chi^2(1) = 4.7$, $p = .030$, *Gattwu**tz***, $\chi^2(1) = 19.0$, $p < .001$, *Seregropi**st***, $\chi^2(1) = 7.8$, $p = .005$, ***Gl**österkeit*, $\chi^2(1) = 5.1$, $p = .024$, and in the KiSS item *Verklasena**ft***, $\chi^2(1) = 6.1$, $p = .013$. No counter-examples were identified, that is, no items with a higher percentage of omissions in the MO subgroup compared with BI. Differences between BI and MO were not statistically significant for the SSV items *De**f**sal*, ***Kr**apselistong*, *Nebatsu**bst***, ***Sk**atagurp*, *Ippazeumeri**nk***, and *Vominlapertu**st***.

Performance of MO and BI With Age-Appropriate German Language Competence on KiSS and NWRT

Next, all children classified as ED and MED in KiSS were excluded. The following calculations comprised only children whose German language skills were age-appropriate. Under this condition, BI scored lower than MO in the KiSS subtest on vocabulary ($Z = -7.0$, $p < .001$, $\hat{p} = .30$; $M = 5.9 \pm 2.0$ vs. 7.2 ± 1.7). In phonology, no significant difference was found ($M = 8.8 \pm 1.4$ vs. 9.0 ± 1.4). In SSV NWRT, BI scored higher than MO ($Z = -2.1$, $p = .035$, $\hat{p} = .44$; $M = 10.9 \pm 3.4$ vs. 10.4 ± 3.4), whereas no significant difference in KiSS NWRT was identified ($M = 15.2 \pm 3.2$ vs. 15.5 ± 3.1).

After the exclusion of ED and MED children, MO scored higher than BI in the SSV item *Kalifeng*, $\chi^2(1) = 4.1$, $p = .042$, and in the KiSS item *Flome*, $\chi^2(1) = 4.3$, $p = .039$. BI scored higher than MO in the SSV items *Skatagurp*, $\chi^2(1) = 4.9$, $p = .027$, *Kabusaniker*, $\chi^2(1) = 10.7$, $p = .001$, and *Vominlapertust*, $\chi^2(1) = 4.7$, $p = .031$. In the cross-tables on the omission of consonants from consonant clusters in BI and MO subgroups, no significant differences were found.

Performance of Children Speaking European and Non-European Languages on KiSS and NWRT

In a similar way, possible differences within the BI subgroup were analyzed for children speaking European and non-European languages. In KiSS, the former were classified less often as ED than the latter, $\chi^2(1) = 26.8$, $p < .001$; $n = 60/34.0\%$ vs. $99/63.5\%$, without significant difference in the MED classification. "Europeans" scored higher than "non-Europeans" in both NWRT: SSV ($Z = -2.8$, $p = .006$, $\hat{p} = .41$; $M = 9.4 \pm 4.0$ vs. 8.2 ± 4.0) and KiSS ($Z = -2.8$, $p = .004$, $\hat{p} = .33$; $M = 15.1 \pm 3.6$ vs. 12.4 ± 4.6). After the exclusion of ED and MED children, no significant differences between "Europeans" and "non-Europeans" were found in both NWRT: SSV ($M = 10.8 \pm 3.5$ vs. 10.9 ± 3.3) and KiSS ($M = 15.7 \pm 3.5$ vs. 14.9 ± 3.2). Whereas before the exclusion of ED and MED children "non-Europeans" omitted consonants more frequently than "Europeans" in three SSV items, *Nebatsu**bst***, $\chi^2(1) = 5.0$, $p = .025$, ***Sk**atagurp*, $\chi^2(1) = 15.9$, $p < .001$, and *Ippazeumeri**nk***, $\chi^2(1) = 5.7$, $p = .017$, after the exclusion there was only one left, *Nebatsu**bst***, $\chi^2(1) = 7.4$, $p = .006$, without counter-examples.

Discussion

Although NWRT in numerous language tests, including SSV, were designed to identify children with language-related medical issues, the respective cut-off values often implicitly

presuppose monolingual background. Performance on working memory tasks depends on the command of and input in the respective language (Szewczyk et al., 2018). In case of BI, limited German language skills and influence of the L1 phonotactics can result in low total scores of correctly repeated non-words even in children without any language-related impairments. As a consequence, BI can be misidentified as having a language impairment and prescribed unnecessary medical examinations and therapies.

