



**HAL**  
open science

## The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German

Solveig Chilla, Cornelia Hamann, Philippe Prévost, Lina Abed Ibrahim, Sandrine Ferré, Christophe dos Santos, Racha Zebib, Laurice Tuller

### ► To cite this version:

Solveig Chilla, Cornelia Hamann, Philippe Prévost, Lina Abed Ibrahim, Sandrine Ferré, et al.. The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German. *LITMUS in Action: Comparative Studies across Europe*, 29, John Benjamins Publishing Company; John Benjamins Publishing Company, pp.228-262, 2021, Trends in Language Acquisition Research, 9789027209795. 10.1075/tilar.29.08chi . hal-03404644

**HAL Id: hal-03404644**

**<https://hal.science/hal-03404644>**

Submitted on 5 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

This author-accepted manuscript is under copyright.

Solveig Chilla, Cornelia Hamann, Philippe Prévost, Lina Abed Ibrahim, Sandrine Ferré, et al.. The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German. *LITMUS in Action: Comparative Studies across Europe*, 29; John Benjamins Publishing Company, pp.228-262, 2021, Trends in Language Acquisition Research, 9789027209795.  
( <https://benjamins.com/catalog/tilar.29.08chi>)  
(<https://doi.org/10.1075/tilar.29.08chi>)

**The influence of different first languages on LITMUS nonword-repetition and sentence repetition in second language French and second language German: a crosslinguistic approach**

Solveig Chilla\*, Cornelia Hamann<sup>+</sup>, Philippe Prévost<sup>#</sup>, Lina Abed Ibrahim<sup>+</sup>, Sandrine Ferré<sup>#</sup>,  
Christophe dos Santos<sup>#</sup>, Racha Zebib<sup>#</sup> & Laurice Tuller<sup>#</sup>

\*University of Flensburg, Flensburg, Germany

<sup>+</sup>University of Oldenburg, Oldenburg, Germany

<sup>#</sup>UMR 1253, iBrain, Université de Tours, Inserm, Tours, France

**Key words**

specific language impairment, bilingualism, cross–linguistic influence, nonword repetition, sentence repetition

**Abstract**

The study investigates the appropriateness of the French and German LITMUS–Sentence Repetition (LITMUS–SR) and the LITMUS–Quasi–Universal–Nonword–Repetition task (LITMUS–QU–NWR) for the assessment of bilingual children with diverse language backgrounds. We examine the quantitative and qualitative performance of 117 bilingual typically developing children (BiTD), and 34 bilingual children with specific language impairment (BiSLI) with Arabic, Portuguese and Turkish as their first language (L1). These

data are compared to 47 monolingual typically developing children (MoTD). Given the typological and linguistic differences between the languages involved, the comparisons of the performance of bilingual and monolingual children in two typologically different L2s are intended to provide a baseline for typical L2 development and may shed light on morphosyntactic difficulties that are due to cross-linguistic influence.

## 1. Introduction

Language assessment of bilinguals is particularly difficult because children acquiring a second language (L2) in childhood differ from monolingual age-matched peers in several respects. It has been shown that typically developing bilingual children (BiTD), and minority language children in particular, perform significantly lower than their monolingual peers (MoTD) on standardized tests of their L2 normed on monolingual children (e.g., Restrepo, 1998).

In the area of morphosyntax, certain linguistic patterns deviating from those of typically developing monolingual children have been reported for children acquiring L2 French or German. These distinctive patterns often overlap with those known for children with Specific Language Impairment (SLI) (Paradis, 2010), a developmental disorder affecting language in the absence of other clinically diagnosed deficits, and whose prevalence is about 5–7% among school-age children, monolinguals as well as bilinguals (de Abreu et al., 2013).

For German, determiner use, case and gender assignment, the acquisition of subject-verb agreement (SVA), and verb placement are argued to be affected by SLI and also occur as typical errors in L2 acquisition (e.g.; Chilla, 2008; Chilla & Bonnesen, 2011; Clahsen, 1990; Clahsen et al., 1983; Clahsen et al., 2014; Haberzettl, 2003; Hamann, 2012; Hamann et al., 1998; Montanari, 2010; Paradis, 2010; Wegener, 1994).

For French SLI and L2 children, the use of accusative pronominal clitics (Grüter & Crago, 2012; Jakubowicz et al., 1998 and many others) and question formation can differ from monolingual age-matched peers (Chilla & Bonnesen, 2011; Granfeldt & Schlyter, 2004; Meisel et al., 2011; Prévost, 2009; Prévost et al., 2014; Scheidnes & Tuller, 2014). Since the differentiation of SLI-specific versus L2-specific patterns is complex, BiTD children may be misdiagnosed as having SLI, even though several studies focusing on different languages have shown that the quality and the quantity of errors differ in BiTD and monolingual

children with SLI (MoSLI) (e.g. Armon–Lotem, 2014; Meir et al., 2016; Paradis et al., 2008; Tuller et al., 2018).

Within Cost Action IS0804 (Armon–Lotem et al., 2015), new tools for cross–linguistic application were constructed and became known as the **LITMUS** tasks (**L**anguage **I**mpairment **T**esting in **M**ultilingual **S**ettings). These tools were carefully designed according to recent results provided by linguistic theory and psycholinguistic studies, with the aim of disentangling bilingual typically developing children from children with SLI. The present study brings new insights as to the appropriateness of LITMUS tasks for bilingual populations from two typologically distinct societal/majority languages by exploring performance on nonword repetition (LITMUS–QU–NWR) and sentence repetition (LITMUS–SR) tasks in two groups of bilingual children (BiTD and BiSLI) with similar L1s (Arabic, Portuguese, and Turkish) in France and Germany. We compare the performance of these children to typically developing monolingual peers in the two countries. As Marinis et al. (2017) point out, LITMUS–NWR and SR tasks have been proven to tease apart BiTD from MoSLI and from BiSLI in several countries and for several language combinations. However, before the generalized use of these tools can be recommended, the potential influence of L1 and L2 properties, language dominance, and age of onset (AoO) needs to be tested in several language pairs, since these factors have been argued to influence the performance of bilingual children on assessment tasks (Marinis et al., 2017; Thordadottir, 2015). Moreover, detailed analyses comparing different L1 groups on substructures of the LITMUS tasks are rare, with relevant variables including (i) L1 influence on the language dependent (LD) and language independent (LI) items of the LITMUS–NWR task, and (ii) transfer effects on specific syntactic constructions in the LITMUS–SR task.

The investigation of Arabic, Portuguese, and Turkish as minority languages in France and Germany is of particular interest, since these bilingual children represent a sizable

number of children in each of these countries. Moreover, these languages are typologically very different. Turkish is an agglutinative language that expresses morphosyntactic relations through suffixation. Arabic is a Semitic language, which follows root-pattern morphology. European Portuguese is a Romance language that displays, like other Romance languages, a great amount of inflectional morphology and word order variation. The L1–Portuguese–L2–French pair, for example, represents two typologically closely related languages in contrast to the L1–Portuguese–L2–German pair, which are typologically more distant. These language pairs thus allow for pertinent exploration of L1 influence on L2 development and assessment.

## **2. L1 influence in child bilingualism and consequences for the LITMUS–QU–NWR and LITMUS–SR tasks**

Minority language learners as a group consist both of simultaneous and successive bilinguals. For these two types of bilinguals, cross-language influence at the phonological, syntactic, as well as morphological levels has been reported to varying degrees. It is by now uncontroversial that simultaneous bilingual children do not show interference between one phonological system and the other, despite initial delays in establishing the phoneme systems of both their languages (Sebastián–Gallés & Bosch, 2005). Successive bilingual children, however, are more heterogeneous, since phonological working memory and the phoneme system of the L1 have been argued to develop during the first year of life (Bosch et al., 2000; Pierce et al., 2017; Werker & Tees, 1983). Age of Onset (AoO) and Length of Exposure to the L2 (LoE) have often been discussed as having an influence on the L2 development and performance in L2 tasks. Thus, the nonword repetition tasks for bilingual children as a group have to be carefully constructed in order to avoid a disadvantage of bilingual children with diverse input and language backgrounds.

The typological differences or similarities between the L1 and the L2, the Age of Onset, language dominance, and the Length of Exposure seem to be crucial for predicted transfer from one grammar to the other. Cross-linguistic morphosyntactic influence has been observed for simultaneous bilinguals (Müller & Hulk, 2001), and has been argued to be an important factor for the interpretation of error patterns in sentence repetition tasks for bilinguals. In studies of successive bilinguals, transfer of L1 properties has been reported for L2 acquisition of verb-placement with respect to negation in Turkish learners of English (Haznedar, 1997), scrambling in Dutch (Unsworth, 2005) and double-object constructions in English (Whong-Barr & Schwartz, 2002). For German, the basic OV order, subject-verb agreement, and the assignment of case and gender have equally been argued to be sensitive to L1 transfer (Haberzettl, 2003; Wegener, 1994), whereas difficulties with clitic placement and with question formation are reported for L2 French (Hamann & Belletti, 2006; Prévost et al., 2014).<sup>1</sup>

Transfer models adopted for child L2 acquisition are often related to what has been proposed for adult L2 acquisition, e.g. in terms of early availability or not of functional categories and features, such as the feature assembly account (Lardiere, 2009) or the Interpretability Hypothesis (Tsimplici & Dimitrakopoulou, 2007). These are of interest here since they predict that features of functional categories, and uninterpretable features (such as case, number agreement on V, or grammatical gender) in particular, may not be assembled in a target-like manner. It is the uninterpretable features which might be impaired in children with specific language impairment and second language acquisition, leading to overlap effects between bilingual children and children with SLI.

---

<sup>1</sup>Other studies have shown, however, that basic word order parameters are not subject to transfer or are acquired very fast, see Belletti & Hamann (2004) for lack of transfer of V2 (verb second placement) to French in a German early child learner of French, or Blom (2008) for early mastery of V2 and V-final order in Dutch by children of Moroccan or Turkish origin.

With respect to a possible identity between bilingual development patterns in children and adult L2, it has to be considered that, for early L2 learners, the L1 may still be incomplete when L2 development begins. It follows that L1 features that are acquired late could not be responsible for some L2 errors. This is especially true for children with an early AoO of the L2. An example is the prevalence of article omission in L2 English in Arabic, Spanish, and Romanian L1 children, even though these L1s have article systems similar to English (Zdorenko & Paradis, 2011).

A further crucial factor for the assessment and prediction of bilingual influence is language dominance, since structures from the dominant language might be incorporated into the weaker language (Paradis & Genesee, 1996; Yip & Matthews, 2000). This scenario is more frequent than transfer from the non-dominant to the dominant language. Specifically, studies have suggested that bilingual children may experience developmental difficulties in the domain of morphosyntax in their non-dominant language, again depending on the amount and type of exposure to the target language. As previous studies indicate, we might expect effects at the level of morphological spell-out (Prévost & White, 2000) and in feature assembly of functional categories (Lardiere et al. 2009), specifically for interface and uninterpretable features, i.e. grammatical features lacking semantic content (Tsimpili & Dimitrakopoulou, 2007), but also for features that influence word order properties such as verb placement.

It is furthermore challenging to disentangle the effects of L1 influence from the effects of reduced L2 input (Scontras et al., 2015). However, detailed analyses of the impact of language dominance on the French and German LITMUS tasks have shown that this variable did not compromise the diagnostic accuracy of the tasks (cf. Tuller et al., 2018; Abed Ibrahim & Fekete, 2019).



### **3. The LITMUS nonword repetition and sentence repetition tasks**

Sentence repetition and nonword repetition tasks (SR, NWR) are widely recognized as tools for the identification of specific language impairment in monolingual and bilingual children (Conti–Ramsden et al., 2001; Klem et al., 2015; Vinther, 2002).

