International Journal of Language & Communication Disorders

INT J LANG COMMUN DISORD, JULY–AUGUST 2018,
VOL. 53, NO. 4, 888–904

Research Report

Identifying language impairment in bilingual children in France and in Germany

Laurice Tuller†, Cornelia Hamann‡, Solveig Chilla§, Sandrine Ferré†, Eléonore Morin†, Philippe Prevost†, Christophe dos Santos†, Lina Abed Ibrahim‡ and Racha Zebib†

†UMR 1253, iBrain, Université de Tours, Inserm, Tours, France
‡University of Oldenburg, Oldenburg, Germany
§Europa-Universität Flensburg, Flensburg, Germany

(*Received July 2017; accepted April 2018*)

Abstract

Background: The detection of specific language impairment (SLI) in children growing up bilingually presents particular challenges for clinicians. Non-word repetition (NWR) and sentence repetition (SR) tasks have proven to be the most accurate diagnostic tools for monolingual populations, raising the question of the extent of their usefulness in different bilingual populations.
Aims: To determine the diagnostic accuracy of NWR and SR tasks that incorporate phonological/syntactic complexity as discussed in recent linguistic theory. The tasks were developed as part of the Language Impairment Testing in Multilingual Settings (LITMUS) toolkit, in two different national settings, France and Germany, and investigated children with three different home languages: Arabic, Portuguese and Turkish.
Methods & Procedures: NWR and SR tasks developed in parallel were administered to 151 bilingual children, aged 5;6–8;11, in France and in Germany, to 64 children in speech–language therapy (SLT) and to 87 children not in SLT, whose first language (L1) was Arabic, Portuguese or Turkish. Children were also administered standardized language tests in each of their languages to determine likely clinical status (typical development (TD) or SLI), and parents responded to a questionnaire including questions about early and current language use (bilingualism factors) and early language development (risk factors for SLI). Monolingual controls included 47 TD children and 29 children with SLI. Results were subjected to inter-group comparisons, to diagnostic accuracy calculation, and to correlation and multiple regression analyses.
Outcomes & Results: In accordance with previous studies, NWR and SR identified SLI in the monolingual children, yielding good to excellent diagnostic accuracy. Diagnostic accuracy in bilingual children was fair to good, generally distinguishing children likely to have SLI from children likely to have TD. Accuracy was necessarily linked to the determination of clinical status, which was based on standardized assessment in each of the child's languages. Positive early development, a composite risk factor for SLI, and not variables related to language exposure and use, generally emerged as the strongest predictor of performance on the two tasks, constituting additional, independent support for the efficacy of NWR and SR in identifying impairment in bilingual children.
Conclusions & Implications: NWR and SR tasks informed by linguistic theory are appropriate for use as part of the diagnostic process for identifying language impairment in bilingual children for whom the language of assessment is different from the home language, in diverse sociolinguistic contexts.

Keywords: specific language impairment, bilingualism, diagnostic accuracy, sentence repetition, non-word repetition.

What this paper adds

What is already known on the subject

One of the major clinical challenges associated with identifying language impairment in bilingual children is that, in many settings, language assessment can plausibly be accomplished only in the country language. A growing body

Address correspondence to: Laurice Tuller, UMR Inserm U930, Université François Rabelais de Tours, Tours, France; e-mail: tuller@univ-tours.fr.

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online © 2018 Royal College of Speech and Language Therapists
DOI: 10.1111/1460-6984.12397

of research has pointed to tools that might be useful in this situation, including parental questionnaires and specific types of language tasks.

What this paper adds to existing knowledge
This study reports on a direct comparison of use of the diagnostic tools with the same design in bilingual children in two countries, with varying sociolinguistic settings (Arabic/Portuguese/Turkish communities in France and Germany). Bilingual children recruited both in SLT and not in SLT were assigned to clinical groups based on scores in both L1 and in L2, and then compared for their performance on linguistically informed NWR and SR tasks. Diagnostic accuracy was calculated and risk factors for language impairment versus bilingualism factors were compared to determine which best explained task performance. Results, but also method, were compared with those obtained in other studies of such tools.

What are the potential or actual clinical implications of this work?
The same kinds of language assessment tasks (NWR and SR) that have been found to yield the best diagnostic accuracy in monolingual children can profitably be used with bilingual children for assessment in their L2.

Introduction

With growing migration, more and more children in Europe speak at least two languages, their first language (L1), which they might use predominantly at home and in some other contexts, and the majority language (L2), which they use in (pre-)school. A study of bilingualism, a term used to refer to children who have acquired skills in more than one language, either since birth or later, has therefore become a focus of interest not only for linguists but also for educators and speech–language therapists. Several studies have shown that children growing up bilingually display specific difficulties and error patterns in the majority language (see Paradis 2010 for an overview) which may persist for periods of time that vary from child to child, and, at least for some children, may be quite long (Paradis *et al.* 2016, Tuller *et al.* 2015). Several factors influencing the rate of L2 acquisition and error types are currently under investigation: age of onset (AoO), length of exposure (LoE), quantity and quality of input (rich or reduced input), language status (high or low prestige), L1–L2 language typology, and parents' socioeconomic status (SES). The interaction of these factors makes it notoriously difficult to define what is typical for L2 development.

This is all the more so given that it has been reported that error patterns found in bilingual children, such as the omission of object clitics in French or the difficulties with verbal agreement in German (Paradis 2010, Hamann 2012), tend to overlap with error patterns that have been found to be diagnostic markers for the identification of what is variously referred to in the literature as primary language disorder, developmental language disorder or specific language impairment (SLI). The latter term is employed here in the interest of continuity with other work on otherwise unexplained language impairment, in this paper, in particular, in children growing up bilingually, referred to as Bi-SLI. A language disorder of this type is extremely detrimental to the individual's social and academic development and needs careful diagnosis in order to allow targeted language intervention. However, the language difficulties displayed by bilingual children demonstrably lead to both frequent over- and under-diagnosis, and, more generally, situations in which the nature of a child's language difficulties remains unclear for teachers, clinicians and parents (Grimm and Schulz 2014, Tuller *et al.* 2015). This fundamental problem has received increasing attention among researchers (for reviews, see Grimm and Schulz 2014, Armon-Lotem *et al.* 2015 and Marinis *et al.* 2017). The basic question is quite simply whether and how it can be determined that language difficulties in a bilingual child are due to SLI and not the reflex of a particular stage of typical L2 development.

Since SLI necessarily affects language development in all a child's languages, assessing each of them would appear to be the best solution and is indeed what is recommended by professional organizations such as the American Speech and Hearing Association (ASHA), the Royal College of Speech and Language Therapists (RCSLT) and many research groups (see the evaluation of procedures in De Lamo White and Jin 2011 and Thordardottir 2015). Testing the L1, however, may not be feasible, as standardized tests are often not available or cannot be administered/interpreted by a clinician who does not speak the language and may not have access to an interpreter or cultural advisor (see also Boerma *et al.* 2017). Even where possible, L1 testing may be unreliable because of potential L1 attrition (e.g., Montrul 2008) and because the L1 variety spoken outside the country of origin may have undergone linguistic change due to contact phenomena compared with the variety tested by the standardized L1 tools (e.g., Chilla and Şan 2017).

