Identifying Specific Language Impairment (SLI) across Different Bilingual Populations: German Sentence Repetition Task (SRT)

Lina Abed Ibrahim, Cornelia Hamann, and David Öwerdieck

1. Introduction

Recent research in language disorders has focused on the problem of identifying specific language impairment (SLI) in bilingual populations. This is indeed a problem since difficulties in the acquisition of a second language (L2) or of two first languages (2L1) may concern the same linguistic structures as those difficult for monolingual children with SLI (MoSLI). Bilingual children are therefore often misdiagnosed as language impaired (Genesee, Paradis & Crago 2004). In order to address this problem, LITMUS tools were developed (Language Impairment Testing in Multilingual Settings, see Armon-Lotem, de Jong & Meir 2015) during COST Action IS0804, among which are sentence repetition tasks (SRTs) incorporating structures known to be cross-linguistically difficult for children with SLI and structures identified as difficult in a particular language (Marinis & Armon-Lotem 2015).

Repetition tasks like nonword repetition and sentence repetition have proved to have high diagnostic accuracy for identifying SLI in monolingual children (Conti-Ramsden et al. 2001). Although such repetition tasks usually evaluate working-memory (Archibald & Gathercole 2006), they also involve phonological and/or syntactic derivation or representation. Such linguistic derivations/representations, and their complexity in particular, have been shown...
to crucially influence performance on these tasks (Szterman & Friedmann 2015, Gallon et al. 2007) so that it can be argued that such tasks are not mere measures of working memory (Polišenská et al. 2014; Vinther 2002). In the present paper, we focus on sentence repetition because it targets morphosyntax, which is a main area of difficulty for children with SLI (Leonard 2014). Additionally, SRTs are versatile and allow testing (or contrasting) particular structures while minimizing avoidance strategies (see Marinis & Armon-Lotem 2015, Hamann et al. 2017). Testing for mastery of structures vulnerable in children with SLI but usually available for typical bilinguals, such as subordination or Wh-movement (with or without intervening elements), can thus easily be accomplished. Finally, SRTs are fast and easy to administer and offer different rating methods, which allow global scores as well as fine-grained analysis.

It has recently been shown for different language combinations that LITMUS-SRTs have good diagnostic accuracy in bilinguals (Armon-Lotem & Meir 2016 for Russian/Hebrew bilinguals, de Almeida et al. 2017 for Arabic/French, Portuguese/French and Turkish/French bilinguals, Hamann & Abed Ibrahim 2017 for Arabic/German, Portuguese/German and Turkish/German bilinguals). In particular, Abed Ibrahim & Hamann (2017) and Hamann & Abed Ibrahim (2017) have shown that different rating methods, especially those discounting lexical errors and other bilingual errors such as gender assignment or case2 in German, achieve good diagnostic accuracy in the bilingual populations investigated so far.

In the current study, we want to verify these findings on larger groups of children whose status as language impaired or typically developing has been determined by a comprehensive battery of standardized baseline measures3 in both L1 and L2. Crucially; however, we want to know whether testing in the majority language (L2 German) could disadvantage children with specific home languages, here Arabic, Portuguese or Turkish. In other words, we want to explore if there is an influence of typological differences on the performance in the German SRT and how such an influence could manifest itself in particular German structures or in specific error patterns. As to the latter, it is possible that L1 influence sometimes observed in spontaneous production can also be observed in repetition tasks. Such influence could be expected when certain language features of the L2 differ from L1 features. Thus, problems could be expected for Turkish L1-speakers with overt auxiliaries and determiners since Turkish lacks both categories. German determiners might also be used erroneously because features such as case or gender are overtly marked on this category in German, whereas they are morphologically marked elsewhere in the L1 as in the case of Arabic. For the same reason, the differentiation of German complementizers and relative pronouns could be a locus of difficulty for children with L1 Portuguese because Portuguese uses the complementizer *que* in both contexts. We therefore

---

2 Case errors are not disregarded if they are necessary for the realization of the targeted structure, e.g. object relatives and topicalized sentences.