NWR tasks are known to measure (phonological) working memory (WM) capacities (Archibald & Gathercole, 2006), and storage, but, and this is crucial, they can also provide information about underlying phonological representations (Marshall & van der Lely, 2009). Several researchers have argued that both WM and phonological complexity are sources of difficulties for children with SLI (dos Santos & Ferré, 2018; Ferré et al., 2012; Gallon et al., 2007). Monolinguals outperform bilinguals in NWR tasks when they contain L2–like nonwords (Kohnert et al., 2006), and, in such tasks, typically developing bilinguals (BiTD) might pattern with monolingual children with SLI (Schöler, 2006; Windsor et al., 2010). At the same time, the performance of bilinguals is better when tested in their L1 (Gutiérrez–Clellen & Simon–Cereijido, 2010; Summers et al., 2010). QU–NWRs with quasi-universal nonwords do not disadvantage bilinguals when compared to monolinguals (Boerma et al., 2015). Since the primary aim of the LITMUS–QU–NWR is to assess the phonological competence of bilingual children without disadvantaging them, the task under discussion here does not contain L2–like nonwords in order to limit the effects of prior linguistic knowledge of the L2 (Chiat, 2015; dos Santos & Ferré, 2018; Grimm & Hübner, to appear). An advantage of such an NWR over other expressive or receptive tasks is that it does not rely on prior knowledge of vocabulary and morphosyntax. NWR tasks thus seem to be promising clinical assessment tools for the identification of SLI in bilingual contexts (e.g., Klem et al. 2015; Hamann & Abed Ibrahim, 2017; dos Santos & Ferré, 2018; Tuller et al., 2018, Abed Ibrahim & Fekete, 2019).

Sentence repetition tasks are easy to use in clinical settings and have been shown to assess underlying grammatical representations (Polišenská & Kalpakova, 2014) as well as

language processing (Archibald & Gathercole, 2006). More recently, they have been used in bilingual contexts for distinguishing bilingual children with and without SLI (de Almeida et al., 2017; Hamann & Abed Ibrahim, 2017; Meir et al., 2016). SR tasks have been argued to be more reliable than other language-dependent expressive and receptive language tasks, such as (for English) third-person singular or past tense tasks for the assessment of SLI (Conti-Ramsden et al., 2001; Stothard et al., 1998).

### **3.1. Nonword repetition and contrastive linguistic features relevant for the LITMUS–QU–NWR**

The LITMUS–QU–NWRs for French and German (dos Santos & Ferré, 2018; Grimm & Hübner, to appear) were explicitly designed to avoid the effects of L2 lexical and morphosyntactic competence and those L2 phonological properties that could penalize bilingual children (Chiat & Polišenská, 2016; En et al., 2014). These NWRs were constructed using vowels and consonants attested in most of the world’s languages; they contain the same 30 language independent (LI) nonwords in addition to language dependent (LD) nonwords for French and for German, respectively (for a detailed presentation of the items and discussion see dos Santos & Ferré, 2018; Grimm & Hübner, to appear). The maximal nonword length is limited to three syllables in order to limit working memory effects. LI test items consist of cross-linguistically frequent phonemes and conform to language-independent phonotactic properties. Unlike the cross-linguistic–NWR (CL-NWR) discussed in Chiat & Polišenská (2016), the nonwords in the French and German LITMUS–QU–NWRs contain “C(onsonant)V(owel)” syllables, as well as more complex syllables with branching onsets (“#CCV”) and syllables with consonantal codas (“CVC#”) (cf. Appendix A). The latter syllable types are expected to pose difficulties for language-impaired children, due to their complexity. At the same time, they should not challenge bilingual typically developing

children, due to their cross-linguistic frequency and the presumably intact phonological loop in typical bilinguals. The language-dependent items were built using the same vowels and consonants as for the language independent nonwords (/a, i, u/ and /p, k, f, l/), adding the coronal fricatives /s/ (French and German) and /ʃ/ (German only), which gave rise to syllables of the types #sCV, #sCCV, sC#, and Cs#, plus nonwords with an internal coda in the LD part. These syllable types render the LD items more complex than the LI ones. Contrary to expectations, dos Santos and Ferré (2018) found that the more complex syllables in the LD items, which are characteristic of French, but not frequent in other languages, did not have a negative effect. Likewise, the German and French LD items more adequately separated typically developing children from children with SLI than the LI items, in monolingual and bilingual populations (Grimm & Hübner, to appear; dos Santos & Ferré, 2018; Abed Ibrahim & Fekete, 2019). However, other studies have argued for a bilingual advantage in the phonological loop (Pierce et al., 2017) so that we might even expect that bilingual children with SLI perform slightly better than monolinguals with SLI on both parts of the task.

Although the French and the German tests have been shown to reliably differentiate bilingual children with and without SLI (de Almeida et al., 2017; dos Santos & Ferré, 2018; Grimm & Hübner, to appear; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018, Abed Ibrahim & Fekete, 2019), systematic cross-linguistic comparison of different L1 groups is lacking for these NWRs. Such a comparison is crucial given that several studies have argued for transfer effects in the phonological acquisition of bilingual children (i.e. Barlow, 2014; Fabiano-Smith & Barlow, 2010; Paradis, 2001).

### **3.2 Sentence repetition and contrastive linguistic features relevant for LITMUS-SR**

Assessment with SR tasks has been shown to be either sensitive (Schöler, 2003, for German) or insensitive to the overlap of child L2 and monolingual SLI (Paradis et al., 2013, for English; Thordardottir & Brandecker, 2013, for French). Accordingly, the French and

German LITMUS–SR tasks were carefully designed to sidestep disadvantages for bilingual children by avoiding, as far as possible, grammatical features that are known to be specifically difficult for children acquiring French or German as L2. The French SR, for example, does not include accusative pronoun clitics, and the German SR does not explicitly test for subject-verb agreement (SVA) or case.

Each LITMUS–SR contains a core set of structures, and additional language specific constructions (cf. Appendix B and Tuller et al., 2018, for detailed description). Among the structures incorporated into the two tasks are simple sentences with canonical word order for French and German, Object–Constituent–questions involving movement of the (object) interrogative pronoun to the front (left periphery) of the clause, complement clauses (both infinitival and finite, the latter arguably more complex), and relative clauses, constructions involving embedding and movement (both subject and object, the latter arguably more complex).

Although both LITMUS–SR tasks are capable of disentangling BiSLI from BiTD children (de Almeida et al., 2017; Fleckstein et al., 2016; Hamann & Abed Ibrahim, 2017). certain aspects of these tasks may constitute possible loci of L1 transfer due to the typology of the languages involved (Arabic, Portuguese, or Turkish) and L2 interference (French or German) (see Table 1). Note that we list properties for which at least one of the languages is different from the others.

Table 1: LITMUS–SR: Relevant word orders, morphosyntactic features, and expected loci of detectable interference

	<b>Arabic</b>	<b>Portuguese</b>	<b>Turkish</b>	<b>French</b>	<b>German</b>
Word Order	VSO (SVO alt.)	SVO (VSO alt.)	SOV (SVO alt.)	SVO	V2 / OV
Pro-drop	yes	yes	yes	no	no
Wh-movement	left periphery	left periphery	in situ / immed. left of V	left periphery/ in situ	left periphery
Subordinate Clauses	verbal	verbal	nominal (more common)/ verbal	verbal	verbal
Definite D	yes	yes	no	yes	yes
Gender	yes	yes	no	yes	yes
Gender on D	no	yes	no	yes	yes
Number on V	(yes)	yes	yes (optional with overt 3.ps. subject)	yes	yes
Aux	yes	yes	no	yes	yes
C	yes	yes	(yes)	yes	yes
Case	no/(yes)	no	yes	no	yes
Case on D	no	no	no	no	yes

Note: V = Verb; O=Object ; S= Subject ; V2 = Verb in second position; D = Determiner; Aux = Auxiliary

If there are L1 influences on LITMUS–SR performance, they should become obvious in those aspects where the L2s French and German align, but differ from at least one L1 (cf. Table 1). First of all, the L1s are pro-drop languages, in contrast to French and German. The chosen language pairs will allow us to investigate the possible influence of word order, that is VO/OV and SV/VS differences, since Turkish and German are underlying OV, whereas the other languages are not. Second, auxiliaries or determiners do not exist in Turkish, but are overt in French and German, which should result in difficulties with auxiliaries and determiners in the Turkish L1 groups. The omission of L2 articles has been shown to be associated with the realization of definiteness and specificity in the L1 (at least at some

stages), in the sense that L1s without articles may lead children to make more omissions in the L2 in comparison to children whose L1 encodes specificity and definiteness with articles (e.g. Zdorenko & Paradis, 2008; see also Schönenberger, 2011 for adult L2).

The overview in Table 1 furthermore allows us to expect errors in the use of gender on the determiner and in case. Arabic, Portuguese, French, and German have determiner systems, while Turkish does not. Cross-linguistic influence, however, may surface in the grammatical features spelled out by the determiner (D).

Arabic, for example, marks definiteness only, in contrast to German and French, which both mark gender. For L1 Portuguese children, we would postulate difficulties with the use of gender (i.e. two versus three grammatical genders) and with case marked object relatives in German. For relative clauses, higher error rates for German than for French are possible since the French object relative pronoun *que* is not marked for gender, contrary to German (*den, die, das*). We would thus expect to see L1 influence in German but not necessarily in French. Turkish differs from all of the other languages in several characteristics, predicting that Turkish children should show errors in the production of auxiliaries and determiners, and on gender assignment, in both the German and the French SRs, if performance is L1 driven. Finally, the German case system is known to be particularly difficult for all L2 children (including children of Turkish origin, see Schönenberger et al., 2012), and constitutes a difference between the German and French SRs that cannot be easily removed, since case marking is crucial for the realization of the targeted constructions in German, as in sentences with object relatives and topicalization. Given these properties of German, we might expect that, generally, typically developing bilinguals in Germany are disadvantaged in their performance on L2 SR when compared to the SR performance of bilinguals in French, a language, which does not have systematic case marking.

#### **4. Research questions and aims of the study**

The LITMUS-sentence repetition and the LITMUS–nonword–repetition tasks aim to differentiate children with SLI from BiTD children. The current study adds a more detailed view on the applicability of these tasks for diverse language minority groups. Considering the different properties and features of the three L1s and the two L2s of the children in our study, we ask the following research questions:

(1) Do typically developing bilingual children (BiTD) with typologically different L1s differ from each other and from monolingual typically developing children (MoTD) and bilingual children with SLI (BiSLI) on the LITMUS–QU–NWR and LITMUS–SR for German and French?

(2) Given that the LITMUS–QU–NWR contains language dependent and language independent items, do children with typologically different L1s perform differently on these items?

(3) Are differences in performance on LITMUS–SR systematically driven by L1 influence?

(4) Given that the LITMUS–SR necessarily has language specific features for either German or French as an L2, can the performance of bilingual children with the same L1 be explained by the inherent structures of the L2?

#### **5. Participants and procedures**

The French and the German versions of the LITMUS–QU–NWR and SR were administered as per description (cf. Hamann & Abed Ibrahim, 2017; Tuller et al., 2018) and in line with the COST IS0804 recommendations (cf. Armon–Lotem et al., 2015). We investigated 117 typically developing bilingual children and 34 bilingual children with SLI with Arabic, Portuguese or Turkish as their home language. These data are compared to 47 MoTD children (cf. Table 2).

Participants were recruited in several institutions, e.g. kindergartens, community associations and speech-language therapy centres in different parts of Germany and France, representing a wide spectrum of the L1 communities in question. The age range of our participants (5;6 to 9;0) covers the last year of kindergarten and the crucial first 2–3 years of primary school that are relevant for the assessment of SLI in both countries.