A battery of comparable tools was developed for cross-linguistic use during COST Action IS0804 under the name of Language Impairment Testing in Multilingual Settings (LITMUS; Armon-Lotem *et al.* 2015). To collect background information on bilingualism factors, but also risk factors for SLI, a parental questionnaire, the Parents of Bilingual Children Questionnaire (PABIQ; Tuller 2015), was developed. To investigate language abilities, we used non-word-repetition tasks (LITMUS-NWR; Chiat 2015) and sentence-repetition tasks (LITMUS-SR; Marinis and Armon-Lotem 2015) which were specifically designed to target phonological and syntactic structures that have been identified as problematic for children with SLI cross-linguistically. One reason these particular types of tasks were chosen as a focus here was because each has been demonstrated to be the most discriminatory for identifying SLI in monolingual children (Conti-Ramsden *et al.* 2001), and, in bilingual populations, evidence has been provided that both may be less sensitive to previous exposure than other language measures such as receptive vocabulary (Thordardottir and Brandeker 2013).

Results on the diagnostic accuracy of the LITMUS repetition tasks have appeared (e.g., de Almeida *et al.* 2017, Armon-Lotem and Meir 2016, Boerma *et al.* 2015, Chiat and Polišenská 2016, Hamann and Abed Ibrahim 2017, Marinis *et al.* 2017), and are encouraging in that they have shown fair to excellent diagnostic accuracy. Still lacking, however, are large-scale studies using the same methodology in different bilingual contexts, with participant samples displaying the wide range of diversity that is common in most clinical settings. Moreover, existing studies vary in terms of how clinical status of the bilingual participants was established and how participants were recruited, leading to questions about interpretation of reported diagnostic accuracy. As outlined in the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) initiative (Bossuyt *et al.* 2015), diagnostic accuracy compares a diagnostic tool with a *reference standard*, 'the best available method for establishing the presence or absence of the target condition' (Bossuyt et al. 2015:3). For bilingual children there is no agreed-upon reference standard for detecting language impairment (Thordardottir 2015); indeed, this is the basic research problem. Some studies have established clinical status exclusively on the basis of L2 testing (e.g., Boerma *et al.* 2015), others have combined L2 and L1 assessments, sometimes along with early developmental history information (e.g., Armon-Lotem and Meir 2016). STARD has warned against potential sources of bias in diagnostic accuracy studies, including spectrum bias (when subjects included do not reflect the complete spectrum of the targeted populations), which can lead to overestimation of test accuracy. In the case of language impairment in bilingual children, it is particularly important that sampling methods do not in some way exclude children whose diagnosis is the most uncertain. Comparison of reference standard methods and, more generally, participant characteristics is wanting.

We explicitly addressed these issues in the Franco-German project reported here. In this paper, we focus on whether and how well similarly constructed, linguistically targeted repetition tools, LITMUS-NWR and LITMUS-SR, can identify SLI in 5–8-year-old bilingual children having one of the three home languages (Arabic, Portuguese or Turkish), but growing up in different sociocultural and L2 settings, in France and in Germany, and thus whether the French and German versions of these tasks have adequate diagnostic accuracy. Furthermore, we wanted to know how well these repetition tasks fared with children whose exposure and use of the L2 varied. Finally, we wanted to see if performance on the two repetition tasks was related to risk factors for SLI, as this would constitute a further indication that these tools could help in detecting language impairment in bilingual children.

Materials and methods

This section presents the materials and procedures used, and the bi- and monolingual participants tested. The likely clinical status of the bilingual participants was the result of a procedure based on assessment in both of their languages in conjunction with an estimate of language dominance derived from the PABIQ. We thus first present the referenced tools chosen for each language, then the three LITMUS tools used, NWR, SR and PABIQ, and, finally, the participants and their clinical status.

Materials

Standardized L1 and L2 tests

Standardized L1 and L2 tests were selected based on the availability of appropriate norms, frequency of use in speech–language therapy (SLT) and language areas tested. With the exception of Turkish, the following five domains were evaluated in each language: phonology, receptive and expressive vocabulary, and morphosyntax in production and in comprehension. Details on the subtests used to assess each of the different areas in each of the languages are provided in appendix A. For French, we selected subtests from the BILO and N-EEL. For German, we used the LiSe-DaZ, the only test in our protocol having bilingual norms, along with the PLAKSS-II and WWT. For Portuguese, we used the PALPA-P and the GOL-E. For Arabic, we used the ELO-L, standardized and normed in Lebanon, but

Table 1. LITMUS-NWR: content of language independent and dependent items

	Vowels	Consonants	Syllable types	Examples
Language independent: 30 items	/a, i, u/	/p, k, f, l/	CV CCV CVC#	faku klipafu, fupla fuk, kafip
Language dependent—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—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

translated (by native speakers) into other varieties of Arabic for this study, with pre-recorded oral stimuli for all versions. For Turkish, we used the TEDIL, which only has composite norms, for reception and for expression (and which does not include a phonological subtest).

The LITMUS tools

The experimental protocol consisted of three LITMUS tools: LITMUS-NWR, an NWR task, LITMUS-SR, an SR task, and LITMUS-PABIQ, a parental questionnaire for bilingual families. The parental questionnaire used in the two countries was identical. The French and German versions of LITMUS-NWR and LITMUS-SR contain common cores, but differ in terms of the number of linguistic features targeted, in order to take into account those that have been shown to be difficult for children with SLI in each of the languages, whose phonology and syntax differ.

Non-word-repetition tasks. Non-word-repetition tasks have a long history of being employed as measures of working memory (Archibald and Gathercole 2007) and have been demonstrated to reliably identify SLI in monolingual children (Conti-Ramsden *et al.* 2001). More recently it has been shown that it is not only working memory, measured by increasing numbers of syllables, but also phonological complexity which is difficult for children with SLI (Gallon *et al.* 2007, Ferré *et al.* 2012, dos Santos and Ferré 2018). Building on these results, new NWR tasks have been constructed in order to tap more directly into phonological competence (Chiat 2015, dos Santos and Ferré 2018, Grimm and Hübner in press). At the same time, these tasks were constructed to avoid phonological properties of the L2 that could pose problems in bilingual contexts by using non-words which rely on phonological properties that are largely universal. Such quasi-universal non-words should not be harder for bilingual children compared with monolingual children. They should therefore allow for reliable assessment of phonology in bilingual children, even after only a short LoE to the L2 (Chiat and Polišenská 2016).

The LITMUS-NWR tasks for French and German were designed according to these principles (dos Santos and Ferré 2018, Grimm and Hübner in press). They contain the same language independent (LI) non-words, as well as language dependent (LD) non-words for French and for German respectively. All the non-words are either mono-, bi- or trisyllabic, in order to minimize memory effects. The 30 language independent non-words are made up of cross-linguistically frequent phonemes and respect cross-linguistically frequent phonotactic properties (Maddieson *et al.* 2011). Besides simple CV (consonant–vowel) syllables, there are also syllables having initial consonant clusters (branching onsets, CCV) and a final consonant (-CVC#), since these exist in many languages and should therefore be unproblematic for (most) typical bilinguals, but difficult for children with SLI. The LD items, 41 in the French version and 36 in the German version, include one or two additional consonants, and several more complex syllable types (table 1),[1] to follow the available results on specific phonological difficulties in SLI in each of these languages.