3 Spontaneous speech samples were consulted in borderline cases.
want to know whether such possible L1-L2 interference will influence performance on a sentence repetition task.

Our most important finding is that even though particular structures may be more (or less) challenging for children with specific home languages, the different groups of bilinguals do not differ in their overall performance in the German SRT. The SRT can therefore be used confidently across language combinations and will prove a valuable tool for language assessment in the majority language.

The paper is structured as follows. The method section first provides more detailed information about the different conditions used in the German LITMUS SRT in 2.1. and, in 2.2., proceeds to describe the participants and the way their clinical status was determined by norm referenced L1 and L2 tests. 2.3. gives details about the data analysis and the statistical tests applied. Section 3 presents the results and section 4 provides discussion and conclusion.

2. Method
2.1. The German LITMUS Sentence Repetition Task

The German SRT (Hamann et al. 2013), developed within COST Action IS0804, was constructed according to the LITMUS (Language Impairment Testing in Multilingual Settings) principles (Marinis & Armon-Lotem 2015). It was developed in parallel to the French LITMUS-SRT (Fleckstein et al. 2016; de Almeida et al. 2017) and incorporates complex constructions known to challenge children with SLI cross-linguistically, in addition to structures representing crucial milestones in the acquisition of the properties of German morphosyntax. The former include object questions, subject and object relative clauses, finite complement clauses and long passives. The latter include structures reported to be vulnerable in monolingual German children with SLI, such as topicalization and the sentence bracket, which reflects the V2 and OV properties of German (see Hamann et al. 2017 and Hamann & Abed Ibrahim 2017 for details).

The German LITMUS-SRT under investigation consists of 45 sentences divided in three levels of increasing complexity (five conditions per level and three test items per condition). The sentences of each level are controlled for syllable number and vocabulary. The level of structural complexity in the task depends on the presence of complexity factors such as Wh-movement, clausal embedding, intervention in addition to the presence of two propositions. Level 1 contains simple monoclausal declaratives focusing on SVA, tense and the sentence bracket (see (1)). Level 2 contains object questions with either a non-discourse-linked wh-operator (Wen ‘who-masc-acc’) or a D-linked wh-operator (Welchen ‘which-masc-acc.’) with an intervening lexical noun phrase (see (2a) & (2b)). Level 2 further includes finite (see (3)) and non-finite complement clauses, which are contrasted with coordinate clauses. Level 3 tests long passives, subject and object relative clauses with (see (4)) and without intervention as well as topicalization (see (5)). A summary is given in table 1.
(1) **Sentence bracket**
Der Prinz *hat* die Prinzessin umarmt
The/nom. prince *has* the/acc. Princess hugged
‘The prince hugged the princess’

(2a) **Bare-WH**
Wen beißt der große Löwe immer
Who/acc. bites the/nom. big lion always?
‘Whom does the big lion always bite?’

(2b) **Which-NP**
Welchen Bauern ärgert der Affe?
Which/acc. farmer annoys the/nom. monkey?
‘Which farmer does the monkey annoy?’

(3) **Finite complement clause**
Der Wikinger glaubt, dass die Hexe ihn mag
The/nom. viking believes that the/acc. witch him likes
‘The viking believes that the witch likes him’

(4) **Object relative with intervention**
Ich sehe den Vogel, *den der Pinguin* weckt
I see the/acc. bird who/acc. the/nom. penguin wakes-up
‘I see the bird who(m) the penguin wakes up’

(5) **Topicalization**
Den Arzt fotografiert der Bauer gerne
The/acc. doctor photographs the/nom. farmer gladly
‘The doctor, the farmer photographs gladly’