The assignment of the label “SLI” to bilingual children followed international recommendation (i.e. Leonard, 2014; Thordadottir, 2015) in combination with a comprehensive assessment procedure, including standardized tests of the L1 and L2. Relevant background information was gathered with the Questionnaire for Parents of Bilingual Children (PaBiQ, Tuller, 2015). Based on the PaBiQ, a language dominance index (LDI) is calculated by subtracting the L1–Exposure\_Index from that of L2 (for detailed description cf. Tuller, 2015; Tuller et al., 2018; Abed Ibrahim & Fekete, 2019). After comparing multiple LDI cut–offs around LDI=0 (optimal balanced bilingual) against impressions of bilingual investigators of the children’s spontaneous speech samples in both of their languages (see also Almeida et al., 2017), an LDI between –5 and +5 was set as a cut–off separating dominant from balanced bilinguals. Children with LDIs ranging from –5 to +5 are classified as ‘balanced’, children whose LDI is below –5 are considered to be L1-dominant, while children with an LDI above +5 are classified as dominant in the L2. All groups were comparable with respect to socio-economic status (SES) as measured with the PaBiQ which includes questions providing information about the family’s SES based on the mother’s and the father’s educational levels. Only children who scored above percentile rank 9 (IQ score  $\geq$  80 according to Wechsler’s IQ scale) were included in the study. Children who were unable to complete even receptive subtests in the L1 were also excluded, due to lack of functional bilingualism. All children had a minimum exposure to the L2 of 12 months.



Table 2: Participants in France and in Germany: MoTD, Bi-TD and Bi-SLI

	Germany							France						
	MoTD	BiTD			BiSLI			MoTD	BiTD			BiSLI		
Total N	10	48			8			37	69			26		
			LoE	AoO		LoE	AoO			LoE	AoO		LoE	AoO
L1 Arabic		11	58.45 (23.30)	35.45 (17.47)	3	63.4 (21.7)	19.5 (16.5)		27	60.93 (22.35)	20.19 (26.30)	9	68.4 (19.5)	15.3 (25.5)
L1 Portuguese		20	66.45 (24.00)	17.65 (23.80)	1				18	63.83 (24.71)	19.50 (23.29)	8		
L1 Turkish		17	58.12 (18.52)	23.82 (16.24)	4				24	76.96 (19.32)	10.83 (15.66)	9		
<b>Age (mo.)</b>														
M	75.0	83.8			81.0			84.3	84.0			83.7		
(SD)	(9.0)	(13.3)			(16.7)			(10.7)	(12.5)			(13.3)		
Range	66–92	66–108			64–108			67–101	64–106			66–107		
<b>RPM</b>														
M	81.2	64.48			59.1			52.2	40.61			28.8		
(SD)	(14)	(25.6)			(26.6)			(24.2)	(25.8)			(20.6)		
<b>SES<sup>2</sup></b>														
M	n.a. <sup>3</sup>	13.3			12.5			n.a.	10.73			9.7		
(SD)		(3.99)			(2.97)				(3.88)			(2.65)		
<b>Gender</b>														
	8 F	26 F			2 F			14 F	28 F			8 F		
	2 M	22 M			6 M			23 M	41 M			18 M		

<sup>2</sup> Maternal education (years of mother's education) is used as a metric for SES in this table.

<sup>3</sup> SES data are only available for bilingual children.

## 6. Results

We first compare the general characteristics of the bilingual children in the two countries, in terms of non–language variables (age and gender) and bilingualism measures (AoO and LoE to the L2, early L1 and L2 exposure, current L1 and L2 richness). Second, we focus on quantitative QU–NWR and SR results on the performance comparing BiTD, MoTD and BiSLI children in both countries. Since both tasks reliably disentangle BiSLI and BiTD children, and the BiSLI L1 subgroups were too small for statistical analyses of L1 influence, further error analyses on the SR compare BiTD and MoTD groups, only. IBM SPSS 24 (2016) was used for all statistical analyses. We used *t*–tests for background comparisons. Due to unequal group sizes and since explorative statistics revealed a violation of the assumption of normality in our data set, non–parametric statistical tests were used for further group comparisons. The *r* Statistic is used to report on effect sizes as follows (Field, 2018):

$r = 0.10 - < 0.30$  ( small effect),  $r = 0.30 - < 0.50$  (medium effect),  $r = \geq 0.50$  (large effect).

### 6.1 Background comparisons

Comparing typically developing bilingual children in Germany with BiTD children in France for non–linguistic variables, no significant differences were found regarding age ( $t(115) = 0.532, p = 0.596$ ); however, a significant difference was found between the bilingual groups in the two countries concerning SES as measured by years of maternal education ( $t(112) = 3.4111, p = 0.001$ ). Considering linguistic variables, such as early and current exposure to the L2 and current use of the L1 and the L2 (Tuller 2015), the BiTD children in France had a higher L1 richness (current use outside of the immediate family) score ( $t(115) = -5.068, p < 0.001$ ) than the children in Germany, but no reliable significant difference was observed between the children in the two countries with respect to L2 richness ( $t(115) = 1.921, p = 0.057$ ). More importantly, no significant contrasts were found between the two groups concerning the Age of Onset of systematic exposure to the L2 ( $t(157) = 1.752, p = 0.082$ ) or

the Length of Exposure to the L2 ( $t(115) = -1.023, p = 0.308$ ). The bilingual children with specific language impairment in the two countries did not differ significantly for gender, age, AoO, LoE, or on their linguistic richness scores in the L1 or the L2, and the MoTD children in Germany were younger ( $t(45) = -2.264, p < 0.05$ ) than the MoTD children in France.

Concerning the typically developing children within each country, the MoTDs were significantly younger than their typically developing bilingual peers in Germany ( $t(56) = -2.084, p < 0.05$ ) but not in France ( $t(104) = -0.072, p = 0.943$ ). No significant differences emerged between the BiTD and BiSLI groups in Germany, except for early exposure to L2, where BiSLI children had significantly more early exposure to German than the BiTDs [ $\chi^2(1, N = 56) = 6.796, p < 0.05$ ]. Comparing BiTDs to BiSLIs in France only revealed a significant difference concerning linguistic variables, namely Linguistic Richness L1 in favor of the BiTD group ( $t(93) = 3.151, p < 0.05$ ).

## **6.2 Quantitative Comparisons on the QU–NWR and the SR results**

### **6.2.1 Comparison of typically developing children and children with SLI on the QU–NWR**

The QU–NWR task was scored by whole item accuracy, i.e. counting the whole response to a test item as correct (=1) or incorrect (=0) following the procedures as described in Tuller et al. (2018) and Hamann & Abed Ibrahim (2017). In order not to disadvantage bilingual children whose L1 lacks certain voicing contrasts, such as /p, b/ in Arabic, voicing substitutions were disregarded. We present whole item accuracy over the test, and performance on the language dependent (LD) and language independent (LI) items.

We first ran Omnibus Kruskal–Wallis tests using scores on NWR\_correct, NWR\_LI and NWR\_LD as dependent variables to detect an effect of clinical status. All measures yielded significant results for QU–NWR for both German (GE) and French (FR) data ( $p <$

0.001; cf. Appendix C). Between-group pairwise comparisons (Mann-Whitney-U test) were then carried out on the dependent measures applying Bonferroni-correction of  $p$ -values. Typical bilinguals and monolinguals in both countries performed significantly better than the bilingual children with SLI ( $p < 0.001$ ; cf. Appendix D).

In order to evaluate whether bilingualism per se disadvantaged children in performance on QU-NWR, we compared the MoTD and BiTD groups. No significant differences were found between these two groups of children on either the language dependent or the language independent items of the task, in either of the two languages.

Figure 1: Performance of MoTD ( $n = 37$ ), BiTD ( $n = 48$ ) split by L1 (Arabic=Bi-TD-A, Portuguese=Bi-TD-P, and Turkish=Bi-TD-T), and BiSLI children ( $n = 8$ ) in the German LITMUS-QU-NWR

□ NWR\_% correct   ▨ NWR\_LI-% correct   ▩ NWR\_LD-% correct

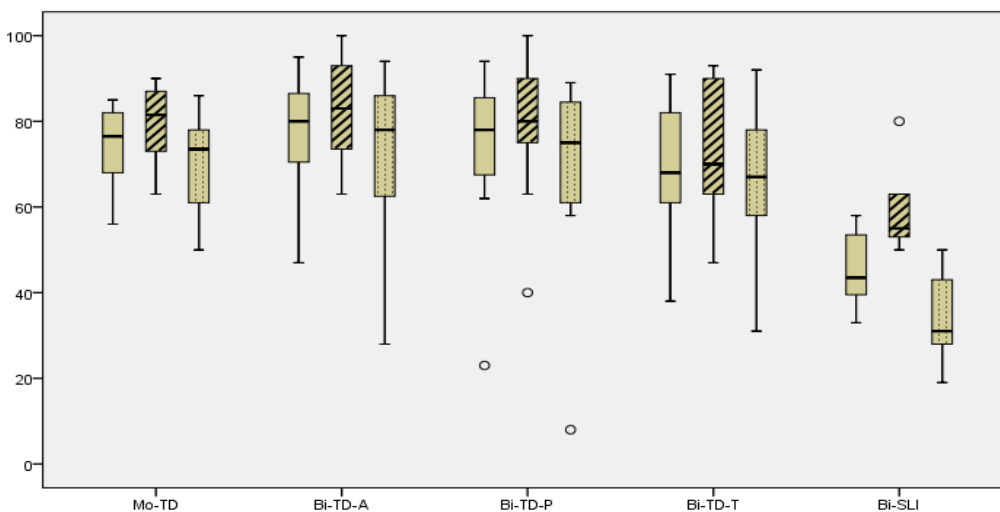
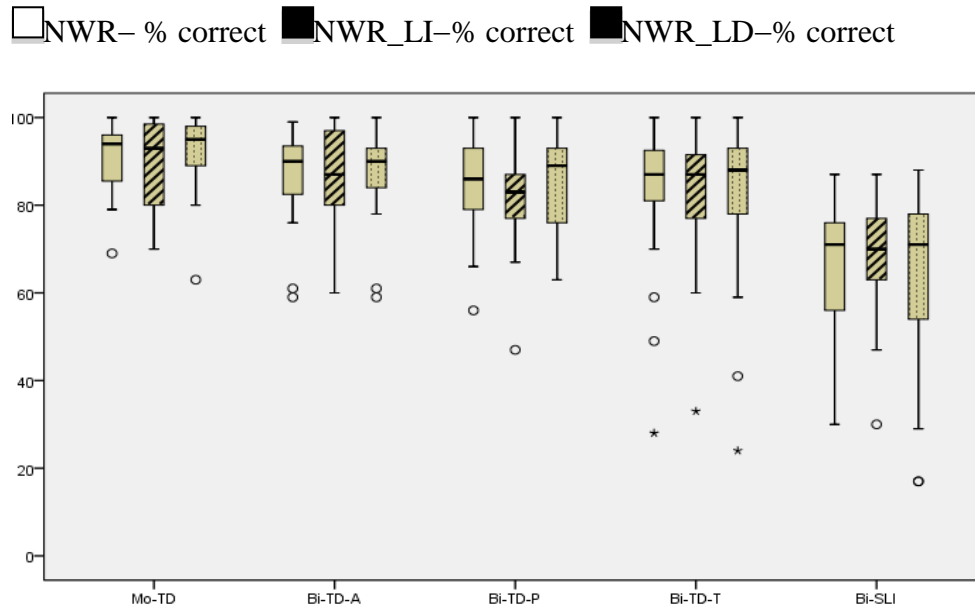


Figure 2: Performance of MoTD ( $n = 37$ ), BiTD ( $n = 69$ ) split by L1 (Arabic=Bi-TD-A, Portuguese=Bi-TD-P, and Turkish=Bi-TD-T), and BiSLI children ( $n = 26$ ) in the French LITMUS-QU-NWR



As can be seen in Figures 1 & 2, each of the three language groups in each of the countries performed equally well. The L1 subgroups did not significantly differ from the monolingual typically developing group or from each other in the QU-NWR. It is important to state that errors were too infrequent to allow for meaningful comparison of errors or of L1 influence on performance in the LITMUS-QU-NWR between groups. We also controlled for language dominance (see section 5). Notably, L1 dominant children did not perform worse than balanced or L2-dominant children in the QU-NWR. Partial correlations on the BiTD groups in Germany and France controlled for age at testing showed no significant correlation with L2 language dominance, except for a weak correlation on the NWR\_LI measure in Germany ( $r = 0.299, p < 0.05$ ).