The task was administered in a child-friendly, prerecorded, computerized format. Administration time generally did not exceed 5 min. Children's responses were audio-recorded for later transcription. The task was scored by whole-item accuracy; voicing substitution on consonants and substitution of a vowel minimally different from the target vowel were not counted as errors (dos Santos and Ferré 2018).

Sentence repetition. Sentence repetition has been shown to be particularly reliable for identifying SLI in monolingual children (e.g., Conti-Ramsden *et al.* 2001). It has been argued to be a sensitive indicator of children's morphosyntactic abilities since repeating a sentence involves both sentence processing and sentence reconstruction (Marinis and Armon-Lotem 2015, Polišenská *et al.* 2015). The LITMUS-SR tasks were designed to include complex constructions known to cause difficulties for children with SLI in many languages, including French and German (Jakubowicz and Tuller 2008, Hamann *et al.* 1998), as argued explicitly by Marinis and Armon-Lotem (2015) and Fleckstein *et al.* (2018). These include constructions displaying a word order different from the typical word order found in the language, such as object wh-questions involving

Table 2. 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	Who-question	*Wen umarmt der Pinguin heute?* Who.ACC hugs the.NOM penguin today 'Whom does the penguin hug today?'
	Which N-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'

movement of the interrogative pronoun to the front of the clause, constructions with a subordinate clause, such as clausal complements (both infinitival and finite, the latter arguably more complex), and relative clauses, constructions having both embedding and movement (both subject and object, the latter arguably more complex). Another known difficulty in SLI, in both languages, was also targeted: verbal morphology. LITMUS-SR-French and LITMUS-SR-German contain a core of similar constructions, involving the same syntactic variables (presence of a subordinate clause or not, type of subordinate clause, wh-movement of the object etc.) presented in table 2 (with alternating example sentences drawn from the two tests).[2]

There are 30 items in LITMUS-SR-French and 45 in LITMUS-SR-German. The tasks are presented in a computerized version with pre-recorded stimuli, and take 5–10 min to administer. The child's responses were recorded, transcribed, verified and analyzed with different scoring methods, including *identical repetition*, which requires verbatim repetition of the stimulus, disregarding only phonological errors, and which yields a score of 0 (the child's production contained at least one repetition error) or 1 (no error occurred). We focus on this score here, as it constitutes the simplest and quickest coding procedure for clinicians.[3]

Interpretation of language data from bilinguals requires taking into account the individual's previous and current language exposure and use. Detection of language impairment typically makes use of information about early language development and family history for language difficulties. For both of these purposes, *parental questionnaires* have been developed and have proved to be reliable tools not only for assessing risk factors but also for establishing age of first systematic language exposure (AoO), LoE, and quantity and quality of input at home and in other settings. We used the PABIQ (Tuller 2015), which was based on questionnaires created by Paradis *et al.* (2010) and Paradis (2011).

As laid out in table 3, an index of the child's exposure to each of his/her languages was calculated from information on AoO, LoE, language use and richness before and after the age of 4, at home, at school and in extracurricular activities, each of which was attributed points. Therefore, for example, for AoO, the maximum of 4 points was allotted for exposure beginning at birth (0 months), 0 points were allotted for an onset at 96 months or more, and all 12-month intermediate intervals changed by 0.5 point (AoO of 12 months = 3.5 points, of 24 months = 3 points etc.). The difference between the L2 Exposure Index and the L1 Exposure Index yielded an estimate of L2 Language Dominance.[4]

In addition, a rate of Early Language Exposure, for the L1 and for the L2, was calculated based on which language was used over a maximum total of eight different contexts (mother, father, grandparent, babysitter,

Table 3. Calculation of language exposure indices and the L2 dominance index

	L1 (Arabic/Portuguese/Turkish)	L2 (French/German)
AoO	/4	/4
Frequency of exposure > age 4	/4	/4
Contexts of exposure > age 4	/8	/8
LoE	/4	/4
Current use in the family	/16	/16
Current L1/L2 Richness (other current-use activities, friends)	/14	/14
Language exposure index	/50 maximum points	/50 maximum points
L2 dominance index (–50, . . . , 50)	(Language exposure index for L2) – (language exposure index for L1)	

other adult, siblings, nursery school/daycare centre and kindergarten) relevant for each child before age 4.

Information related to risk factors for SLI obtained via the PABIQ included age of first word and first sentence, early parental language concerns (yes/no), and language difficulties in immediate family members (yes/no for mother, father and siblings, for receptive/expressive/written language difficulties). This information was synthesized into a Positive Early Development Index (/14, with 6 maximum points for age of first word, 6 maximum points for age of first sentence and 2 points for parental concerns about language) and a global No Risk Index (/23) (Positive Early Development /14 added to the score /9 for absence of language difficulties in the family) (Tuller 2015, de Almeida *et al.* 2017).

Participants

The children investigated were L1 Arabic, Portuguese or Turkish and L2 French or German. These particular L1 languages were chosen because they are significant languages of immigration in France and in Germany, either deep-rooted or more recent, and thus correspond to clinical reality in these countries. Moreover, they differ in linguistic typological proximity to the L2. Finally, these language communities differ in the two countries in terms of sociocultural factors such as L1 transmission and community cohesion (Condon and Régnard 2010, Tucci and Groh-Samberg 2008).

Recruitment of children

In order to increase chances of finding bilingual participants both with and without language impairment, we recruited children in ordinary schools, through community associations, clubs and other gathering places (e.g., places of worship), but also in language diagnostic centres and in speech–language pathology (SLP) private practice. The recruitment criteria were based on age (5;6–8;11),[5] and having either Arabic, Portuguese or Turkish spoken at home. Children who scored below percentile 9 on Raven's Progressive Matrices (the cut-off for 'low average' non-verbal intelligence) were excluded in the absence of any other available normal non-verbal score. 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 L2 LoE of 12 months. A total of 151 bilingual children were retained for the study, 95 in France (47 not in SLT and 48 in SLT) and 56 in Germany (40 not in SLT and 16 in SLT) (table 4). Note that the number of participants in the two countries was not the same, and furthermore that while equal numbers were recruited in SLT and not in SLT in France, in Germany there were more than twice as many children not in SLT as in SLT. We return to these differences below.