**Table 1: Overview of German LITMUS-SRT test items**

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVO-Present</td>
<td>Object Who-question (bareWH)</td>
<td>Passive</td>
</tr>
<tr>
<td>SVO-Simple past</td>
<td>Object Which-question (Wh-NP)</td>
<td>Topicalization</td>
</tr>
<tr>
<td>Sentence bracket (Aux)</td>
<td>Coordination (Coord)</td>
<td>Subject relatives (SR)</td>
</tr>
<tr>
<td>Sentence bracket (Particle)</td>
<td>Non-finite Complement Cl</td>
<td>Object relatives without intervener (OR –intv.)</td>
</tr>
<tr>
<td>„Werden“ control</td>
<td>Finite complement Cl</td>
<td>Object relatives with intervener (OR +intv.)</td>
</tr>
<tr>
<td></td>
<td>(CompFin)</td>
<td></td>
</tr>
</tbody>
</table>
The stimuli sentences are pre-recorded and presented to the child in a pseudorandomized order using a child friendly PowerPoint presentation. The task is to be administered in full length with no abort criteria and takes about 10 minutes. Before administering the task, the child is provided with two practice items. The test sentences can be scored by ‘identical repetition’, i.e. scoring the whole response to a test item as correct or incorrect (0 = incorrect; 1 = correct). This type of rating only disregards phonological errors and is considered to be the fastest and easiest method of scoring. Alternatively, the task can be scored by ‘target structure’, which measures the mastery of a particular structure and compensates for L2-errors such as lexical substitutions and systematic recurrent case and gender errors not crucial for the realization of the targeted structure (see Hamann & Abed Ibrahim 2017 for details).

2.1.1. German LITMUS SRT and potential L1-influenced error patterns

The German LITMUS SRT was thoroughly designed in order not to penalize bilingual children. This was achieved by avoiding areas of overlap between BiTDs and MoSLIs as far as possible. For this reason, the task does not explicitly test for SVA or case and gender. However, given the typological differences between German and the three L1s under investigation (Arabic, Portuguese and Turkish), potential loci of L1-influence in the SRT could be predicted (cf. Chilla et al. submitted). For example, both Arabic and Turkish lack auxiliaries and the latter lacks determiners (D) as well. Although determiners are available in both Arabic and Portuguese, they differ from German in the grammatical features overtly marked on D. German determiners are overtly marked for definiteness, gender and case. Arabic, on the other hand, possesses only definite determiners nonbearing of case or gender marking, while Portuguese determiners are marked for definiteness and gender but not for case. Based on the aforementioned typological differences, we decided to look into the following possible error patterns: Aux-omission, D-omission, case and gender errors on D. Here, error rates were calculated as percentage of erroneous responses out of all possible contexts produced by a child.

2.2. Participants

The present study reports on the performance of 78 children recruited for the BiLaD project on the German LITMUS SRT. The age range of the participating children is 5;6-9;4 yrs. covering the last year of kindergarten and the crucial first 2-3 primary school years. The subjects of the current study are divided into four groups: monolingual typically developing children (MoTD), monolingual children with specific language impairment (MoSLI), bilingual typically developing children (BiTD) and bilingual children with specific language impairment (BiSLI). The latter two groups include both simultaneous and

---

4 The subjects in this study were drawn from a larger participant pool (N=92) recruited for the BiLaD project in Germany.
sequential bilingual children with Arabic, Turkish or European Portuguese as first language (L1). All of the bilingual children were exposed to L2 German for at least 24 months except for one German-Portuguese BiTD child (LoE= 18 months). Monolingual and bilingual children clinically diagnosed with SLI were recruited from specialized speech-language pathology centers and kindergartens. Verification of their status as MoSLI or BiSLI started with a control for non-verbal intelligence using the German version of Raven’s Coloured Progressive Matrices (CPM; Bullheller & Hacker 2002). Only children who scored above percentile rank 9 (IQ score ≥ 80 according to Wechsler’s IQ scale) were included in the study. Given the risk of over- and underdiagnosis especially in case of bilinguals (Grimm and Schulz 2014), the clinical status of the subjects as ± language impaired (refined clinical status) was verified using a battery of baseline language measures in both their L1 and L2 (see Hamann & Abed Ibrahim 2017 for particulars). Background information about the children’s early language development, early and current exposure to and use of L1/L2 necessary for determining language dominance and type of child bilingualism were collected using the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller 2015). We adapted the criteria for MoSLIs outlined in Leonard (1998) and assigned a child to the BiSLI group, if she scored below dominance-adjusted norms (Thordardottir 2015) in two language domains (on norm-referenced tests) in both of her L1 and L2.