## 6.2.2 Comparison of typically developing children and children with SLI on the LITMUS–SR

For the SR, all responses were recorded, transcribed, verified and analyzed with different scoring methods. Two different rates are reported here: the percentage of correct identical repetitions (SR\_Id) and the percentage of correctly repeated target structures (SR\_Tar). SR\_Id requires verbatim repetition of the stimulus, disregarding only phonological errors. SR\_Tar represents whether the sentence preserves the syntactic construction targeted, and thus disregards repetition errors that do not affect the structure of the stimulus sentences. For this measure, lexical substitutions and gender errors were disregarded in both French and German. Systematic case errors were overlooked in German if they did not obliterate the characteristics of the structure as in topicalization or object relative clauses.

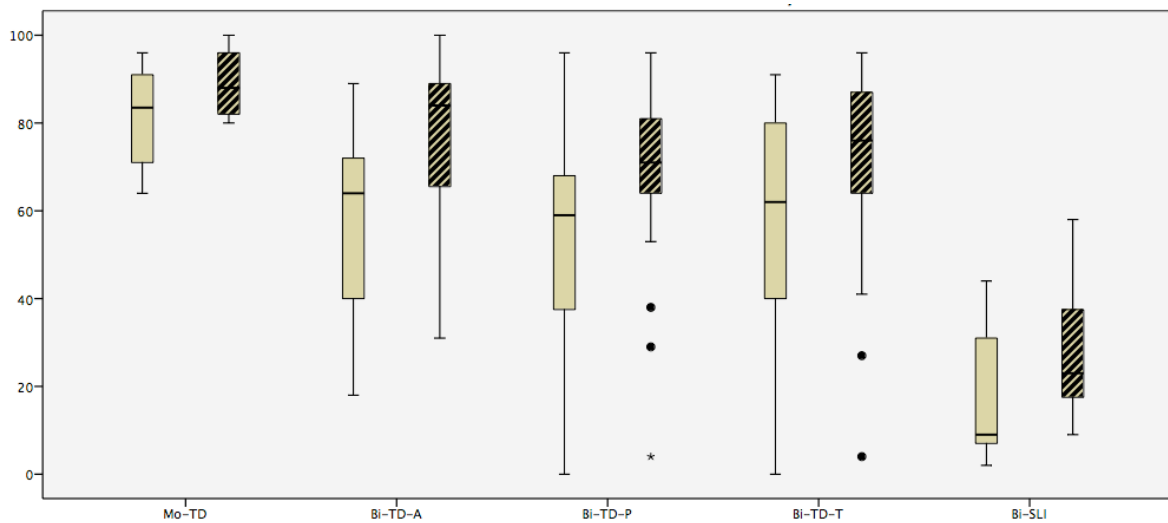
Since language dominance has been argued for to influence performance on LITMUS tasks, we ran partial correlations that revealed a moderate correlation with language dominance, when age is controlled. Dominance had an impact on the performance of bilingual children in France and Germany for both identical repetition (FR:  $r = 0.462, p < 0.001$ ; GER:  $r = 0.440, p < 0.05$ ) and target structure (FR:  $r = 0.445, p < 0.001$ ; GER:  $r = 0.412, p < 0.05$ ). In Germany, L2 dominant children outperformed L1 dominant children on the German version of the LITMUS–SR (SR\_Id [ $\chi^2(2, N = 46) = 10.692, p < 0.05$ ], SR\_Tar [ $\chi^2(2, N = 46) = 9.531, p < 0.05$ ]). The same holds for France, where pairwise comparisons reveal significant differences between the L1 and the L2 dominant children on the identical repetition measure ( $\chi^2(2, N = 79) = 12.932, p < 0.05$ ) as well as on the “target structure” measure ( $\chi^2(2, N = 79) = 15.653, p < 0.001$ ).

## LITMUS-SR in L2 German

Omnibus Kruskal-Wallis tests were undertaken based on scores for the percentage of correct identical repetitions (SR\_Id) and the percentage of correctly repeated target structures (SR\_Tar). SR\_Id and SR\_Tar as dependent variables showed a positive effect of clinical status ( $p < 0.001$ ; cf. Appendix C and F). The overall performance of the different groups in Germany is given in Figure 3.

Figure 3: Performance of MoTD ( $n = 10$ ), BiTD ( $n = 48$ ) split by L1 (Arabic=Bi-TD-A, Portuguese=Bi-TD-P, and Turkish=Bi-TD-T), and BiSLI children ( $n = 8$ ) on the German SR

□ SR\_Id- % correct    ■ SR\_Tar-% correct



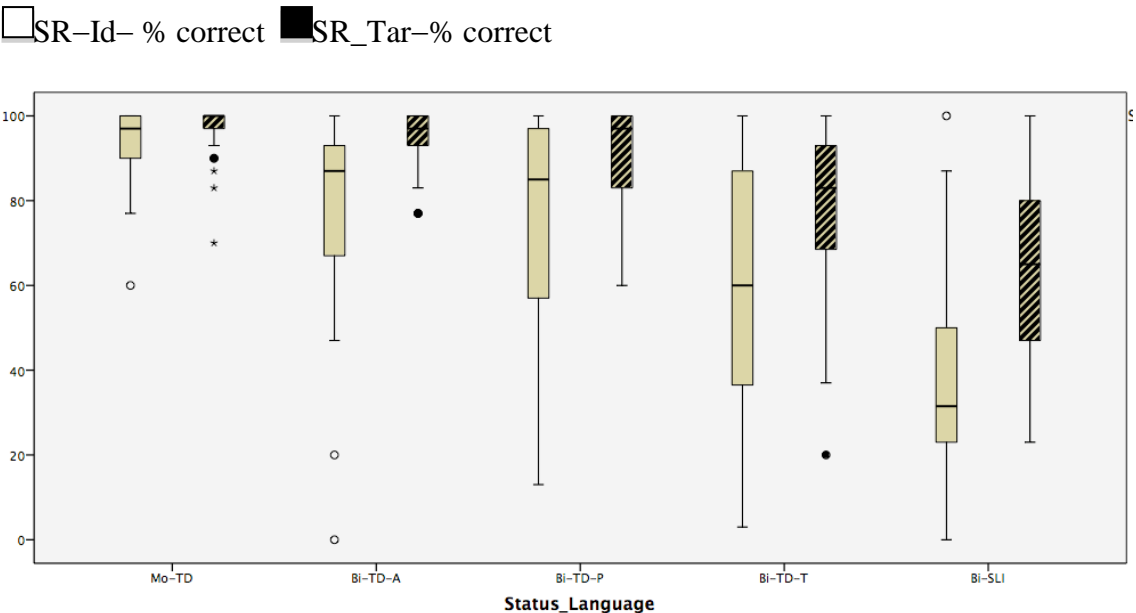
Pairwise comparisons using Mann-Whitney- $U$  test applying Bonferroni-correction showed that the German SR can reliably distinguish between the BiTD and BiSLI groups on both SR measures ( $p < 0.001$ ; cf. Appendix D and F). However, significant differences with medium effect sizes were found between the MoTD and the three BiTD subgroups (for identical repetition (SR\_Id):  $U = 111$ ,  $p < 0.05$ ,  $r = 0.416$ ; and for target structure (SR\_Tar):  $U = 86$ ,  $p < 0.05$ ,  $r = 0.469$ ), where MoTD children outperformed the bilingual groups.

Detailed analyses revealed that only one language subgroup in Germany, namely the typically developing Portuguese L1 children (BiTD-P,  $n = 20$ ), significantly differed from MoTDs, with large effect sizes, in both SR\_Id ( $U = 28.5, p < 0.05, r = 0.575$ ) and SR\_Tar ( $U = 31, p < 0.05, r = 0.555$ ). However, no significant differences were found between the performance of the three BiTD groups with L1 Turkish, Portuguese or Arabic on the German LITMUS-SR.

**LITMUS-SR in L2 French**

The Omnibus Kruskal-Wallis test revealed positive results for the effect of SLI for the French data as well ( $p < 0.001$ ; cf. Appendix C and F). Post-Hoc Mann-Whitney-U test pairwise comparisons applying Bonferroni adjustments showed significant differences with medium to large effect sizes between BiTD and BiSLI children ( $p < 0.001$ ; cf. Appendix D and F). The overall performance of the different groups in France is given in Figure 4.

Figure 4: Performance of MoTD ( $n = 37$ ), BiTD ( $n = 69$ ) split by L1 (Arabic=Bi-TD-A, Portuguese=Bi-TD-P, and Turkish=Bi-TD-T), and BiSLI children ( $n = 26$ ) on the French SR





In France, significant differences with medium effect sizes were found between the MoTD, who performed better, and the BiTD children on both measures (identical repetition (SR\_Id):  $U = 578, p < 0.001, r = 0.452$ ; target structure (SR\_Tar):  $U = 766, p < 0.001, r = 0.341$ ).

The BiTD children in France performed more heterogeneously and with a broader variance than the monolingual children, so that measures revealed statistically different performance in the results for identical repetition (MoTD/BiTD–A:  $U = 281, p < 0.05, r = 0.357$ ; MoTD/BiTD–P:  $U = 184.5, p < 0.05, r = 0.365$ ; MoTD/BiTD–T:  $U = 112.5, p < 0.001, r = 0.632$ ). However, for target structure, the Turkish L1 children ( $n = 24$ ) performed lower and thus diverged significantly from the MoTDs ( $U = 137.5, p < 0.001, r = 0.602$ ). Further comparisons between the three Bi–TD groups revealed significant differences with medium effect sizes only between the BiTD Arabic ( $n = 27$ ) and the Turkish children ( $n = 24$ ), since the Arabic L1 children produced fewer errors than the Turkish group (BiTD–A/BiTD–T, SR\_Id:  $U = 184.5, p < 0.05, r = 0.369$ ; SR\_Tar:  $U = 149.0, p < 0.05, r = 0.470$ ).

### **6.2.3 Error analyses: SR**

In a first step, we compared BiSLI and BiTD children on all substructures included in the SRs. As Table 3 reveals, only a few of these structures covered by the French and German LITMUS–SR were difficult for both BiTDs and BiSLIs in both countries ( $p > 0.05$ ; cf. Appendix E and G), including structures such as topicalization (identical repetition and target structure), case on the determiner and auxiliaries. Likewise, BiTD and BiSLI children in France and Germany encountered difficulties with gender marking on determiners (D).

Since the BiSLI L1–subgroups are too small and the different error types were too idiosyncratic to allow for systematic L1 comparisons, the following analyses concentrate on L1 influence on SR substructures and specific grammatical features, comparing MoTD and BiTD children, only. Considering the two L2s, we first analysed those structures and features

that are similar in German and French but are different in one of the three different L1s under consideration. We ran Omnibus Kruskal–Wallis tests using scores on the dependent variables gender on determiners (Gen\_D), omission of the determiner (D\_Om), omission of auxiliaries (Aux\_Om) and C errors. Between–group pairwise comparisons (Mann–Whitney–U test, with Bonferroni correction) were carried out on the dependent measures. Second, all structures that were part of the French and German SR were statistically analysed for both scorings, that is “identical repetition” (SR\_Id) and “target structure” (SR\_Tar). For functional categories such as auxiliaries, complementizers, and determiners as well as case and gender marking features on determiners, error rates were calculated as percentage of erroneous responses out of all possible contexts produced by a child.