Table 4. Bilingual children recruited in France and in Germany, in SLT and not in SLT

	France		Germany	
	Not in SLT	In SLT	Not in SLT	In SLT
L1 Arabic	20	16	7	4
L1 Portuguese	14	12	21	1
L1 Turkish	13	20	12	11
Total number of participants	47	48	40	16
Age (months), mean (SD), range, gender	82.3 (11.6) 64–107 20 F, 27 M	85.6 (13.5) 64–106 16 F, 32 M	83.4 (14.4) 61–119 22 F, 18 M	84.6 (14.9) 64–108 6 F, 10 M

Assignment of children to Bi-TD and Bi-SLI groups

In order to establish a benchmark against which we could evaluate the usefulness of the experimental tasks, a procedure was applied to estimate the clinical status of children, those most likely to be typically developing bilingual children (Bi-TD) and those most likely to be bilingual children with SLI (Bi-SLI). The point of departure for the procedure used was the recommendation emanating from the COST Action IS0804 assessment committee, as presented by Thordardottir (2015). This recommendation followed Tomblin *et al.* (1996) in considering that SLI can be identified in monolingual children when their language performance is below norms in two different language areas. Thordardottir (2015)

Table 5. Bilingual participants in France and in Germany: Bi-TD and Bi-SLI

	France		Germany	
	Bi-TD	Bi-SLI	Bi-TD	Bi-SLI
L1 Arabic	27	9	8	3
L1 Portuguese	18	8	21	1
L1 Turkish	24	9	19	4
Total number of participants	69	26	48	8
Age (months),	84.0 (12.5)	83.7 (13.3)	84.2 (14.2)	81.0 (16.7)
mean (SD), range,	64–106	66–107	61–119	64–108
gender	28 F, 41 M	8 F, 18 M	26 F, 22 M	2 F, 6 M

Table 6. Monolingual participants in France and in Germany: Mo-TD and Mo-SLI

	France		Germany	
	Mo-TD	Mo-SLI	Mo-TD	Mo-SLI
Total number of participants	37	17	10	12
Age (months),	84.3 (10.7)	91.2 (8.6)	75.9 (9.0)	81.8 (13.4)
mean (SD), range,	67–101	75–104	66–92	68–112
gender	14 F, 23 M	11 F, 6 M	8 F, 2 M	5 F, 7 M

suggested adjusting Tomblin *et al.*'s (1996) –1.25 SD (standard deviation) cut-off according to the status of the language being tested. Adopting this system, we set cut-offs at –1.5 SD if the language tested was the dominant language of the child, at –2.25 SD if it was the weaker language, and at –1.75 SD if the language was one of equally dominant languages, as proposed by Thordardottir (2015).[6] Since children in our study were tested in both their languages, only those with scores below language-dominance-adjusted cut-offs in two language domains *in each language* were assigned to the Bi-SLI group. The language domains in question were phonology, receptive and expressive vocabulary, and morphosyntax in production and in comprehension (see the Materials and methods above and appendix A); however, the lexicon was counted as a single domain, and was not considered to be impaired if only expressive vocabulary was below the appropriate cut-off.[7]

This procedure gave rise to the groups presented in table 5. Comparison of table 5 with table 4 reveals that roughly half the children recruited in SLT in each country were assigned to the Bi-TD group; only one child not in SLT was assigned to the Bi-SLI group. Note that the proportion of Bi-SLI children in the bilingual children in France (27%) is twice that of children in Germany (14%) (though these proportions are not statistically different: χ^2 (1, $N = 151$) = 2.75, $p = .097$, and reflect the differences in the proportion of children recruited in SLT and not in SLT in the two country samples shown in table 4.

We also administered these tasks to 71 monolingual children in France and in Germany, 29 with SLI (Mo-SLI) and 47 with TD (Mo-TD) (table 6). Note that monolingual children in Germany were a bit younger than those in France, both Mo-TD ($W = 100$, $p = .028$) and Mo-SLI ($W = 54$, $p = .035$).[8]

Data analysis

Children's NWR and SR responses were recorded with digital audio recorders, and then transcribed, verified and scored by at least two independent, trained research assistants, as well as by one or more of the project linguists. Statistical analyses were conducted using a variety of tests. *T*-tests were used for inter-country comparisons of the entire bilingual samples, and for some French intra-country comparisons. Since normality of distribution (as per the Shapiro–Wilk test) and homogeneity of variances (as per the Levene's test) could not be established for certain measures, non-parametric statistical tests were run as well (Kruskal–Wallis (K-W), Mann–Whitney (MWW) or Wilcoxon (W)). Since these resulted in the same conclusions, we report *t*-tests throughout. Non-parametric tests are reported for all comparisons involving small sample sizes (e.g., Germany intra-country comparisons).

Results

One goal of the participant recruitment was to include the complete spectrum of the targeted populations, including, for example, bilingual children for whom SES and bilingualism measures varied. We report first therefore on the background measures in our bilingual samples. We then directly address whether French and German LITMUS-NWR and LITMUS-SR show adequate diagnostic accuracy in the bilingual samples in each of the countries. Finally, we report on whether performance on these tasks is related to factors related to bilingualism or to risk factors for SLI.

Background measures

We begin by comparing the general characteristics of the entire bilingual sample, and thus all the children, regardless of clinical status, in the two countries, in terms of non-language variables (age and SES), bilingualism measures (AoO and LoE of the L2, early L1 and L2 exposure, current L1 and L2 richness, and, more globally, degree of dominance for the L2), risk factors for SLI, and performance on the L1 standardized language tasks. We sought to establish the degree of diversity within the two country samples of bilingual children of the same age, diversity which we believe is characteristic of the population diversity and thus of the diversity observed

Table 7. Age and SES (mother's education): mean (SD)

	France			Germany		
	Bi- ($n = 95$)	Bi-TD ($n = 69$)	Bi-SLI ($n = 26$)	Bi- ($n = 56$)	Bi-TD ($n = 48$)	Bi-SLI ($n = 8$)
Age (months)	83.94	84.01	83.73	83.75	84.21	81.00
	(12.66)	(12.50)	(13.33)	(14.44)	(14.18)	(16.66)
SES (mother's education)	10.45	10.73	11.27	12.78	12.84	12.50
(years)	(3.60)	(9.72)	(9.96)	(4.09)	(4.28)	(2.98)

Note: SES, socioeconomic status.

Table 8. Bilingualism measures: mean (SD)

	France			Germany		
	Bi- ($n = 95$)	Bi-TD ($n = 69$)	Bi-SLI ($n = 26$)	Bi- ($n = 56$)	Bi-TD ($n = 48$)	Bi-SLI ($n = 8$)
L2 AoO (months)	16.36	16.75	15.31	21.02	21.27	19.50
	(23.15)	(22.41)	(25.46)	(19.08)	(19.62)	(16.59)
L2 LoE (months)	67.58	67.26	68.42	63.64	63.69	63.38
	(21.88)	(22.85)	(19.47)	(22.65)	(23.03)	(21.71)
Early L1 Exposure (%)	71.45	73.06	67.20	71.59	71.75	70.69
	(17.44)	(14.87)	(22.72)	(20.73)	(21.94)	(11.97)
Early L2 Exposure (%)	58.96	58.46	60.27	50.33	46.91	70.91
	(50.33)	(25.35)	(26.39)	(26.27)	(26.59)	(10.68)
Current L1 Richness (/14)	5.85	6.43	4.31	4.09	4.10	4.00
	(3.07)	(2.72)	(3.44)	(1.76)	(1.9)	(0.0)
Current L2 Richness (/14)	9.19	8.88	10.00	9.59	9.69	9.00
	(2.83)	(2.87)	(2.59)	(1.88)	(2.01)	(0.00)
L2 Dominance (–50, . . . , 50)	–1.83	–3.30	2.06	–1.58	–2.00	0.94
	(12.90)	(11.92)	(14.74)	(12.52)	(12.47)	(13.40)

Table 9. Language dominance index (–50, . . . , 50): mean (SD) and number of participants

	France			Germany		
	Bi- ($n = 95$)	Bi-TD ($n = 69$)	Bi-SLI ($n = 26$)	Bi- ($n = 56$)	Bi-TD ($n = 48$)	Bi-SLI ($n = 8$)
Arabic	4.21 (13.72)	0.26 (13.29)	16.06 (6.27)	0.36 (12.37)	–3.13 (11.12)	9.67 (12.42)
	$n = 36$	$n = 27$	$n = 9$	$n = 11$	$n = 8$	$n = 3$
Portuguese	–1.42 (12.61)	–1.11 (11.34)	–2.12 (15.98)	0.68 (12.97)	0.74 (13.29)	–0.50
	$n = 26$	$n = 18$	$n = 8$	$n = 22$	$n = 21$	$n = 1$
Turkish	–8.74 (8.25)	–8.94 (8.60)	–8.22 (7.69)	–4.67 (12.04)	–4.55 (12.07)	–5.25 (13.72)
	$n = 33$	$n = 24$	$n = 9$	$n = 23$	$n = 19$	$n = 4$

in clinical settings. Diagnostic tools clearly must remain operative in these settings.