Table 2 provides a participant overview based on refined clinical status. In this table, the BiTD children are subgrouped based on their L1 into BiTD-A, BiTD-T and BiTD-P. The BiSLI children, on the other hand, are grouped together regardless of their home language due to the small sample size.

Table 2: Participants including monolingual children and verified clinical status of bilingual children as BiTD/BiSLI: age & Length of exposure (mo.), colored progressive matrices (CPM) scores (percentile ranks)

<table>
<thead>
<tr>
<th></th>
<th>MoTD (n=10)</th>
<th>MoSLI (n=12)</th>
<th>BiTD (n=45)</th>
<th>BiSLI (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simult./total</td>
<td></td>
<td>BiTD-A (n=10)</td>
<td>BiTD-P (n=19)</td>
</tr>
<tr>
<td>Age at testing</td>
<td>75.90 (8.99)</td>
<td>81.75 (13.41)</td>
<td>88.60 (13.5)</td>
<td>83.10 (14.38)</td>
</tr>
<tr>
<td>Length of Exposure (Gr.)</td>
<td>75.90 (8.99)</td>
<td>81.75 (13.41)</td>
<td>53.2 (19.68)</td>
<td>66.53 (24.70)</td>
</tr>
<tr>
<td>CPM</td>
<td>81.20 (13.98)</td>
<td>53.41 (23.39)</td>
<td>47.34 (30.91)</td>
<td>71.63 (20.59)</td>
</tr>
</tbody>
</table>

5 In case of sequential bilinguals, systematic exposure to L2 (within an institutional setting, e.g. kindergarten) started at approximately age three or later (cf. Hamann & Abed Ibrahim 2017).
The four groups (MoTD, MoSLI, BiTD, BiSLI) did not statistically differ in age at testing or in non-verbal intelligence as revealed by omnibus Kruskall-Wallis test: age at testing \[\chi^2 (3, N = 78) = 5.492, p = 0.139\], CPM: \[\chi^2 (3, N = 78) = 7.402, p = 0.06\]. This also holds for the BiTD groups when split by L1 and compared to the BiSLI group: age at testing \[\chi^2 (3, N = 56) = 2.286, p = 0.515\], CPM: \[\chi^2 (3, N = 56) = 7.425, p = 0.06\]. Furthermore, no significant differences were found between the BiTD groups in terms of length of exposure to the L2 (LoE): \[\chi^2 (3, N = 56) = 3.640, p = 0.303\].

2.3. Data Analysis

The children’s SRT responses were audio-recorded with special dictaphones. The data were transcribed offline, verified and coded for errors by at least two independent linguistically trained raters. Both scoring methods ‘identical repetition’ and ‘target structure met’ were used for the evaluation of the SRT data. IBM SPSS 24 (2016) was used to conduct all statistical analyses. Non-parametric tests were used for group comparisons throughout the paper due to unequal sample sizes and violation of normality assumption (Shapiro-Wilk test). Receiver Operating Characteristic (ROC; Dunn 2011) curve analysis was carried out to determine the optimal cut-off scores associated with the highest diagnostic accuracy for each of the monolingual and bilingual groups separately. This enabled calculating the sensitivity (probability that a language impaired child will be identified as such by a test, i.e. true positive rate) and specificity (probability that a typically developing child will be identified as such by a test, i.e. true negative rate). Moreover, likelihood ratios (LRs), which incorporate both the specificity and sensitivity of a test, were manually calculated for the cut-off scores obtained by the ROC analysis. LRs have the advantage that they are less likely to be influenced by variations in the properties of the test sample (Dollaghan 2004). An LR+ (sensitivity/(1-specificity)) is indicative of the likelihood of a score below a cut-off criterion to be present in language impaired children; whereas an LR- ((1-sensitivity)/specificity) indicates the likelihood of a score above a cut-off criterion to occur in children with typical language development (see Hamann & Abed Ibrahim 2017 for details). Partial correlations and multiple linear regression analysis were used to determine the potential influence of AoO, LoE and language dominance on the performance of the BiTDs.