In Germany, no significant differences between MoTD ( $n = 10$ ) and BiTD groups split by L1 ( $n = 48$ ) were found for gender on the determiner (Gen\_D), omissions of determiners (D\_Om), auxiliary omissions (Aux\_Om) or C errors. Between–group comparisons for the bilingual children revealed no significant differences between the Turkish, Arabic or Portuguese L1 TD–children in their performance on the German version of the LITMUS–SR.

Comparing the results of the French BiTD ( $n = 69$ ) and MoTD ( $n = 37$ ) children on the LITMUS–SR, we found significant differences for all four error types, omission of determiner ( $\chi^2(3, N=106) = 8.940, p < 0.05$ ), gender on determiner ( $\chi^2(3, N=106) = 69.225, p < 0.001$ ), auxiliary omission ( $\chi^2(3, N=106) = 19.116, p < 0.001$ ), and C errors ( $\chi^2(3, N=106) = 27.765, p < 0.001$ ). French MoTD children outperformed BiTD children, but further analyses on these four aspects revealed no significant differences for Arabic– or Portuguese–speaking children when compared to monolingual typically developing children. The Turkish L1 group, however, performed significantly lower on all measures, except for omissions of the determiner, with large effect sizes (MoTD/BiTD–T: gender on determiners:  $U = 55.5, p < 0.001, r = 0.867$ ; auxiliary omission:  $U = 223.5, p < 0.001, r = 0.554$ ; C\_Err:

$U = 175, p < 0.001, r = 0.659$ ). When the bilingual typically developing groups split by L1 were compared for error types, no significant differences were found between Arabic- and Portuguese-speaking children, but the Turkish L1 group ( $n = 24$ ) had significantly more gender errors on determiners than the Arabic L1 group ( $n = 27$ ) ( $U = 81, p < 0.001, r = 0.696$ ), and the Portuguese L1 group ( $n = 18$ ) ( $U = 34, p < 0.001, r = 0.757$ ).

#### **6.2.4 Analysis on specific features of the German and French LITMUS-SR**

Further structure-by-structure analyses of the performance of MoTD and BiTD children shed light on substructures and grammatical features susceptible to L1 influence and revealed important insights into the composition of the German and the French SR's.

Significant differences between MoTD and BiTD children (split by L1) in the German version of the LITMUS-SR were found with respect to identical repetition of object who-questions [ $\chi^2(3, N=58) = 13.427, p < 0.05$ ], identical repetition of finite complement clauses [ $\chi^2(3, N=58) = 8.356, p < 0.05$ ] and identical repetition of subject relatives [ $\chi^2(3, N=58) = 9.636, p < 0.05$ ]. However, these significant differences were no longer visible when comparisons were made on correct target structure.

Further analyses revealed that it is the Arabic-speaking children who showed significant difficulties in the identical repetition of subject relatives ( $U = 15, p < 0.05, r = 0.657$ ) and object wh-questions ( $U = 7.4, p < 0.001, r = 0.784$ ). Within-group comparisons, however, showed no differences between the L1 Turkish, Portuguese or Arabic groups. These findings somewhat contradict the above-mentioned results from the BiTD-P children ( $n = 20$ ) in Germany, which displayed strong effect sizes on identical repetition, with significantly lower performance than the Mo-TD group. As Figure 3 illustrates, however, there were three BiTD-P children who might have lowered the overall scores of the BiTD-P group (for a detailed discussion of these participants cf. Hamann & Abed Ibrahim, 2017). It is also

important to note that, the correct production of case on D, a specific feature of German, was difficult for all bilingual groups. Pairwise comparisons revealed no significant differences between MoTD and bilingual children with Turkish L1 or bilingual children with Arabic L1. Case on the determiner was even more difficult for Portuguese children ( $U = 37, p < 0.05, r = 0.509$ ).

For the structures covered by the French LITMUS–SR, significant differences between MoTD and BiTD children split by L1 were documented in most of the structures under discussion, except for the identical repetition of object who-questions (Wh\_bare\_Tar). Within-group comparisons detected object which-questions, object relatives, as well as subject relatives and finite complement clauses (ComFin\_Tar) as being difficult for bilingual children.

The subgroup comparisons showed that Arabic–L1 children ( $n = 27$ ) struggled more with identical repetition of finite complement clauses in the L2 French LITMUS–SR than MoTD children ( $U = 304, p < 0.05, r = 0.350$ ) and had difficulties with identical repetition of subject relatives ( $U = 281, p < 0.05, r = 0.376$ ) and object relatives ( $U = 314, p < 0.001, r = 0.368$ ), with medium effect sizes. Identical repetition of these structures was also difficult for the 18 Portuguese L1 children (SR:  $U = 184.5, p < 0.05, r = 0.365$ ; OR:  $U = 208, p < 0.05, r = 0.363$ ), in addition to object which-questions using both scoring measures (Wh–NP\_Id,  $U = 242, p < 0.05, r = 0.363$ ; Wh–NP\_Tar:  $U = 241, p < 0.05, r = 0.359$ ). When compared with each other, these two bilingual groups only differed in one aspect, namely in the repetition of Object–Which–questions, where BiTDs with L1 Arabic performed significantly better than BiTDs with L1 Portuguese (BiTD–A/BiTD–P:  $U = 170, p < 0.05, r = 0.400$ ).

Compared to the MoTD group, the BiTD–T children performed significantly lower on both identical repetition and target structure of finite complement clauses (CompFin\_Id:  $U = 183.5, p < 0.001, r = 0.516$ ; CompFin\_Tar:  $U = 230.0, p < 0.001, r = 0.459$ ), subject relatives

(SR\_Id:  $U = 112.5, p < 0.001, r = 0.632$ ; SR\_Tar:  $U = 138.0, p < 0.001, r = 0.604$ ), object relatives (OR\_intv\_Id:  $U = 130.0, p < 0.001, r = 0.651$ ; OR\_intv\_Tar:  $U = 177.0, p < 0.001, r = 0.589$ ), and object which-questions: (Wh-NP\_Id:  $U = 175.0, p < 0.001, r = 0.632$ ; Wh-NP\_Tar:  $U = 265.5, p < 0.001, r = 0.472$ ).

The Turkish L1 group also displayed difficulties with both the identical repetition and the target structure measures of object Which-questions with medium to strong effect sizes when compared to the Arabic group (Wh-NP\_Id:  $U = 160.5, p < 0.001, r = 0.493$ ; Wh-NP\_Tar:  $U = 185.5, p < 0.001, r = 0.494$ ) and subject relatives (SR\_Id:  $U = 184.5, p < 0.05, r = 0.369$ ; SR\_Tar:  $U = 185, p < 0.05, r = 0.445$ ), as well as with meeting the target structure of OR ( $U = 147.5, p < 0.001, r = 0.519$ ).

Table 3 summarizes the results with respect to significant differences between the MoTD children and the BiTD subgroups in France and Germany and allows for an overview of L1 language specific errors.

Table 3: Statistically significant differences between BiTD children (split by L1) compared to their MoTD peers on specific features and SR–substructures

L1/structure	LITMUS–SR–French			LITMUS–SR–German		
	Arabic	Portuguese	Turkish	Arabic	Portuguese	Turkish
Gender on Determiner	no	no	yes	<i>no</i>	<i>no</i>	<i>no</i>
Omission of Determiner	no	no	no	no	no	no
Omission of Auxiliaries	no	no	yes	no	no	no
Complementizer Errors	no	no	yes	no	no	no
Finite Complement Clauses	(yes)*	(yes)	yes	no	no	no
Subject Relatives	(yes)	(yes)	yes	(yes)	no	no
Object Relatives	(yes)	(yes)	yes	<i>no</i>	<i>no</i>	<i>no</i>
Object–Who–Question	yes	no	no	(yes)	no	no
Object–Which–Questions	no	yes	yes	no	no	no
Case on Determiners	n.a.	n.a.	n.a.	no	yes	no

Colored cells:

Parentheses: Significant difference were found for Identical repetition, but not for Target structure.

Italics: All groups (including MoTD) produced errors in this category (cf. Appendix G).

## 7. Discussion

The detailed cross–linguistic results based on three different first and two typologically different second languages allow for new insights on the applicability of the LITMUS–QU–NWR and the LITMUS–SR to children with diverse language backgrounds in different countries. The above–mentioned analyses confirm that the German and the French versions of these tasks reliably distinguish bilingual children with and without SLI (see Figures 1–4). Importantly, it is not the case that some minority language groups were disadvantaged per se by an L1–driven influence on L2 German or L2 French.

In our study, L1 influence on the L2s was expected to be visible in those structures where French and German share the same features since the effects would be seen by children in the same L1 groups in both countries. The Turkish group indeed showed L1 influence in several grammatical aspects, but, contrary to assumptions, these showed up predominantly in French. Our comparisons lead to the impression that on the whole, bilingual children had more difficulties with L2-specific structures than difficulties due to L1 influence. Not all error types predicted by L1 influence appeared in our data, and L1 group errors differed between countries.

First, for the QU–NWR, BiTD children performed like MoTD children. Contrary to earlier findings that showed a disadvantage of bilingual children in NWRs that contained language dependent items (Schöler, 2006; Windsor et al., 2010), our systematic analyses strengthened the impression that even on the language dependent items of both parts of the LITMUS–QU–NWR, typically developing bilingual children with typologically different L1s performed equally well on each task. Error rates in TD children were low, and they were shown to be independent from language dominance. Moreover, no significant differences were found between monolingual and bilingual typically developing children on either the language independent or the language dependent items of the task, in either the French or the German versions of the LITMUS–QU–NWR. Since both LITMUS tools were carefully designed to prevent errors caused by limited L2-experience, our results confirm earlier findings (Grimm & Hübner, to appear; Hamann & Abed Ibrahim, 2017; Tuller et al., 2018; among others) in that the French and German versions of the NWR are applicable and easy to administer tasks for the separation of BiSLI and BiTD children in their second language.

Second, we found no general influence of L1 on the L2 performance in the LITMUS–SR, though some residual L1 influence is possible. Given the typological differences between the three L1s and the grammatical similarities between French and German (cf. Table 1). Arabic L1–bilinguals would have been expected to show difficulties

with gender marking on determiners. Portuguese L1 children should have shown more inaccuracies with German gender, and Turkish children would have been expected to display a variety of errors such as gender marking on determiners, omission of definite determiners and auxiliaries, and *wh*-in situ instead of *wh*-movement.

Sentences in the French and German SRs contained determiners which are marked for gender and some had auxiliaries and/or complementizers. No significant differences between monolingual and bilingual typically developing children were found in the German SR, but did occur in the French SR. With respect to the performance of bilingual children on the two sentence repetition tasks, our analysis reveals that some language-specific structures, such as overt case and gender marking on the determiner in German, could indirectly affect the performance scores on, for example, object relatives or object questions. This is because case marking is necessary for the correct realization of the latter structures. Due to case marking, Arabic-speaking children struggled with subject relatives and object questions in both L2 tasks. As expected, Portuguese children had difficulties assigning case in German, but, at the same time, did not differ from monolingual children with respect to determiner use or object relatives in that language. However, as Abed Ibrahim & Hamann (2017) point out, Portuguese L1 children might show insecurity about the choice of relative pronoun in German, e.g. overuse of feminine gender, substitution of the relative pronoun in object relatives. The identical repetition of object relatives was also difficult for Portuguese-speaking children in French. Note, that differences between MoTD and BiTD in Germany were not detected, since both the bilingual and the monolingual children had difficulties with object relatives in the German SR.

Globally, compared to their monolingual typically developing peers, the bilingual children in France produced more errors on finite complement clauses and subject relatives than the German bilinguals, irrespective of their L1, when the task was scored by identical repetition. Object relatives were difficult for bilingual children in both countries. When the



task was scored by “target structure met”, only the BiTD–Turkish children performed significantly worse than their MoTD peers on the aforementioned structures.