Comparing bilingual children in France and in Germany for non-language variables (table 7), no significant age differences were observed ($t(149) = .082$ $p = .934$). SES (years of the mother's education) did differ in the two countries, mothers in France having had fewer years of education than mothers in Germany ($t(141) = 3.570$, $p < .001$).[9]

As for bilingualism measures (table 8), the two country samples (again, including all the bilingual children in each of the countries regardless of clinical status) were comparable for AoO ($t(149) = 1.27$, $p = .205$), LoE ($t(149) = –1.05$, $p = .294$), Early L1 Exposure ($t(149) = –.044$, $p = .965$), and Current L2 Richness ($t(149) = 1.04$, $p = .299$). However, for the bilingual children in France, more early language contexts were in French than was the case for L2 contexts for the bilingual children in Germany (Early L2 Exposure: $t(149) = 1.98$, $p = .049$), and at the same time the French bilinguals tended to have a greater variety of current L1 contexts (Current L1 Richness: ($t(149) = –4.49$, $p > .001$). Focusing on the Bi-TD and Bi-SLI subgroups in the bilingual samples from each of the countries, the Bi-SLI children in Germany had higher rates of Early L2 Exposure compared with their Bi-TD counterparts (MWW: $W(56) = 85$, $p = .012$); no such difference was found in France ($t(93) = .308$, $p = .759$).[10]

Degree of L2 dominance varied according to L1 group (table 9). In France, this difference was significant (K-W: χ^2 (2, $N = 95$) = 16.60, $p < .001$). The L1 Portuguese children tended to be more balanced

Table 10. Risk Factors for SLI: mean (SD)

	France			Germany		
	Bi- (n = 95)	Bi-TD (n = 69)	Bi-SLI (n = 26)	Bi- (n = 56)	Bi-TD (n = 48)	Bi-SLI (n = 8)
No risk index (/23)	18.73 (4.96)	19.96 (4.20)	15.46 (5.43)	19.55 (4.52)	20.54 (3.63)	13.62 (5.07)
Positive Early Development (/14)	10.57 (4.40)	11.59 (3.62)	7.85 (5.16)	11.00 (4.53)	12.00 (3.57)	5.00 (5.24)

Table 11. Diagnostic accuracy of NWR and SR in monolingual children

	NWR		SR	
	France	Germany	France	Germany
AUC[a]	0.966	0.954	0.981	1.000
Cut-off (%)	77.5	59.9	78.3	63.3
Sensitivity (%)	88	92	93	100
Specificity (%)	92	90	92	100
LR+[b]	10.63	9.17	11.52	–
LR–[c]	.13	.09	.07	.00

Notes: [a]AUC values indicate whether accuracy is excellent (.9–1.0), good (.80–.90), fair (.70–.80), poor (.60–.70) or worthless (.50–.60) (Swets *et al.* 2000).
[b]LR+ values of 1–2 indicate a minimal increase in likelihood that a score below the cut-off corresponds to language impairment, 2–5 a small increase, 5–10 a moderate increase and > 10 a large increase in likelihood (Grimes and Schultz 2005).
[c]LR– values of 0.5–1.0 indicate a minimal decrease in the likelihood that a score over the cut-off corresponds to typical language development, 0.2–0.5 a small increase, 0.1–0.2 a moderate decrease and < 0.1 a large decrease in that likelihood (Grimes and Schultz 2005).

bilinguals and the L1 Arabic children to be either 'balanced' (Bi-TD) or clearly L2 dominant (Bi-SLI), whereas the L1 Turkish children were largely L1 dominant, significantly more so than the L1 Arabic children (p < .001). In Germany, overall, children tended to be less L1 dominant than the French bilinguals, and did not differ for dominance according to L1 group (K-W: χ^2 (2, N = 56) = 1.84, p = .398). This variety in L2 dominance reflects the diversity common in clinical settings in Europe today, supporting the clinical relevance of the participant samples in the two countries.

No evidence was found for between-country differences regarding risk factors for SLI (table 10), as measured by the global No Risk Index (t(149) = 1.05, p = .297) or by the Positive Early Development Index (t(149) = –.58, p = .566). As expected, differences between Bi-SLI and Bi-TD groups were significant in each country. This was true for the No Risk Index (France: t(93) = 3.82, p = .001; Germany: W(56) = 336, p < .001) and for Positive Early Development (France: t(93) = 3.41, p = .002; Germany: W(56) = 328, p < .001).

LITMUS-NWR and LITMUS-SR: performance and diagnostic accuracy

Using the most basic measures for NWR (global % correct) and SR (% identical repetition), diagnostic accuracy in monolingual children in France and in Germany was excellent (table 11). Optimal cut-offs for each of the two tests were lower for the German sample (NWR, 60% and SR, 63%) compared with the French sample (NWR, 78% and SR, 78%); recall that monolingual children in Germany are somewhat younger.

Focusing henceforth on the bilingual samples, performance on both NWR (figure 1) and SR (figure 2) was significantly different in the Bi-TD and Bi-SLI groups in France (NWR: t(92) = 5.81, p < .001 and SR: t(93) = 4.85, p < .001) and in Germany (NWR: W(56) = 359.5, p < .001 and SR: W(56) = 344.5 p < .001).

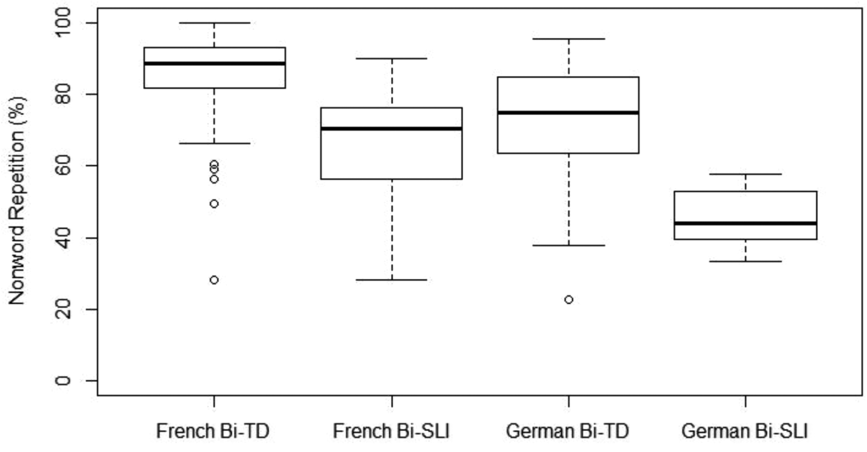

Figure 1. NWR-French and NWR-German: Bi-SLI versus Bi-TD.