---

6 This applies to global scores, scores on SRT substructures as well as error patterns.
7 Good diagnostic accuracy is given when sensitivity and/or specificity rates are ≥90%. Sensitivity and/or specificity rates between 80% and 89% are considered fair (Plante and Vance 1994).
8 LRs+ ≥10 are highly indicative of the presence of language impairment, LRs- ≤0.10 are highly indicative of the absence of language impairment, LRs+ ≥3.0 and LRs- > 3.0 are clinically suggestive, whereas LRs+ < 3.0 and LRs- > 0.3 are clinically uninformative (cf. Dollaghan 2007).
3. Results

As stated in the introduction, the main goal of this study is to determine whether the German LITMUS-SRT (using global scores obtained by both scoring measures ‘identical repetition’ and ‘target structure met’) is capable of identifying SLI in bilingual children from different L1 backgrounds. A further objective of the study is to investigate whether bilingual typically developing children differ in their global performance on the SRT from their monolingual peers and from each other when split by L1. We further want to know if any of the SRT substructures pose more difficulties for BiTDs with a particular L1 background and whether specific error patterns found in the performance of the BiTDs are influenced by their L1. Finally, we want to know whether scoring by the realization of target structure rather than by whole item accuracy is better suited for the assessment of bilingual children.

3.1. Overall results on the German LITMUS-SRT

We first ran Omnibus Kruskal–Wallis tests using the overall scores SRT_identical repetition (SRT_Id) and SRT_target structure met (SRT_Tar) as dependent variables to determine whether clinical group membership had a statistical effect. Both measures gave significant results SRT_Id: $[\chi^2 (3, N = 78) = 40.636, p < 0.001]$, SRT_Tar: $[\chi^2 (3, N = 78) = 42.548, p < 0.001]$. As the next step, pairwise comparisons were run using post hoc Mann-Whitney U test. In order to reduce Type I error arising due to multiple comparisons, Bonferroni-adjustment was applied to the obtained $p$-values. The overall performance of the groups is depicted in Figure 1. On the whole task, monolingual and bilingual typically developing children performed significantly better than their peers with SLI. The task sharply distinguishes between MoTD and MoSLI: SRT_Id ($U= 0.000, p < 0.001, r=0.847$), SRT_Tar ($U= 0.000, p < 0.001, r=0.846$), BiTD and BiSLI SRT_Id ($U= 47.000, p < 0.001, r=0.553$), SRT_Tar ($U= 29.500, p < 0.001, r=0.601$) and most importantly between MoSLI and BiTD: SRT_Id ($U= 35.500, p < 0.001, r=0.584$), SRT_Tar ($U= 24.500, p < 0.001, r=0.615$). No significant differences were found between MoSLI and BiSLI. A significant difference with medium effect size was found between the MoTDs and BiTDs for the scoring method ‘identical repetition’: SRT_Id ($U= 86.500, p < 0.05, r=0.408$). On the other hand, performance in SRT does not significantly differ in MoTDs and BiTDs if the task is scored by ‘target structure met’. It should be noted that three bilingual children (33, 53 and 41) perform within the BiSLI range on SRT_Id but only 41 performs BiSLI-like on SRT_Tar.
Fig. 1: SRT: % identical repetition and correct production of target structure

ROC curve analyses revealed excellent diagnostic accuracy with 100% sensitivity and specificity for both SRT_Id and SRT_Tar in monolingual children. As can be seen in Table 3, it emerges that the optimal cut-off point for SRT_Id (63.3%) is lower than that for SRT_Tar (77.7%). ROC analyses conducted on the bilingual groups are also given in Table 3. Similar to the monolinguals, the cut-off point for ‘identical repetition’ was lower than that for ‘target structure met’ (SRT_Id: 36.6%; SRT_Tar: 47.7%) but the optimal cut-off points for bilinguals were generally much lower compared to monolinguals. Diagnostic accuracy was fair for SRT_Id and good for SRT_Tar. Due to better specificity, diagnostic accuracy for SRT_Tar was higher in case of bilingual children.