This study aimed to analyze differences in the repetition of language-specific (here, German-based) non-words by MO and BI living in Germany, including factors of influence related to the language input, and to find the most appropriate, that is, "culture-fair" items for the assessment of working memory in BI. As expected, BI scored significantly lower than MO in both chosen NWRT (those from SSV and KiSS), probably (a) because they needed educational assistance in acquiring German significantly more often than MO and (b) due to transfers from their non-German L1. After the exclusion of children with limited German language skills, BI outperformed MO in SSV NWRT and scored on the same level as MO in KiSS NWRT.

Weaker results of BI, compared with MO, in NWRT can be explained in terms of BI's weaker command of the German phonotactics and vocabulary. BI not only scored lower in the respective KiSS subtests but also tended to simplify consonant clusters in the non-words more often than MO. Consonant clusters are highly language-specific and therefore are sometimes criticized as inappropriate for NWRT (Grimm, 2016). Because most non-words in SSV and KiSS contain consonant clusters, apart from German-specific phonemes such as /ʃ/, /ʔ/, and /ŋ/, this might have contributed to lower scores of correct answers in NWRT in the BI subgroup compared with MO. BI scored significantly lower in 6 out of 18 SSV items and in five out of 20 KiSS items. It should be noted that in some alternative NWRT such as "quasi-universal" non-words suggested by Chiat (2015) consonant clusters do not occur at all to avoid this disadvantage caused by consonant clusters for BI.

The dependence of NWRT results on German language skills was confirmed by significant associations between performance on SSV and variables related to the quality and quantity of the German language input. In case of MO, only one very weak correlation was identified: children attending kindergartens for a full day yielded higher scores of correctly repeated non-words than children attending kindergartens only for half a day. In case of BI, most correlations demonstrated significant results. Children with lower total scores of correct answers in NWRT attended kindergartens, nursery schools, associations, and study groups less often than children with higher total scores. They played less often with German-speaking children both in kindergartens and beyond and spoke out less often when playing. A late onset of German language acquisition was

also associated with low results in NWRT. To sum up, in contrast to MO, BI with a limited German language input beyond their household indeed scored lower in NWRT although such tasks were actually designed to assess not German language skills but working memory.

After the exclusion of ED and MED children, BI still scored significantly lower than MO in the vocabulary task but outperformed MO in SSV non-words and performed on the same level as MO in KiSS non-words. As was shown in classification trees and confirmed by Spearman correlations, the SSV NWRT total score was most closely associated with the total score of correct answers in the KiSS subtest on phonology, and KiSS NWRT total score with the KiSS subtest on vocabulary. Because BI still lagged behind in the command of the German vocabulary, but not in phonology, even after the exclusion of ED and MED children, they could not outperform MO in KiSS non-words. However, a comparable level of command of the German phonotactics contributed to the better result of BI, compared with MO, in SSV non-words. Additionally, the rates of omission of consonants in consonant clusters in the BI and MO subgroups became comparable. The exclusion of ED and MED even resulted in higher rates of correct answers in three non-words in the BI subgroup, compared with MO, although only in the items where differences were not statistically significant before the exclusion of children with weak German language skills. MO still scored higher than BI in two out of 38 items.

The finding that BI with normal language skills in L2 can score in working memory tasks on the same level with MO supports results of some previous studies on other language tests. Schöler et al. (2005) found no significant differences between BI and MO in NWRT from the German tests "Heidelberger Auditives Screening in der Einschulungsuntersuchung" (HASE; Brunner & Schöler, 2001) and "Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten" (BISC; Jansen et al., 2002). In some comparatively rare studies, BI even outperformed MO in repetition tasks (Bastian et al., 2018). Because working memory is trainable (Lee et al., 2017), the double pressure imposed by bilingualism on executive functions can obviously contribute to a better development of its subsystems including working memory. In other words, significantly better results of BI, compared with MO, after the statistical deletion of test subjects with limited German language skills in the current study can be attributed to more developed working memory due to the acquisition of two or more languages (Blom et al., 2014).