Language-group comparisons showed that these cannot be due exclusively to L1 influence, but rather could be traced back to the specific features of object and subject relatives in French, and embedding appeared to be more fault–prone for L2 learners with Turkish as an L1 in France. The same holds for German since errors explainable by L1 influence were only seen in Portuguese children, with a strong effect of case, and for Arabic-speaking bilinguals in Germany, who showed marginal difficulties with subject relatives. As a whole, most errors were not specific to one language group.

Considering hypotheses on L1 transfer at the morphosyntactic level, it is relevant that specific L2 features causing difficulties in L2 children and which are known to also be sources of difficulty in SLI, such as verbal morphology (SVA) and pronominal clitics, were a priori expressly not targeted by these SRs. Faulty feature–assembly (Lardiere, 2009) could explain auxiliary omission and difficulties with gender or case assignment on determiners.

Even though mainly the Turkish L1 children in France showed many structures that align with the idea of errors being caused by L1 feature transfer, the picture is not consistent. It is not true that L1 transfer from Turkish manifests itself in both the German and the French SR as predicted (cf. Table 1). The Turkish groups did not exhibit the same errors in the German and in the French SR. Turkish children were expected to struggle with the determiner systems, auxiliaries, gender assignment, and word order in both L2s. Turkish L1 children in France produced errors with embedding, and, more specifically, with the correct repetition of French relative clauses. Whereas the Arabic and the Portuguese L1 children had higher target structure scores, difficulties remained for the Turkish children in France even when scoring the SR by target structure.

Both SR tasks were also tested for language dominance effects. L2 dominance correlated positively with SR performance, but no significant differences were found with

respect to exposure to the L2, AoO, or LoE on the BiTD groups in the two countries considered here. Since the Turkish group in France showed many more errors that pattern with the expected cross-linguistic differences between Turkish as L1 and the L2 (Chiat et al., 2014), these difficulties might be traced back to different ages of onset or lengths of exposure when compared to Turkish L1 children in Germany. In support of this hypothesis, Turkish children in France constituted the biggest L1-dominant group within the French BiTDs (20 of the 36 L1 dominant children were L1 Turkish). Considering the aspect of an earlier or different L2 contact between the Turkish groups in France and Germany, it is thus not surprising that the Turkish L1 group in France showed more L1-specific errors than the Turkish children in Germany. Since transfer effects are commonly assumed to influence predominantly early stages of acquisition, it might also be true that most of the children in Germany had already acquired the L2 structures used in the SRs, since the children were L2 dominant. Probably, the higher error rates of French BiTD children in comparison to German BiTD children with Turkish as L1 on sentences that require embedding could also be explained by sociolinguistic influence on heritage language use. Our bilateral comparison opens a new perspective to different language practices in the Turkish communities in the two countries and their influence on language assessment tasks. Most Turkish-German bilinguals differ from their French peers with respect to the amount of early L2 exposure and, probably more importantly, the exposure to a variety of Turkish (“Immigrant Turkish”) that is different from Standard Turkish. Further comparative analyses of performance on standardized L1 tests indicate that the Turkish bilingual children that were tested in France spoke a standard variety when compared to the Turkish bilingual children in Germany, who predominantly had acquired Immigrant Turkish (i.e. Akinci, 2001, 2014; Chilla & Şan, 2017; Hamurcu, 2015). Turkish children of the 2<sup>nd</sup> and 3<sup>rd</sup> generations in Germany tend to construct relative clauses in Turkish using two clauses and having the head noun precede the clause. Children in Germany who acquire such a contact-induced L1 variety of Turkish might profit from this in

the German SR since embedding in the L2 and in their variety of the L1 align. The Turkish-French children in our study, however, were able to produce relative clauses in their L1 that conform to the standard variety in Turkish (Şan & Chilla, 2018), leading to the conclusion that subject and object relatives in the French version of the LITMUS–SR might disadvantage L1 dominant children acquiring standard Turkish. Further sociolinguistic studies on the minority language status in the two countries might focus on the influence of language communities and language prestige for the assessment of L1 and L2 in bilingual minority children and give rise to the interpretation of language-variety dependent errors in sentence repetition tasks expected by different language minority groups within one first language.

The results provided here shed light on the applicability of LITMUS SR tasks for clinical use. Coding procedures should reflect that identical repetition of sentences is easier to code, on a practical level, but will necessarily include error types such as case and gender assignment. Tasks for bilingual children with diverse L1 backgrounds cannot be narrowed down to a few particular structures as in established tests for the assessment of SLI in monolingual children, and the error patterns vary. As a consequence, assessment of bilingual children would appear to require bilingual norms, should furthermore incorporate a variety of morphosyntactic structures, and should be designed after careful consideration of the child L2–SLI overlap in order to correct for difficulties caused by potential L1 influence. Revised versions of the SR tasks might, for instance, consider auxiliaries, determiners, and features spelled out on determiners to be sensitive to L1 influence more carefully. For example, in German, if all nouns in the sentences were feminine, this would make it possible to avoid case marking (as the definite determiner *die* is invariant for case).

Clinical practice should carefully investigate language background measures. Background questionnaires ought to include questions about language varieties spoken at home since they might have an impact on the recommended scoring procedure for bilingual children. The L1 background could influence the performance of certain language groups, as

the difference between 1<sup>st</sup> and 2<sup>nd</sup> generation Turkish L1 children in France and 2<sup>nd</sup> and 3<sup>rd</sup> generation children in Germany indicates. In SRs, error-prone items should be listed as possible outliers in the scoring manual.

## **8. Conclusion**

Our detailed cross-linguistic comparison focused on L1-driven influence, taking as its point of departure the fact that both LITMUS-tools were carefully designed to limit L1 influence. In our study, these tools have been shown to be capable of reliably assessing bilingual children from typologically different minority language groups in different countries. The two QU-NWRs were especially resilient to impact from L1 influence or from language dominance. The French and German LITMUS-SRs are also capable of differentiating between children with SLI and typically developing bilingual children, though some structures and properties proved to be equally difficult for these groups, such as case and gender assignment in German, and gender on the determiner in French. Our cross-linguistic comparison further leads to the impression that different L1 varieties spoken in the family might have an impact on the performance and the scoring of substructures in SR tasks. Some L2 structures involving embedding appeared to be more fault-prone for L2 learners with Turkish as an L1 in France, but not in Germany. In addition, thorough and sensitive scoring procedures (such as a choice between identical repetition and production of the targeted structure) are crucial. As argued in Hamann & Abed Ibrahim (2017) scoring should be adapted to L2 properties and needs further investigation with different language pairs.

**Acknowledgements:**

The authors are grateful to the children and parents who participated in this study. The Bilingual Language Development: Typically Developing Children and Children with Language Impairment (BiLaD) research project was funded by a bilateral grant (French ANR-12-FRAL-0014-01 and German DFG: HA 2335/6-1, CH 1112/2-1, RO 923/3-1).

## 9. References

- Abed Ibrahim, L., Fekete, I. 2019. What Machine Learning can tell us about the (non-)role of language dominance in the diagnostic accuracy of German LITMUS nonword and sentence repetition tasks. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2018.02757.
- Akinci, M.A. 2001. Développement des compétences narratives des enfants bilingues turc-français en France âgés de 5 à 10 ans, München.
- Akinci, M.A. 2014. Fransa'daki Türk Göçmenlerinin Etnik ve Dinî Kimlik Algıları, *Bilig*, 70, 29-58.
- Alloway, T.P., Gathercole, S.E. 2005. Working memory and short-term sentence recall in young children. *The European journal of cognitive psychology* 17(2), 207–220.
- Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, Ch., Tuller, L., Zebib, R., and Barthez, M. –A. 2017. Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism* 7(3/4). doi: 10.1075/lab.15019.alm.
- Archibald, L.M.D., Gathercole, S.D. 2006. Nonword Repetition. A comparison of tests. *Journal of Speech, Language, and Hearing Research* 49, 970–983.
- Armon-Lotem, S. 2014. Between L2 and SLI: Inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. *Journal of Child Language* 41(1), 3–33.
- Armon-Lotem, S, de Jong, J., Meir, N. (eds.) 2015. *Assessing multilingual children: Disentangling bilingualism from language impairment* Bristol, UK: Multilingual Matters.

- Barlow, J.A. 2014. Age of acquisition and allophony in Spanish–English bilinguals. *Frontiers in Psychology*, 5, 1–14.
- Belletti A., Hamann C. 2004. On the L2/bilingual acquisition of French by two young children with different source languages. In P. Prévost, J. Paradis, (eds.). *The acquisition of French in different contexts: Focus on functional categories*, Amsterdam/Philadelphia: John Benjamins, 147–174.
- Belletti, A., Bennati, E., Sorace, A. 2007. Theoretical and developmental issues in the syntax of subjects: Evidence from near–native Italian". *Natural Language and Linguistic Theory*.
- Blom, W.B.T. 2008. Testing the Domain–by–Age Model: inflection and placement of Dutch verbs. In B. Haznedar & E. Gavrusseva (eds.). *Current trends in child second language acquisition: a generative perspective*, Amsterdam: Benjamins, 271–300.
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., and Blom, E. (2015). A quasi–universal nonword repetition task as a diagnostic tool for bilingual children learning dutch as a second language. *JSLHR* 58(6), 1747–1760.
- Chilla, S., Bonnesen, M. 2011. A cross–linguistic perspective on child SLA: The acquisition of questions in German and French. *Linguistische Berichte* 228, 413–442.
- Bosch, L., Costa, A., & Sebastián–Gallés, N. 2000. First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12(2), 189–222.
- Chiat, S. 2015. Non–word repetition. In: Armon–Lotem, S., de Jong, J., Meir, N. (eds.), *Assessing multilingual children. Disentangling bilingualism from language impairment. Communication disorders across languages* 13. Bristol, UK: Multilingual matters, 125–150.

- Chiat, S., Polišenská, K. 2016. A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research* 59(5), 1179–1189.
- Chiat, S., Torgerson, C., Snowling, M.J., Nag, S. 2014. *Literacy, Foundation Learning and Assessment in Developing Countries*. London: Department for International Development. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/305150/Literacy-foundation-learning-assessment.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/305150/Literacy-foundation-learning-assessment.pdf).
- Chilla, S. 2008. *Erstsprache, Zweitsprache, Spezifische Sprachentwicklungsstörung? Eine Untersuchung des Erwerbs der deutschen Hauptsatzstruktur durch sukzessiv-bilinguale Kinder mit türkischer Erstsprache*. Hamburg: Dr. Kovač.
- Chilla, S., Şan, N.H., 2017. Möglichkeiten und Grenzen der Diagnostik erstsprachlicher Fähigkeiten: Türkisch-deutsche und türkisch-französische Kinder im Vergleich. In C. Yildiz et al. (eds.) *Sprachen 2016: Russisch und Türkisch im Fokus*. Berlin: Lang, 175–205.
- Clahsen, H. 1990. *Child language and developmental dysphasia. Linguistic studies of the acquisition of German*. Amsterdam: John Benjamins.
- Clahsen, H., Meisel, J.M., Pienemann, M. 1983. *Deutsch als Zweitsprache: deutscher Spracherwerb ausländischer Arbeiter*. Tübingen: Narr.
- Clahsen, H., Rothweiler, M., Sterner, F., Chilla, S. 2014. Linguistic markers of specific language impairment in bilingual children: The case of verb morphology. *Clinical Linguistics & Phonetics* 28(9), 709–721.
- Conti-Ramsden, G., Botting, N., Faragher, B. 2001. Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry* 42(6), 741–748.