Table 3: Diagnostic accuracy of the German LITMUS-SRT in monolingual and bilingual children: identical repetition vs. correct target structure

<table>
<thead>
<tr>
<th></th>
<th>Monolinguals</th>
<th>Bilinguals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SRT_Id</td>
<td>SRT_Tar</td>
</tr>
<tr>
<td>Cut-off</td>
<td>63.3%</td>
<td>77.7%</td>
</tr>
<tr>
<td>AUC</td>
<td>1.000</td>
<td>0.996</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>100% (12/12)</td>
<td>100% (12/12)</td>
</tr>
<tr>
<td>Specificity</td>
<td>100% (10/10)</td>
<td>100% (10/10)</td>
</tr>
<tr>
<td>LR+</td>
<td>undefined</td>
<td>undefined</td>
</tr>
<tr>
<td>LR-</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Partial correlation controlling for age at testing revealed a strong positive correlation between L2 dominance and performance on SRT in the BiTD group: SRT_Id: \( r = .558, p < 0.001 \), SRT_Tar: \( r = .580, p < 0.001 \). Moderate correlations

\(^9\) *If specificity = 100% then LR+ ratios are undefined.*
were found between performance on SRT and AoO: SRT_Id: $r = -0.371, p = 0.013$, SRT_Tar: $r = -0.315, p = 0.037$ as well as for LoE: SRT_Id: $r = 0.370, p = 0.014$, SRT_Tar: $r = 0.311, p = 0.040$. A multiple linear regression model was developed for predicting the performance of the BiTDs in SRT_Tar based on AoO, LoE and L2 dominance index. The three predictor model was significant ($F(3,41) = 8.704, p < 0.001$) with an $R^2$ of 0.389. The results show; however, that only the L2 dominance index had significant partial effects in the full model: L2_dominance_index ($\beta = 0.528, T = 3.972, p < 0.001$) AoO ($\beta = 0.230, T = 1.302, p = 0.200$), LoE ($\beta = 0.352, T = 1.981, p = 0.054$).

In a next step, we compared the performance of bilingual typically developing children with L1 Arabic (BiTD-A), Portuguese (BiTD-P) and Turkish (BiTD-T) to MoTDs, MoSLIs and BiSLIs. Recall that the bilingual children with SLI were collapsed together due to the small sample size. As can be seen in Figure 2, the three BiTD groups perform significantly better than either MoSLIs or BiSLIs on both SRT_Id and SRT_Tar. Moreover, no significant differences were found between the BiTD subgroups on either SRT_Id or SRT_Tar. However, the MoTD group performed significantly better than the BiTD-P group on both measures. Table 4 gives a summary of the pairwise between group comparisons with significant results.

It is noteworthy that the statistical outliers (Figure 2) 30, 31, 33, 41, 48, and 53 are all dominant in the minority language (L1-dominant) and two of them, namely 48 and 31 are late successive bilinguals with AoO 6;3 and 9;0 yrs. respectively.