BI do not constitute a homogeneous group in regard to the quality and quantity of the contact to the German language, and, consequently, in regard to their German language skills. In previous research, children from non-European countries, such as Arabs and Turks, were shown to acquire German under unfavorable sociodemographic conditions and to have

a limited command of the German language, compared with Greek and English-speaking children (Zaretsky & Lange, 2015; Zaretsky et al., 2013). In this study, children speaking European languages also outperformed "non-Europeans" in NWRT, probably due to better German language skills and comparatively small typological differences between their L1 and German. After the exclusion of ED and MED children from both subgroups, no significant differences between "Europeans" and "non-Europeans" in NWRT were found. Omissions of consonants in consonant clusters still occurred more often in the non-European subgroup, but only in one word, compared with three before the exclusion of children with a weak command of German.

Most effect sizes in the comparison of chosen subgroups—BI vs. MO, "Europeans" vs. "non-Europeans"—were low. The same is valid for the correlations between NWRT results and questionnaire items on the quality and quantity of the language input (see Table 1). Indeed, numerous intra- and extralinguistic variables, such as length and phonological complexity of test items (Cilibrasi et al., 2018), children's biological sex (Lange & Zaretsky, 2021), and familial socioeconomic status (Korecky-Kröll et al., 2019), just to name a few, contribute to the performance on NWRT. Obviously, bilingualism is only one of many factors influencing the results of working memory tests.

Performance on tests on working memory in the BI subgroup resembled, to a certain degree, performance of children with SLI from other studies. Phonological representations have not yet been completely formed for L2 in BI and are deficient in children with SLI due to word learning difficulties (Szewczyk et al., 2018). Just like test subjects with SLI (Jones et al., 2010), BI from the current study scored low in the reproduction of nonwords with a high lexical load: items with German morphemes (suffix -*keit* in *Dilecktichkeit*) and those with a high probability of re-etymologization (*Toschlander*, cf. popular sausage trademark *Deutschländer*). In MO, such items activate frequently used phonological representations that do not exist in BI yet.

The following non-words were shown to be comparatively "culture-fair," that is, they put neither BI nor MO at a disadvantage and, thus, probably measured working memory more than German language skills: (a) SSV: *Billop*, *Defsal*, *Ronterklabe*, *Entiergent*, *Krapselistong*, *Nebatsubst*, *Seregropist*, *Pristobierichkeit*, *Ippazeumerink*, (b) KiSS: *Gune*, *Lamposta*, *Kaulen*, *Triser*, *Siraf*, *Schabu*, *Feilister*, *Posalin*, *Verwumpeln*, *Gestrabelt*, *Schrimane*, *Kirumander*, *Wasubritzen*, *Misküranok*, *Nabolira*.

One of the limitations of the study is that BI's competence in their non-German L1 was not assessed. Such tests were not available for most languages spoken by children in the sample and also were not of interest for the original test validation studies that provided data for the retrospective analyses presented here. Therefore, it remained unknown how many children lagged behind in the acquisition of their non-German L1, which would have been an important indicator of language impairments.

To conclude, NWRT delivered in case of BI misleading results. Insufficient German language skills and transfers from L1 impede the interpretation of NWRT results and interfere with a reliable identification of MED children. Numerous German-based non-words from the two chosen tests were shown to put BI at a disadvantage. Therefore, the use of "quasi-universal" non-words such as those proposed by Chiat (2015) is recommended instead of language-specific ones. However, "quasi-universal" items are not yet validated for German, which should be done in future research. Alternatively, comparatively "culture-fair" items such as those identified in this study should be included in working memory tests.

Declaration of Conflicting Interests

Funding

References

Asbrock, D. (2009). Sprachentwicklungsdiagnostik bei mehrsprachigen Kindern im Vorschulalter: Erfassung der Zeitsprache mit dem SETK 3-5 [Diagnostics of the language development in multilingual preschoolers: Language assessment by SETK 3-5]. *Die Sprachheilarbeit* [Works on Speech and Language Therapy], *5*, 197–203. https://doi.org/10.3224/diskurs.v11i1.22252

Bastian, L., Reichenbach, K., Helbig, L., Lenz, K., Rohrbach, S., Pollex-Fischer, D., Schaeuble, M., Gross, M., & Sarrar, L. (2018). Auswirkungen von Bilingualität auf kognitive Funktionen im Vorschulalter [Effects of bilingualism on cognitive functions in early childhood]. *Praxis der Kinderpsychologie und Kinderpsychiatrie* [Practice of the Child psychology and Child Psychatry], *67*(1), 2–17. https://doi.org/10.13109/prkk.2018.67.1.2

Beckman, M. E., Munson, B., & Edwards, J. (2007). Vocabulary growth and the developmental expansion of types of phonological knowledge. In J. Cole & J. I. Hualde (Eds.), *Papers in laboratory phonology 9: Change in phonology* (pp. 241–264). Mouton de Gruyter.