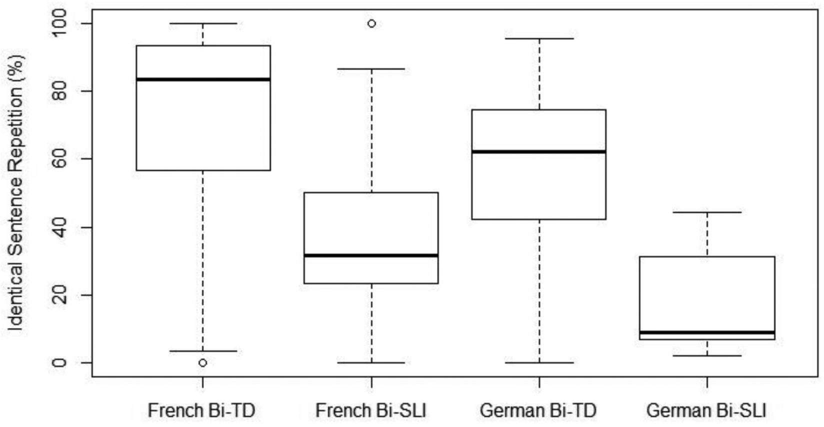

Figure 2. SR-French and SR-German: Bi-SLI versus Bi-TD.

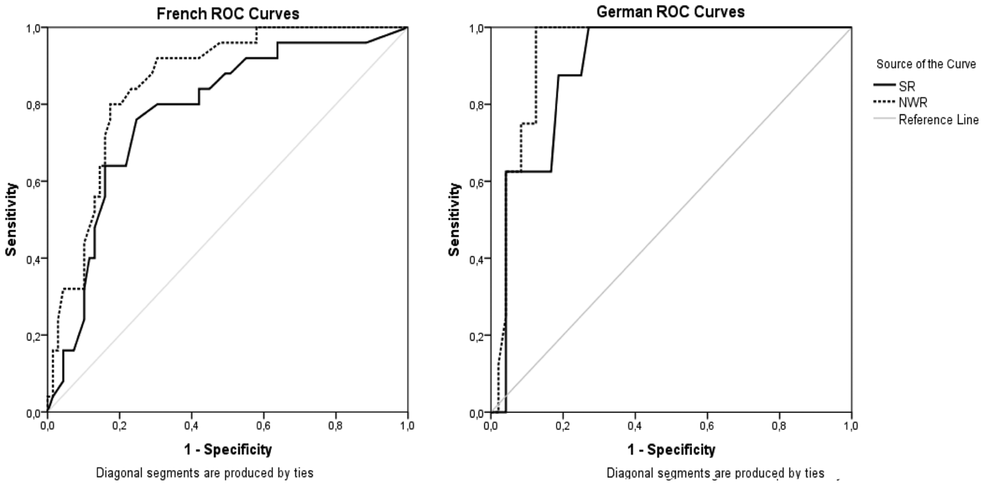

Figure 3. ROC curves for NWR and SR for LITMUS-French (left) and LITMUS-German (right).

Table 12. Diagnostic accuracy of NWR and SR in bilingual children

	NWR		SR	
	France	Germany	France	Germany
AUC[a]	.856	.936	.782	.897
Cut-off (%)	79.6	58.3	53.3	32.2
Sensitivity (%)	84	100	77	88
Specificity (%)	77	88	76	81
LR+[b]	4.80	–	3.27	4.67
LR–	0.28	0.00	0.32	0.15

Notes: [a]AUC values indicate whether accuracy is excellent (.90–1.0), good (.80–.90), fair (.70–.80), poor (.60–.70) or worthless (.50–.60) (Swets *et al.* 2000).

[b]LR+ values of 1–2 indicate a minimal increase in likelihood that a score below the cut-off corresponds to language impairment, 2–5 a small increase, 5–10 a moderate increase and > 10 a large increase in likelihood (Grimes and Schultz 2005).

[c]LR– values of 0.5–1.0 indicate a minimal decrease in the likelihood that a score over the cut-off corresponds to typical language development, 0.2–0.5 a small increase, 0.1–0.2 a moderate decrease and < 0.1 a large decrease in that likelihood (Grimes and Schultz 2005).

Receiver Operating Characteristic (ROC) analyses conducted on the French and German bilingual samples are reported in figure 3 and table 12. As was the case for the monolingual samples, bilingual optimal performance cut-offs were lower for the German tests (NWR = 58%, SR = 32%) compared with the French tests (NWR = 80%, SR = 53%). Diagnostic accuracy for NWR yields satisfactory results in both countries. SR-French only approaches minimal levels of accuracy, whereas SR-German meets these levels.

Predicting bilingual children's performance on LITMUS repetition tasks

We have seen that diagnostic accuracy of the LITMUS-French repetition tasks is fair (SR) or good (NWR). Diagnostic accuracy of the LITMUS-German repetition tasks is good (SR) or excellent (NWR), although cut-offs are lower and the German Bi-SLI sample is quite limited. Measures of diagnostic accuracy are very sensitive to the characteristics of the population in which the test accuracy is evaluated. In this study, diagnostic accuracy was affected by how well the bilingual children had been assigned to the Bi-TD and Bi-SLI groups. As the assignment procedure depended on results on standardized tests assessing both children's L1 and their

Table 13. Partial Pearson correlation, controlling for age and SES

	% NWR (global score)		% SR (identical repitition)	
	Germany	France	Germany	France
Positive Early Development	.493	.512	.523	.485
	.000	.000	.000	.000
Early L2 Exposure	–.398	.086	–.220	.095
	.004	.424	.124	.378
Current L2 Richness	.189	.030	.266	.111
	.189	.783	.061	.303

L2, the problems associated with using such tests on bilingual children are necessarily inherent in that procedure. Since therefore some children may not have been assigned to the correct group, we sought to confirm/strengthen the conclusions drawn from our analysis of diagnostic accuracy. Another way of exploring the question of the clinical usefulness of these tasks, which were designed to target linguistic properties, is to explore which variables predict performance, in each of the countries. If these tasks really could be reliable clinical tools, then they should be more strongly related to SLI risk factors and less strongly related to bilingualism factors. We therefore examined correlation analyses and then regression analyses on the relevant measures.

Examination of correlation analyses conducted on SR (% identical repetition) and on NWR (% global repetition) in each country sample (see Appendix B) led to selection of the following five variables as candidates for explaining performance in the bilingual children: Positive Early Development, Age, SES (mother's education), Early L2 Exposure and Current L2 Richness. Notably, no correlations were observed between global identical repetition on SR or on NWR and L2 AoO, L2 LoE, or Early or Current L1 Exposure or Richness. Mother's education was retained for SES since a father's education displayed weaker correlations (and was strongly correlated to a mother's education). The No Risk Index was dropped in favour of the Positive Early Development measure, to avoid circularity as the latter is a component part of the former, and because it led to stronger correlations. The same reasoning led us to abandon the Language Dominance Index in favour of two of its component parts, Early L2 Exposure and Current L2 Richness.