### Table 4: Sign. comparisons: SRT_Id & SRT_Tar (BiTDs split by L1)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>N</th>
<th>U</th>
<th>Z</th>
<th>p</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiTD-A/MoSLI SRT_Id</td>
<td>21</td>
<td>6.000</td>
<td>-3.470</td>
<td>&lt;0.001</td>
<td>0.76</td>
</tr>
<tr>
<td>BiTD-A/MoSLI SRT_Tar</td>
<td>21</td>
<td>3.000</td>
<td>-3.669</td>
<td>&lt;0.001</td>
<td>0.80</td>
</tr>
<tr>
<td>BiTD-P/MoSLI SRT_Id</td>
<td>30</td>
<td>10.500</td>
<td>-4.049</td>
<td>&lt;0.001</td>
<td>0.74</td>
</tr>
<tr>
<td>BiTD-P/MoSLI SRT_Tar</td>
<td>30</td>
<td>6.000</td>
<td>-4.242</td>
<td>&lt;0.001</td>
<td>0.77</td>
</tr>
<tr>
<td>BiTD-T/MoSLI SRT_Id</td>
<td>27</td>
<td>19.000</td>
<td>-3.415</td>
<td>&lt;0.001</td>
<td>0.66</td>
</tr>
<tr>
<td>BiTD-T/MoSLI SRT_Tar</td>
<td>27</td>
<td>15.500</td>
<td>-3.580</td>
<td>&lt;0.001</td>
<td>0.69</td>
</tr>
<tr>
<td>BiTD-A/BiSLI SRT_Id</td>
<td>21</td>
<td>9.000</td>
<td>-3.251</td>
<td>0.006</td>
<td>0.71</td>
</tr>
<tr>
<td>BiTD-A/BiSLI SRT_Tar</td>
<td>21</td>
<td>2.000</td>
<td>-3.737</td>
<td>&lt;0.001</td>
<td>0.82</td>
</tr>
<tr>
<td>BiTD-P/BiSLI SRT_Id</td>
<td>30</td>
<td>16.500</td>
<td>-3.791</td>
<td>&lt;0.001</td>
<td>0.69</td>
</tr>
<tr>
<td>BiTD-P/BiSLI SRT_Tar</td>
<td>30</td>
<td>9.000</td>
<td>-4.112</td>
<td>&lt;0.001</td>
<td>0.75</td>
</tr>
<tr>
<td>BiTD-P/MoTD SRT_Id</td>
<td>29</td>
<td>28.5</td>
<td>-2.963</td>
<td>&lt;0.05</td>
<td>0.60</td>
</tr>
<tr>
<td>BiTD-P/MoTD SRT_Tar</td>
<td>29</td>
<td>30.5</td>
<td>-2.965</td>
<td>&lt;0.05</td>
<td>0.55</td>
</tr>
<tr>
<td>BiTD-T/BiSLI SRT_Id</td>
<td>27</td>
<td>21.500</td>
<td>-3.286</td>
<td>&lt;0.001</td>
<td>0.63</td>
</tr>
<tr>
<td>BiTD-T/BiSLI SRT_Tar</td>
<td>27</td>
<td>18.500</td>
<td>-3.432</td>
<td>&lt;0.001</td>
<td>0.66</td>
</tr>
</tbody>
</table>

### 3.2. Error patterns in BiTDs with different source languages

In order to investigate the potential influence of cross-linguistic typological differences between the L1s and L2 German on the performance of BiTDs on German LITMUS-SRT, we compared the results of the three BiTD groups to their monolingual peers on all 15 substructures incorporated in the German LITMUS-
SRT. Although BiTDs with L1 Turkish and Portuguese perform worse than MoTDs and BiTDs with L1 Arabic on both types of object relative clauses and topicalization (see Figure 3), this difference does not reach statistical significance using either scoring method. Interestingly, none of the bilingual groups encountered more difficulties with object wh-questions, which also involve non-canonical word order. A closer look at the error patterns in TOP and ORs revealed that case errors (on the relative pronoun in case of ORs) rather than word order errors were the main source for non-target responses. In addition to case errors on relative pronouns in ORs, BiTD-Ps demonstrated insecurity about the choice of relative pronoun, e.g. overuse of feminine gender, substitution of the relative pronoun by a complementizer or a prepositional phrase. The latter was only observed in two late sequential BiTD-P children. Three bilingual children with an AoO > 48 mo. omitted relative pronouns as well.

We further looked for BiTD error patterns that could be ascribed to morphosyntactic influence from L1, such as omission of auxiliary verbs and determiners as well as case and gender marking on determiners (see Chilla et al. submitted for an overview). Aux omission did not occur in the BiTD data set at all. As depicted in Figure 4, D-omission was found only in 2% of the determiner phrases (DP) produced by BiTD-Ts. Case followed by gender errors were the most frequent error types on DPs in all BiTD groups. It should, however, be emphasized that no significant differences were found between the monolingual and bilingual TD groups concerning errors on D, while both SLI groups omitted significantly more determiners than BiTDs. Additionally, unlike BiSLIs, MoSLIs committed significantly more case errors than BiTDs: MoSLI/BiTD: D omission: \((U = 923.500, p < 0.001, r=0.53)\), case on D: \((U = 81.500, p < 0.001, r=0.46)\), BiSLI/BiTD: D omission: \((U = 1538.000, p < 0.05, r=0.42)\).