- de Abreu, P.M.J.E, Cruz, S., Puglisi, M. L. 2013. Specific language impairment in language minority children from low income families. *International Journal of Language & Communication Disorders* 49(6), 736–747.
- de Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., & Barthez, M. A. (2017). Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism*. 7(3–4), 331–58.
- dos Santos, C. & Ferré. S., 2018. A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition* 25(1), 58–71.
- Fabiano–Smith, L., Barlow, J.A. 2010. Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism* 73(1), 81–97.
- Ferré, S., Tuller, L., Sizaret, E. and Barthez, M.A. 2012 Acquiring and avoiding phonological complexity in SLI vs. typical development of French: The case of consonant clusters. *Consonant Clusters and Structural Complexity*. Berlin: De Gruyter, 285–308.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics, 5. Ed.*. Sage Publications.
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., Zebib, R. 2016. How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition*, 1–17: doi.org/10.1080/10489223.2016.1192635
- Gallon, N., Harris, J., van der Lely, H. 2007. Nonword repetition: An investigation of phonological complexity in children with grammatical–SLI. *Clinical Linguistics and Phonetics* 21.435–55.
- Granfeldt, J., Schlyter, S. 2004. Cliticisation in the acquisition of French as L1 and L2. In Prévost, P., Paradis, J. (eds.), *The Acquisition of French in Different Contexts* 32, Amsterdam: John Benjamins, 333–370.

- Grimm, A. & Hübner, J. (to appear). Nonword repetition by bilingual learners of German. The role of language-specific complexity. In dos Santos, C., de Almeida, L. (eds.) *Bilingualism and Specific Language Impairment*. Amsterdam: Benjamins.
- Grüter, T., Crago, M. 2012. Object clitics and their omission in child L2 French: The contributions of processing limitations and L1 transfer. *Bilingualism: Language and Cognition* 15(3), 531–549.
- Gutiérrez–Clellen, V.F., Simon–Cerejido, G. 2010. Using nonword repetition tasks for the identification of language impairment in Spanish–English speaking children: Does the language of assessment matter? *Learning Disabilities Research & Practice* 25, 48–58.
- Haberzettl, S. 2003. ‘Tinkering’ with chunks: Form-oriented strategies and idiosyncratic utterance patterns without functional implications in the IL of Turkish speaking children learning German. In: Dimroth, Ch., Starren, M. (eds.) *Information structure, linguistic structure and the dynamics of language acquisition*. Amsterdam/Philadelphia: Benjamins (Studies in Bilingualism), 45–63.
- Hamann, C., 2012. Bilingual development and language assessment. In Biller, A.K., Chung, E.Y., Kimball, A.E. (eds.) *Proceedings of Proceedings of the 36<sup>th</sup> Annual*. Somerville, MA: Cascadilla Press, 1–28.
- Hamann, C., Abed Ibrahim, L. 2017. Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Psychology* 2(16).
- Hamann, C., Belletti, A. 2006. Developmental Patterns in the Acquisition of Complement Clitic Pronouns: Comparing Different Acquisition Modes with an Emphasis on French. *Rivista di Grammatica Generativa* 31 (appeared Jan 2008), 39–78.

- Hamann, C., Penner, Z. Lindner, K. 1998. German Impaired Grammar: The Clause Structure Revisited. Special issue 'Specific Language Impairment in Children'. *Language Acquisition* 7, 193–246.
- Hamurcu Süverdem, B. (2015), Développement du turc et du français en situation de bilinguisme précoce. Le cas d'enfants d'origine turque scolarisés en maternelle, Unpublished PhD, Rouen: Université de Rouen.
- Haznedar, B. 1997. Child second language acquisition of English: a longitudinal case study of a Turkish-speaking child. Thesis (Ph.D.) – University of Durham.
- Hulk, A., Müller, N. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism* 3(3), S. 227–244.
- IBM Corp, 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.
- Jakubowicz, C, Nash, L. Rigaut, C., Gérard, C. –L. 1998. Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition* 7(2–4), 113–160.
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.A., Gustafsson, J.E., Hulme, C. 2015. Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science* 18(1), 146–154.
- Kohnert, K., Windsor, J., Yim, D. 2006. Do language-based processing tasks separate children with primary language impairment from typical bilinguals? *Journal of Learning Disabilities Research & Practice* 21(1), 19–29.
- Lardiere, D. 2009. Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25, 173–227.
- Leonard, L.B. 2014. *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.

- Marinis, T., Armon–Lotem, S., Pontikas, G. 2017. Language impairment in bilingual children: state of the art 2017. *Linguistic Approaches to Bilingualism*, 7 (3–4), 265–276.
- Marshall, C.R., van der Lely, H.K.J. 2009. Effects of word position and stress on onset cluster production: Data from typical development, SLI and dyslexia. *Language* 85, 39–57
- Meir, N., Walters, J., Armon–Lotem, S. 2016. Disentangling SLI and bilingualism using sentence repetition tasks: The impact of L1 and L2 properties. *International Journal of Bilingualism* 20(4), 421–452.
- Meisel, J.M., Elsig, M., Bonnesen, M. 2011. Delayed Acquisition of Grammar in First Language Development: Subject–Verb Inversion and Subject Clitics in French Interrogatives. *Linguistic Approaches to Bilingualism* 1(4), 347–390.
- Montanari, E. 2010. *Kindliche Mehrsprachigkeit: Determination und Genus*. Dissertation. Münster: Waxmann.
- Müller, N., Hulk, A. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4, 1–21.
- Paradis, J., 2001. Are object omissions in Romance object clitic omissions? *Bilingualism: Language and Cognition* 4(1), 36–37
- Paradis J., 2010. The interface between bilingual development and specific language impairment. *Applied Psycholinguistics* 31, 227–252.
- Paradis, J., Genesee, F. 1996. Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition* 18, 1–25.
- Paradis, J., Rice, M. L., Crago, M., Marquis, J., 2008. The acquisition of tense in English: Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics* 29(4), 689–722.

- 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, 971–981.
- Pierce L. J., Genesee F, Delcenserie A., Morgan G. 2017. Variation in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics*, 38, 1265–1302.
- Polišenská, K., Kapalkova, S. 2014. Improving child compliance on a computer administered nonword repetition task. *Journal of Speech, Language, and Hearing Research* 57, 1060–1068.
- Prévost, P. 2009. *The acquisition of French: the development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. Amsterdam: Benjamins.
- Prévost, P., Strik, N., Tuller, L. 2014. Wh-questions in child L2 French: Derivational complexity and its interactions with L1 properties, length of exposure, age of exposure, and the input. *Second Language Research* 30(2), 225 – 250.
- Restrepo, M.A. 1998. Identifiers of Predominantly Spanish-Speaking Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, (41), 1398–1411.
- Riches, N. G. 2012. Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders* 47(5), 499–510.
- Scheidnes, M., & Tuller, L. 2014. L2 children embed normally, but children with SLI do not. In A. Fieis, Freitas, M.J. (eds.), *New directions in the acquisition of romance languages: Selected proceedings of the Romance Turn V*. Newcastle upon Tyne: Cambridge Scholars Publishing, 261–284.

- Schöler, H. 2003. Sprachleistungsmessungen. In Bredel, U., Günther, H., Klotz, P., Ossner, J. & Siebert–Ott, G. (eds.) *Didaktik der deutschen Sprache – Ein Handbuch*, 2. Teilband. Paderborn: Schöningh UTB, 898–913.
- Schöler, H. 2006. HASE – ein Screening zur Früherkennung von Lese– und Rechtschreibstörungen. In Schulte–Körne, G. (ed.) *Legasthenie und Dyskalkulie in Wissenschaft, Schule und Gesellschaft*. Bochum: Winkler, 39–46.
- Schönenberger, M. 2011. Prolonged stage of article omission in by successive bilingual children with L1 Turkish. ISB & meeting. Oslo, JuNE, 2011.
- Schönenberger, M., Rothweiler, M., Sterner, F. 2012. Case marking in child L1 and early child L2 German. In Braunmüller, K., Gabriel, C. (eds.) *Multilingual Individuals and Multilingual Societies [Hamburg Studies on Multilingualism 13]*. Amsterdam: Benjamins, 3–22
- Schwartz, B.D. 2003. Child L2 acquisition: Paving the way. In Beachley, B., Brown, A., Conlin, F. (eds.) *Proceedings of the 27th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 26–50.
- Scontras, G., Fuchs, Z., Polinsky, M. 2015. Heritage language and linguistic theory. *Frontiers in Psychology* 6, 1545.
- Sebastián–Gallés, N., Bosch, L. 2005. Phonology and Bilingualism. In Kroll, J.F., De Groot, A.M.B. (eds.) *Handbook of Bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press, 68–87.
- Stothard, S.E., Snowling, M., Bishop, D.V.M., Chipchase, B.B., Kaplan, C.A. 1998. Language–impaired preschoolers: a follow–up into adolescence. *Journal of speech, language, and hearing research* 41(2), 407–418.

- Summers, C., Bohman, T.M., Gillam, R.B., Pena, E.D. & Bedore, L.M., 2010. Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language and Communication Disorders* 45 (4), 480–493. Somerville, MA: Cascadilla Proceedings.
- Şan, H. & Chilla, S. 2018. Identifying Specific Language Impairment in the L1 of bilingual Turkish-speaking children in Germany and France. Paper presented at the 2018 BiSLI conference, Reading.
- Thordardottir, E. 2015. Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In Armon-Lotem, S., Marinis, T., Meir, N. (eds.) *Assessing Multilingual Children Disentangling Bilingualism from Language Impairment*, Bristol: Multilingual Matters, 331–358.
- Thordardottir, E., Brandeker, M. 2013. The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders* 46, 1–16.
- Tsimpli, I., Dimitrakopoulou, M. 2007. The Interpretability Hypothesis : evidence from wh-interrogatives in second language acquisition. *Second Language Research* 23, 215–242.
- Tuller, L., 2015. Clinical use of parental questionnaires in multilingual contexts. In Armon-Lotem, S., de Jong, J., Meir, N. (eds.) *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters, 301–330.
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Abed Ibrahim, L., Zebib, R. 2018. Identifying language impairment in bilingual children in France and

- Germany. *International Journal of Language and Communication Disorders* 53 (4), 888–904.
- Unsworth, S. 2005. *Child L2, Adult L2, Child L1: Differences and Similarities. A study on the acquisition of direct object scrambling in Dutch.* PhD thesis, Utrecht University.
- Vinther, T. 2002. Elicited imitation: A brief overview. *International Journal of Applied Linguistics* 12, 54–73.
- Wegener, H. 1994. Variation in the acquisition of German plural morphology by second language learners. In Tracy, R., Lattey, E. (eds.) *How tolerant is universal grammar? Essays on language learnability and language variation.* Tübingen: Niemeyer, 267–294.
- Werker, J.F., Tees, R.C. 1983. Developmental change across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, 37, 278–286.
- Whong-Barr, M., Schwartz, B.D. 2002. Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition*, 24, 579–616.
- Windsor, J., Kohnert, K., Lobitz, K.F., Pham, G. 2010. Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*. 19, 298–310.
- Yip, V. & Matthews, S. 2000. Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition* 3(3), 193–208.
- Zdorenko, T., Paradis, J. 2008. The acquisition of articles in child second language English: Fluctuation, transfer or both? *Second Language Research*, 24, 227–250.
- Zdorenko T., Paradis, J. 2011. Articles in Child L2: When L1 and L2 acquisition meet at the interface. *First Language* 32(1–2), 38–62.