The final research question was to investigate whether performance on the two repetition tasks was primarily related to risk factors for SLI rather than to bilingualism factors, as this would constitute evidence that these tools could be of use in detecting language impairment in bilingual children. We therefore conducted partial correlational analyses to isolate the SLI risk factor and bilingualism variables, the locus of our research question, and therefore controlling for age and SES. These analyses revealed that the strongest correlations, for both SR and NWR, in each country, were with Positive Early Development, the SLI risk factor variable (table 13). A significant (negative) correlation was also found between Early L2 Exposure and NWR-German.

Pursuing this result by seeking to determine the relative contribution of Positive Early Development versus Early L2 Exposure and Current L2 Richness, a stepwise multiple regression analysis was conducted to predict performance on SR and NWR in each country sample (tables 14–17). Significant regression equations were found showing that Positive Early Development was the major/single predictor for NWR and for SR in both countries. For NWR-French (table 14), Positive Early Development was the sole predictor, and it explained 24% of the variance, excluding the L2 variables. For NWR-German (table 15), Positive Early Development explained 14% of the variance, with Early L2 Exposure

Table 14. Stepwise multiple regression: NWR-French (dependent variable), and Positive Early Development, Early L2 Exposure and Current L2 Richness (independent variables)

NWR-French ($N = 94$)

Variables entered[a]	R^2	ΔR^2	R^2_{adj}	F (d.f.)	p	β	t	p
1 Positive Early Dev.	.243	–	.234	29.47 (1, 92)	< .001	.493	5.43	< .001

Note: [a]Excluded variables in final model: Early L2 Exposure, Current L2 Richness.

Table 15. Stepwise multiple regression: NWR-German (dependent variable), and Positive Early Development, Early L2 Exposure and Current L2 Richness (independent variables)

NWR-German ($N = 56$)

Variables entered[a]	R^2	ΔR^2	R^2_{adj}	F (d.f.)	p	β	t	p
1 Positive Early Development	.137	–	.121	8.60 (1, 54)	.005	.319	2.566	.013
2 Early L2 Exposure	.211	.074	.181	7.07 (2, 53)	.002	–.273	–2.217	.031

Note: [a]Excluded variable in final model: Current L2 Richness.

Table 16. Stepwise multiple regression: SR-French (dependent variable), and Positive Early Development, Early L2 Exposure and Current L2 Richness (independent variables)

SR-French ($N = 95$)								
Variables entered[a]	R^2	ΔR^2	R^2_{adj}	F (d.f.)	p	β	t	p
1 Positive Early Development	.192	–	.184	22.15 (1, 93)	< .001	.460	5.004	< .001
2 Current L2 Richness	.231	.039	.214	13.79 (2, 92)	< .001	.197	2.137	.035

Note: [a]Excluded variable in final model: Early L2 Exposure.

Table 17. Stepwise multiple regression: SR-German (dependent variable), and Positive Early Development, Early L2 Exposure and Current L2 Richness (independent variables)

SR-German ($N = 56$)								
Variables entered[a]	R^2	ΔR^2	R^2_{adj}	F (d.f.)	p	β	t	p
1 Positive Early Devopment	.165	–	.150	10.69 (1, 54)	.002	.362	3.079	.003
2 Current L2 Richness	.279	.114	.252	10.25 (2, 53)	< .001	.340	2.889	.006

Note: [a]Excluded variable in final model: Early L2 Exposure.

weighing in negatively to account for an additional 7%. Positive Early Development explained 19% of the variance in SR-French and 17% for SR-German (tables 16 and 17), with Current L2 Richness adding 4% (France) and 11% (Germany) more.[11]

Discussion and conclusions

We administered two LITMUS repetition tasks incorporating structures of varying phonological and syntactic complexity to bi- and monolingual children in France and in Germany. The results are in line with expectations for monolingual children, confirming that these linguistic tasks discriminate impaired from typical language development. The predictions were also confirmed for these tasks in the bilingual populations in both countries, yielding fair to excellent diagnostic accuracy.

The research design of our study was predicated on the use of tools developed following the same principles, administered in the country language (the children's L2), in two countries with different linguistic, educational and cultural settings and in three different bilingual communities in these countries. We sought to maximize diversity in our population samples in order to reflect as faithfully as possible the population characteristics of bilingual communities in European countries today. The results showing heterogeneity both between the two countries and within each country for a range of variables (SES, quantity and quality of language exposure and use, language dominance) suggest that our samples did indeed include a broad spectrum of bilingual children. We also sought to maximize the variability inherent in SLI by trying to recruit as many bilingual children currently in SLT as not in SLT. The procedure for assignment of likely clinical status based on standardized evaluation in each of the child's languages, with cut-offs set according to language dominance, led to a number of children (about half) in SLT to be moved to the Bi-DT group. We note that this meant that the Bi-DT group therefore included several children for whom concerns about language difficulties had led to placement in SLT. This method most certainly increased the range of L2 difficulties observed in the Bi-DT group.

Direct France–Germany comparisons of the performance of bilingual children revealed several important points. We note first that the optimal cut-off score was different for the French and German LITMUS tasks. As neither of the two task versions were identical, this was not unexpected, and it furthermore is in line with other findings based on comparison of versions of LITMUS tasks (Armon-Lotem and Meir 2016). What is more noteworthy is that different cultural settings in France and Germany did not appear to influence task reliability. Rather, Positive Early Development was the leading predictor of performance in NWR (in line with the findings of Boerma and Blom 2017) and SR over factors related to bilingualism such as AoO, L2 LoE and Early or Current L2 Exposure or Richness, in both countries. This result, in particular, supports the validity of these tasks for identifying SLI in bilingual children. It is important, however, to emphasize that although these independent variables identified as predictors of performance on NWR and SR account for a sizable proportion of the variance, there are clearly other factors involved. Age and SES are two of them, as we noted, and although these two tasks differ from many repetition tests in minimizing memory effects, there clearly are remaining memory and other executive function factors (Armon-Lotem and Meir 2016, de Almeida *et al.* 2017). Other factors that would need to be considered in greater depth in a study specifically addressing explanation for performance on these tasks include the influence of L1 proximity to features of the L2, and sociolinguistic factors specific to different migrant communities in France and in Germany (attitudes about school, about extra-community contact etc.).