Note that scoring by target structure met in case of TOP and ORs does not compensate for L2-errors since case marking is crucial for the realization of the target structure.
4. Discussion and Conclusion

The present study examined the diagnostic accuracy of the German LITMUS-SRT in bilingual children with different L1 source languages. In line with previous research (Hamann & Abed Ibrahim 2017; Tuller et al., to appear), results show that the task does not only reliably distinguish SLI in monolingual children (sensitivity = 100%, specificity = 100% for both SRT_Id and SRT_Tar), but also in bilingual ones with good diagnostic accuracy for SRT_Tar (sensitivity = 91%, specificity = 89%) and fair diagnostic accuracy for SRT_Id (sensitivity = 91%, specificity = 80%). Importantly, BiSLI children do not perform worse than their monolingual peers on either measure in the SRT.

The results further indicate that the task does not disadvantage bilingual children, especially when scored by ‘target structure met’, as this measure compensates for lexical substitutions and systematic recurrent case and gender errors. This emerges clearly from the fact that the significant difference observed between MoTDs and BiTDs in SRT_Id is no longer visible once the task is scored by correct target structure. Thus, SRT_Tar appears to be a more appropriate method for bilinguals in German. Although the majority of L1-dominant children perform above cut-off, language dominance seems to influence the performance of the BiTD children as suggested by the results of regression analysis (see also Hamann & Abed Ibrahim 2017).

A further aim of the study was to investigate a possible typological influence of the different home languages (Arabic, Portuguese, Turkish) by examining quantitative differences in the overall performance of MoTDs and the three different bilingual groups BiTD-A, BiTD-P and BiTD-T. Our results show no significant differences in the performance of these three BiTD groups on SRT_Id and SRT_Tar. The comparison with MoTDs showed a significant difference only to the BiTD-P, not the two other groups. However, each of the BiTD subgroups performed significantly better than MoSLIs and BiSLIs. Given these results, we conclude that the task does not disadvantage bilingual children from any particular home language.

Additionally, we wanted to know whether any SRT substructures were more problematic for bilingual children with a particular L1 background and whether specific types of errors were influenced by the L1. Although no significant between-group (MoTD, BiTD-A, BiTD-P and BiTD-T) differences were observed on any of the 15 structures in either of the scoring measures, we still found that BiTD-T and BiTD-P particularly struggled with topicalization and object relatives. Good performance of these same children in object questions rules out an explanation involving difficulties with non-canonical word order. Error analysis shows that the primary source for non-target responses on the latter structures is incorrect case marking closely followed by gender errors on determiners or relative pronouns. Specific errors, which could be expected due to L1 properties such as aux omission and errors on determiners (D-omission, incorrect case and gender marking) did not influence overall performance of any of the BiTD groups on the SRT. Crucially, such errors were mainly found in the MoSLI and BiSLI data (see Figure 4).
To conclude, the German SRT reliably identifies SLI in monolinguals and bilinguals. In addition to that, children with specific home languages are not disadvantaged by the task, since L1 influences surfaced neither in specific structures of the SRT nor in the expected error patterns. Finally, we do not see a cumulative effect of bilingualism and SLI in this task. These results strongly suggest that SRTs including syntactically complex structures can be developed into promising clinical screening tools.

References


Plante, Elena and Vance, Rebecca (1994). Diagnostic accuracy of two tests of preschool language. *AJSLP* 4, 70–76. doi:10.1044/1058-0360.0402.70


Proceedings of the 42nd annual Boston University Conference on Language Development

edited by Anne B. Bertolini and Maxwell J. Kaplan

Cascadilla Press  Somerville, MA  2018

Copyright information

Proceedings of the 42nd annual Boston University Conference on Language Development © 2018 Cascadilla Press. All rights reserved

Copyright notices are located at the bottom of the first page of each paper. Reprints for course packs can be authorized by Cascadilla Press.

ISSN 1080-692X
ISBN 978-1-57473-186-6 (2 volume set, library binding)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com