## Appendix A

### LITMUS–NWR: Content of Language Independent and Dependent Items

	Vowels (V)	Consonants (C)	Syllable Types	Examples
Language Independent (LI): 30 Items	/a, i, u/	/p, k, f, l/	CV CCV CVC#	faku klipafu, fupla fuk, kafip
Language Dependent (LD) – French: 41 Items	/a, i, u/	/p, k, f, l/  In addition : /s/	In addition : #sCV, #sCCV sC#, Cs#, internal /l/	skafu, skla, pusk piks, filpa
Language Dependent (LD) – German: 36 items	/a, i, u/	/p, k, f, l/  In addition :  /s/ , /ʃ/	In addition :#sCV, #sCCV Cs#, internal /s/	skifapu, ʃplaklu kapifaps, kufiski

## Appendix B

### Common structures (and substructures) in French and German LITMUS–SR

Core structure	Substructure	Example
Monoclausal	Present tense	La maman lit une histoire the mother reads a story 'The mother is reading a story'
	Past tense	Le lapin a mangé la carotte The rabbit has eaten the carrot 'The rabbit ate the carrot'
Object wh-question	<i>Who</i> -question	Wen umarmt der Pinguin heute? who.ACC hugs the.NOM penguin today 'Whom does the penguin hug today?'
	<i>Which</i> <i>N</i> -question	Welchen Clown besucht der Zauberer? which.ACC Clown visits the.NOM magician 'Which clown does the magician visit?'
Clausal complement	Infinitival	Le papa sait très bien conduire la voiture the father knows very well drive-INF the car 'The father can drive the car very well'
	Finite	La fille croit que le papi a fini sa soupe the girl believes that the grandpa has finished his soup 'The girl believes that the grandpa ate up his soup'
Relative clause	Subject relative	Ich sehe den Roboter, der den Cowboy weckt I see the.ACC robot who.NOM the.NOM cowboy wakes up 'I see the robot who wakes up the cowboy'
	Object relative	Je vois le garçon que la fille a poussé I see the boy that the girl has pushed 'I see the boy that the girl pushed'

## Appendix C

### Effect of (clinical) group membership on LITMUS–QU–NWR and LITMUS–SR measures

Kruskal–Wallis Test	Germany (all groups)	France (all groups)
NWR_correct	$\chi^2 (2, N = 66) = 16.492,$ $p < 0.001$	$\chi^2 (2, N = 103) = 31.247,$ $p < 0.001$
NWR_LI	$\chi^2 (2, N = 66) = 11.994,$ $p < 0.001$	$\chi^2 (2, N = 103) = 23.125,$ $p < 0.001$
NWR_LD	$\chi^2 (2, N = 66) = 16.835,$ $p < 0.001$	$\chi^2 (2, N = 103) = 30.403,$ $p < 0.001$
SR_Id	$\chi^2 (2, N = 66) = 22.184,$ $p < 0.001$	$\chi^2 (2, N = 132) = 47.622,$ $p < 0.001$
SR_Tar	$\chi^2 (2, N = 66) = 21.214,$ $p < 0.001$	$\chi^2 (2, N = 132) = 42.649,$ $p < 0.001$

## Appendix D

### Significant pairwise comparisons using Mann–Whitney U–Test (NWR\_correct, NWR\_LI, NWR\_LD, SR\_Id and SR\_Tar)

Mann–Whitney U–Test	Germany		France	
	MoTD vs. BiSLI	BiTD vs. BiSLI	MoTD vs. BiSLI	BiTD vs. BiSLI
NWR_correct	$U= 1.000,$ $p = 0.001,$ $r=0.816$	$U= 24.5,$ $p< 0.001,$ $r=0.524$	$U= 19,$ $p< 0.001,$ $r=0.513$	$U= 204,$ $p< 0.001,$ $r=0.528$
NWR_LI	$U= 5.5,$ $p = 0.001,$ $r=0.727)$	$U= 51,$ $p< 0.001,$ $r=0.443$	$U= 30.5,$ $p< 0.001,$ $r=0.462$	$U= 279,$ $p< 0.001,$ $r=0.458$
NWR_LD	$U= 0.500,$ $p< 0.001,$ $r=0.829$	$U= 0.23,$ $p< 0.001,$ $r=0.529$	$U= 20.5,$ $p< 0.001,$ $r=0.507$	$U= 227,$ $p< 0.001,$ $r=0.507$
SR_Id	$U= 0,$ $p< 0.001,$ $r=0.84$	$U= 44.5,$ $p< 0.001,$ $r=0.461$	$U= 47.5,$ $p< 0.001,$ $r=0.767$	$U= 44.5,$ $p< 0.001,$ $r=0.461$
SR_Tar	$U= 0,$ $p< 0.001,$ $r=0.839$	$U= 33.5,$ $p< 0.001,$ $r=0.496$	$U= 57.5,$ $p< 0.001,$ $r=0.764$	$U= 33.5,$ $p< 0.001,$ $r=0.496$

## Appendix E

### Comparison of BiSLI and BiTD children on substructures and grammatical features relevant for SR (only non-significant measures, $p > 0.05$ )

Country	Structure/Error	Bi-TD M(SD)-%errors	Bi-SLI M(SD)-%errors	<i>U</i>	<i>p</i>	<i>R</i>
Germany	TOP_Id	70.85 (37.46)	95.88 (11.67)	121.5	0.063	.229
	TOP_Tar	67.27 (39.70)	87.50 (24.87)	134	0.140	.182
	Case on D	11.60 (7.52)	15.62 (7.69)	125	0.177	.193
	Aux_Om	0.50 (2.05)	0.0 (0.0)	176	0.401	.103
France	Gender_on D	3.77 (7.48)	3.31 (5.07)	834	0.549	.063

## Appendix F

### Performance of all subgroups (*M* and *SD*) on the French and German versions of the NWR and the SR

	Germany					France				
	MoTD	BiTD-A	BiTD-P	BiTD-T	BiSLI	MoTD	BiTD-A	BiTD-P	BiTD-T	BiSLI
NWR	74.20 (9.47)	76.00 (16.06)	74.95 (15.47)	70.05 (15.38)	45.50 (8.73)	90.33 (8.97)	87.11 (10.24)	84.16 (11.57)	82.33 (16.58)	66.04 (16.62)
NWR_LI	78.6 (9.14)	82.27 (11.68)	79.60 (13.58)	73.64 (15.09)	59.00 (9.75)	88.83 (10.42)	86.77 (11.75)	81.83 (12.30)	82.95 (15.50)	69.00 (13.73)
NWR_LD	70.6 (11.59)	70.73 (21.43)	71.00 (18.33)	66.82 (16.56)	34.12 (10.57)	91.25 (10.68)	87.41 (10.09)	85.33 (11.70)	81.91 (18.5)	63.77 (20.79)
SR_Id	81.2 (11.23)	57.55 (26.99)	53.90 (25.42)	57.00 (28.46)	17.50 (15.49)	93.05 (8.69)	78.55 (24.52)	74.83 (27.62)	58.37 (30.22)	39.23 (26.84)
SR_Tar	88.5 (7.58)	76.45 (22.25)	67.90 (22.79)	68.35 (25.72)	27.87 (16.20)	96.70 (6.16)	93.92 (7.81)	89.27 (14.54)	77.83 (21.75)	65.69 (21.01)

## Appendix G

### Performance of all subgroups (*M* and *SD*) on the French and German substructures of the SR

	Germany					France				
	MoTD	BiTD– A	BiTD– P	BiTD– T	BiSLI	MoTD	BiTD– A	BiTD– P	BiTD– T	BiSLI
Gender on D	1,2 (1.61)	5,45 (7.65)	4.15 (4.90)	7.76 (9.70)	24.87 (27.26)	0.00 (0.00)	1.33 (4.53)	0.22 (0.94)	9.17 (9.71)	3.3 (5.06)
D–om.	0,2 (0,42)	1 (2,09)	1.65 (6.22)	1.47 (5.33)	6.37 (8.65)	0.00 (0.00)	0.81 (2.93)	0.00 (0.00)	2.75 (6.68)	5.07 (10.88)
Aux_om.	0.00 (0.00)	0.00 (0.00)	0.65 (2.47)	0.65 (2.21)	0.00 (0.00)	0.27 (1.15)	1.55 (3.26)	2.94 (7.81)	4.75 (6.40)	6.88 (9.84)
C_errors	0.00 (0.00)	17.45 (33.03)	13.3 (31.32)	19.59 (39.18)	79.12 (39.64)	0.46 (2.79)	7.44 (16.25)	5.61 (12.83)	18.79 (22.63)	20.5 (20.66)
SVO_Id	92.3 (9.05)	69.73 (33.85)	74.45 (28.05)	70.00 (35.76)	29.25 (24.66)	96.83 (8.62)	87.63 (27.18)	86.06 (26.97)	76.37 (35.77)	53.88 (33.66)
SVO_Tar	96.7 (7.42)	85.91 (16.63)	82.35 (22.28)	82.41 (28.06)	51.50 (24.65)	98.19 (6.55)	95.04 (12.08)	94.39 (9.94)	90.21 (16.29)	72.31 (24.84)
Comp_Fin_Id	83.4 (32.35)	66.73 (44.72)	48.35 (39.80)	49.18 (35.69)	0.00 (0.00)	75.75 (31.08)	46.85 (40.63)	51.94 (41.62)	30.54 (39.25)	16.65 (31.64)
Comp_Fin_Tar	90.0 (31.62)	78.73 (37.41)	73.40 (33.53)	68.65 (39.95)	8.25 (15.28)	90.31 (19.00)	81.52 (33.75)	79.67 (30.57)	58.42 (39.67)	46.23 (39.04)
SR_Id	93.4 (13.91)	45.45 (37.42)	61.70 (39.45)	54.88 (40.79)	12.50 (24.87)	93.05 (8.69)	78.56 (24.52)	74.83 (26.63)	58.38 (30.22)	39.27 (26.82)
SR_Tar	96.7 (10.43)	66.73 (33.40)	66.75 (35.90)	60.76 (42.92)	12.50 (24.87)	96.86 (6.07)	93.70 (7.8)	89.28 (14.54)	78.00 (21.90)	65.85 (21.19)
OR_Id	40.0 (34.53)	30.36 (31.61)	21.60 (29.18)	17.65 (31.49)	0.00 (0.00)	89.19 (22.35)	64.22 (39.2)	63.00 (41.08)	37.50 (38.53)	20.50 (32.80)
OR_Tar	43.30 (31.75)	39.45 (41.75)	26.65 (35.26)	19.59 (31.45)	0.00 (0.00)	92.78 (19.52)	88.88 (28.2)	77.78 (32.40)	51.41 (40.56)	41.04 (41.48)
Wh_bare_Id	96.70 (10.43)	36.45 (37.98)	58.30 (43.13)	62.82 (37.09)	4.12 (11.67)	100 (0)	96.3 (19.24)	94.44 (17.19)	90.29 (20.83)	79.54 (28.43)
Wh_bare_Tar	96.70 (10.43)	75.73 (36.86)	78.35 (70.90)	80.41 (35.48)	20.88 (30.66)	100 (0)	97.52 (12.9)	96.28 (15.80)	100 (0)	93.65 (13.26)
Wh_NP_Id	93.30 (13.91)	63.64 (34.91)	56.65 (42.08)	68.71 (38.15)	20.88 (39.65)	96.38 (15.35)	90.11 (25.90)	81.50 (30.76)	59.79 (38.09)	47.42 (42.39)
Wh_NP_Tar	93.30 (13.91)	75.73 (30.27)	63.30 (35.80)	70.65 (38.89)	20.88 (39.65)	97.28 (12.14)	98.78 (25.89)	85.22 (23.51)	77.83 (28.97)	71.81 (36.19)
Top_Id	53.40 (35.92)	51.54 (45.65)	19.95 (33.15)	25.47 (32.39)	4.12 (11.68)	n.a.	n.a.	n.a.	n.a.	n.a.
Top_Tar	66.70 (35.20)	54.55 (47.80)	21.60 (32.92)	35.71 (37.83)	12.50 (24.87)	n.a.	n.a.	n.a.	n.a.	n.a.
Case on D.	5.40 (3.68)	11.36 (6.26)	12.65 (7.36)	10.52 (8.65)	15.62 (7.69)	n.a.	n.a.	n.a.	n.a.	n.a.