Befi-Lopes, D. M., Pereira, A. C., & Bento, A. C. (2010). Phonological representation of children with Specific Language Impairment (SLI). *Pró-Fono Revista de Atualização Científica* [Scientific Update Journal Pró-Fono], *22*(3), 305–310. https://doi.org/10.1590/S0104-56872010000300025

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. *PLOS ONE,*

11(7), Article 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

Bosma, E., Heeringa, W., Hoekstra, E., Versloot, A., & Blom, E. (2017). Verbal working memory is related to the acquisition of crosslinguistic phonological regularities. *Frontiers in Psychology*, *8*, 1487. https://doi.org/10.3389/fpsyg.2017.01487

Brunner, M., & Schöler, H. (2001). *HASE—Heidelberger Auditives Screening in der Einschulungsuntersuchung [HASE—Heidelberg Auditive Screening for the school enrolment examination]*. Westra.

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.

Cilibrasi, L., Stojanovik, V., Loucas, T., & Riddell, P. (2018). The role of noninitial clusters in the children's test of nonword repetition: Evidence from children with language impairment and typically developing children. *Dyslexia*, *24*(4), 322–335. https://doi.org/10.1002/dys.1599

City Frankfurt/Main. (2017). *Kindergesundheit in Frankfurt am Main [Children's health in Frankfurt/Main]*. Public health department of Frankfurt/Main.

Darcy, I., & Krueger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, *40*, 568–581. https://doi.org/10.1016/j.wocn.2012.05.001

de Abreu, P. M., Baldassi, M., Puglisi, M. L., & Befi-Lopes, D. M. (2013). Crosslinguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, *56*(2), 630–642. https://doi.org/10.1044/1092-4388(2012/12-0079)

District Administration Goeppingen. (2019). *Einschulungsuntersuchung 2017/18 im Landkreis Göppingen [School enrolment examination in 2017/2018 in the district Goeppingen]*. Public health department of Goeppingen.

Ferré, S., dos Santos, C., & Almeida, L. (2015). Potential phonological markers for SLI in bilingual children. In E. Grillo & K. Jepsen (Eds.), *Proceedings of the 39th Boston University conference on language development* (pp. 152–164). Cascadilla Press.

Gathercole, S. E. (1995). The assessment of phonological memory skills in preschool children. *British Journal of Educational Psychology*, *65*(2), 155–164. https://doi.org/10.1111/j.2044-8279.1995.tb01139.x

Gathercole, S. E. (2006). Complexities and constraints in nonword repetition and word learning. *Applied Psycholinguistics*, *27*(4), 599–613. https://doi.org/10.1017/S014271640606053x

Germano, G. D., de, C., César, A. B. P., & Capellini, S. A. (2017). Screening protocol for early identification of Brazilian children at risk for dyslexia. *Frontiers in Psychology*, *8*, 1763. https://doi.org/10.3389/fpsyg.2017.01763

Grimm, A. (2016). Quatschwörter nachsprechen—gleiche Anforderungen für alle Kinder? [Repetition of nonwords: The same requirements for all children?]. *Discourse. Journal of Childhood and Adolescence Research*, *11*(1), 113–117. https://doi.org/10.3224/diskurs.v11i1.22252

Grimm, H. (2003). *SSV—Sprachscreening für das Vorschulalter [SSV—Language screening for the preschool age]*. Hogrefe.

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). Routledge/Taylor & Francis Group.

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

Jansen, H., Mannhaupt, G., Marx, H., & Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC) [Bielefeld screening for the early detection of dyslexia (BISC)]*. Hogrefe.

Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1642–1655. https://doi.org/10.1044/1092-4388(2010/09-0222)

Korecky-Kröll, K., Dobek, N., Blaschitz, V., Sommer-Lolei, S., Boniecki, M., Uzunkaya-Sharma, K., & Dressler, W. U. (2019). Vocabulary as a central link between phonological working memory and narrative competence: Evidence from monolingual and bilingual four-year-olds from different socioeconomic backgrounds. *Language and Speech*, *62*(3), 546–569. https://doi.org/10.1177/0023830918796691

Lange, B. P., & Zaretsky, E. (2021). Sex differences in language competence of four-year-old children: Female advantages are mediated by phonological short-term memory. *Applied Psycholinguistics*, *42*(6), 1503–1522. https://doi.org/10.1017/S0142716421000333

Lee, C. S. C., Pei, J., Andrew, G. A., Kerns, K., & Rasmussen, C. (2017). Effects of working memory training on children born preterm. *Applied Neuropsychology: Child*, *6*(4), 281–296. https://doi.org/10.1080/21622965.2016.1161513

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

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, *19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

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

Nowbakht, M. (2019). The role of working memory, language proficiency, and learners' age in second language English learners' processing and comprehension of anaphoric sentences. *Journal of Psycholinguistic Research*, *48*(2), 353–370. https://doi.org/10.1007/s10936-018-9607-2

Ogino, T., Hanafusa, K., Morooka, T., Takeuchi, A., Oka, M., & Ohtsuka, Y. (2017). Predicting the reading skill of Japanese children. *Brain and Development*, *39*(2), 112–121. https://doi.org/10.1016/j.braindev.2016.08.006

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)

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, *108*(1), 113–125. https://doi.org/10.1016/j.jecp.2010.07.011

Pham, G., Ebert, K. D., Dinh, K. T., & Dam, Q. (2018). Nonword repetition stimuli for Vietnamese-speaking children. *Behavior Research Methods*, *50*(4), 1311–1326. https://doi.org/10.3758/s13428-018-1049-0

Rodrigues, A., & Befi-Lopes, D. M. (2013). Short-term phonological memory in preschool children. *CoDAS*, *25*(5), 422–428. https://doi.org/10.1590/S2317-17822013000500005

Savill, N., Cornelissen, P., Whiteley, J., Woollams, A., & Jefferies, E. (2019). Individual differences in verbal short-term memory and reading aloud: Semantic compensation for weak phonological processing across tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1815–1831. https://doi.org/10.1037/xlm0000675

Schöler, H., Haselbach, P., Schaefer, P., Dressler, A., & Engler-Thuemmel, H. (2005). *Zur Wirksamkeit von Maßnahmen zur Förderung der sprachlichen Fähigkeiten in der Vorschule Mannheim [On the efficacy of language courses at a preschool in Mannheim]*. (Arbeitsberichte aus dem Forschungsprojekt "Differenzialdiagnostik," *Vol. 20*). Heidelberg University of Education.

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Willinger, U., Schmoeger, M., Deckert, M., Eisenwort, B., Loader, B., Hofmair, A., & Auff, E. (2017). Screening for specific language impairment in preschool children: Evaluating a screening procedure including the token test. *Journal of Psycholinguistic Research*, *46*(5), 1237–1247. https://doi.org/10.1007/s10936-017-9493-z

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., & Lange, B. P. (2015). Ob Italienisch Deutsch fördert: Warum Italienisch sprechende Kinder schneller Deutsch erwerben als einige andere Migrantengruppen in Deutschland [Whether Italian language competence contributes to the acquisition of German: Why Italian speaking children acquire German more quickly than some other immigrant groups in Germany]. *Dutch Journal of Applied Linguistics*, *4*(1), 122–140. https://doi.org/10.1075/dujal.4.1.11zar

Zaretsky, E., Lange, B. P., & Hey, C. (2020). Sprachstand 4-jähriger Kinder mit Migrationshintergrund in Hessen: quo vadis? [Language competence of four-year-old children with immigration background in Hesse: quo vadis?]. *Laryngo-Rhino-Otologie*, *99*(5), 308–315. https://doi.org/10.1055/a-1119-6753

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 [German plural acquisition by Russian and Turkish speaking children compared to other immigrants and monolingual native speakers]. *Zeitschrift für Slawistik [Journal on Slavic studies]*, *58*(1), 43–71. https://doi.org/10.1524/slaw.2013.0003