The final point we would like to address concerns interpretation of diagnostic accuracy studies of SLI in bilingual children. Accuracy rates depend to a large degree on which children are part of which group, the Bi-TD or the Bi-SLI. Furthermore, generalizing the results of an analysis of diagnostic accuracy depends on how well the population samples upon which this analysis has been carried out reflect the characteristics of the populations in question. These questions are at the heart of the challenge of identifying language impairment in bilingual children, and thus it is fundamental that studies clearly identify how population samples were obtained and distributed between Bi-TD and Bi-SLI (the reference standard method used). In fact, methods vary quite a lot. To illustrate, we compare our study with two recent studies also using LITMUS repetition tasks, and were thus based on similar principles as ours (Armon-Lotem and Meir 2016, Boerma *et al.* 2015). In the Armon-Lotem and Meir (2016) study, bi- and monolingual participants were matched for SES, which was high (mean for mother's education: 14 years), with little variability (SD = 3). The consecutive sampling did not include SES as a criterion, as we sought to maximize diversity; compared with the Armon-Lotem and Meir study, means were lower and there was much greater variability. The reference standard method used in Armon-Lotem and Meir and ours included standardized L1 and L2 language assessment, while the Boerma *et al.* (2015) study used only L2 assessment. Bilingual norms were used for these assessments in the Armon-Lotem and Meir study, but available only for part of the German assessment in our study. The Armon-Lotem and Meir method required children in the Bi-SLI group to be both below norms in each language *and* to have a parental/teacher report of history of SLI (late developmental milestones or language concerns), but not Boerma *et al.* or the present study. L1/sociocultural diversity was integrated into Boerma *et al.* (random) and into ours (controlled), but not in the Armon-Lotem and Meir study. Any of these differences could have affected diagnostic accuracy and most likely reduced spectrum breadth.

We believe that it is striking that, despite these differences, diagnostic accuracy results for LITMUS repetition tasks go essentially in the same direction. Armon-Lotem and Meir (2016) found bilingual diagnostic accuracy to be fair for LITMUS-NWR-Hebrew and good for LITMUS-SR-Hebrew. Boerma *et al.* (2015) found excellent accuracy for LITMUS-NWR-Dutch. The German and French LITMUS repetition tasks described here proved to have satisfactory diagnostic accuracy for the identification of SLI in bilingual children in a wide variety of settings. These tasks are moreover fast to administer, easily scored by the measure of identical repetition, and do not require the clinician to evaluate in a language (s)he does not know. In short, they appear to be very promising alternatives to testing the L2 with standardized tests normed on monolinguals. Clearly, the next step toward the use of these tests as a population screen would require norming over a population including random selection of an appropriately large number of participants, followed by a fresh diagnostic accuracy analysis.

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). ***Declaration of interest:*** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes

1. Since these properties are language dependent, this led to different numbers of items in French and German.
2. The German SR also contained constructions which are typical for German and have been shown to be milestones in the acquisition of German as L1 and L2. These are, in particular, the so-called sentence bracket, consisting of finite and non-finite verbal material bracketing other constituents, as in (i), and topicalization, as in (ii):

   (i) Hans hat dieses Buch gestern gelesen
   Hans has this book yesterday read
   (ii) Dieses Buch hat Hans gestern gelesen
   This book has Hans yesterday read.

   The German version also included coordination as a contrast to subordination, leading to a higher number of items than in the French SR task.
3. An alternative coding procedure was also used, *target structure*, for whether the sentence preserves the syntactic construction targeted, which disregards repetition errors that do not affect the structure of the stimulus sentences (e.g., lexical errors) (Fleckstein *et al.* 2018; Abed Ibrahim and Hamann 2017; Hamann and Abed Ibrahim 2017).
4. This calculation (notably the weight accorded to different variables) represents work in progress. A score for months of primary school (after age 6) in each language /5 (score range observed: 0–2.5) was added to the exposure indices in France, but not in Germany, where primary school starts anywhere from age 6–8.
5. Four German children over age 8;11 were also included: two bilingual children (one typical and one with SLI) aged 9;0, one typical bilingual child aged 9;11, and one monolingual child with SLI aged 9;4. The two children with SLI performed below cut-offs for younger children in all relevant tests. For the two typical bilingual children, cut-offs were adjusted following recommendations of test authors, e.g., –0.5 SD for the LiSE-DaZ (Grimm and Schulz 2014) and for the ELO-L (R. Zebib, personal communication).

6. These cut-offs were originally established for *simultaneous* bilinguals.
7. For the (German) LiSe-DaZ, which has bilingual norms for sequential bilingual children (AoO $>$ 2 years), cut-offs were set at –1.25 SD. For simultaneous bilinguals, we used mono- or bilingual norms according to a child's dominance as Schulz and Tracy (2011) recommend. The cut-off for evaluation of Turkish was set at –1.25 SD (in one or both of the two scores) if the child was Turkish dominant (–1.5 SD for balanced bilinguals, and –2 SD if Turkish was the weaker language), as the test authors set the monolingual cut-off at –1 SD.
8. Mo-SLI children were slightly older than Mo-TD children in France (MWW: $W = 437$, $p = 0.023$), but not in Germany (MWW: $W = 73.5$, $p = 0.390$).
9. SES measured by years of the mother's education is common in studies of child bilingualism (e.g., Paradis 2011). The fathers in France also had significantly fewer years of education than their German counterparts ($t(147) = -2.05$, $p = .042$). A father's education did not differ significantly from a mother's education in France (means = 10.91 and 10.45 respectively; $t(90) = -1.069$, $p = .288$) or in Germany (means = 12.78 and 12.63 years respectively; $t(50) = .263$, $p = .794$).
10. Non-parametric tests were used throughout when the German Bi-SLI group was targeted; these results should be interpreted with caution due to very small group size. Numbers of children in each group are provided in each table.
11. A regression analysis enlarged to include age and SES as well, and thus not focused on the three variables representing the SLI risk factor and bilingualism variables, showed that, as expected, age and SES also make contributions. However, these contributions were generally less strong than Positive Early Development, and neither L2 variable entered into the models (except for a small negative weight for Early L2 Exposure on performance on NWR-German).

References

Abed Ibrahim, L. and Hamann, C., 2017, Bilingual Arabic–German and Turkish–German children with and without specific language impairment: comparing performance in sentence and nonword-repetition tasks. In M. LaMendola and J. Scott (eds), *Proceedings of BUCLD 41* (Somerville, MA: Casscadilla), pp. 1–17.

Archibald, L. M. D. and Gathercole, S. E., 2007, The complexities of complex span: specifying working memory deficits in SLI. *Journal of Memory and Language*, **57**, 177–194.

Armon-Lotem, S, De Jong, J. and Meir, N. (eds), 2015, *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (Bristol: Multilingual Matters).

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

Boerma, T. and Blom, E., 2017, Assessment of bilingual children: what if testing both languages is not possible? *Journal of Communication Disorders*, **66**, 65–76.

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. *Journal of Speech, Language, and Hearing Research*, **58(6)**, 1747–1760.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., . . . Kressel, H. Y., 2015, STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*, **277(3)**, 826–832.

Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R. and Coltheart, M., 2007, *PALPA-P Provas de Avaliação da Linguagem e da Afasia em Português* (Lisbon: CEGOC).

Chevrie-Muller, C. and Plaza, M., 2001, *N-EEL Nouvelles épreuves pour l'examen du langage* (Paris: ECPA).

Chiat, S., 2015, Nonword repetition. In S. Armon-Lotem, J. de Jong and N. Meir (eds), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (Bristol: Multilingual Matters), pp. 123–148.

Chiat, S. and 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**, 1179–1189.

Chilla, S. and Şan, H., 2017, Möglichkeiten und Grenzen der Diagnostik erstsprachlicher Fähigkeiten: Türkisch-deutsche und türkisch-französische Kinder im Vergleich. In C. Yildiz, N. Topaj, R. Thomas and I. Gülzow (eds), *Die Zukunft der Mehrsprachigkeit im deutschen Bildungssystem: Russisch und Türkisch im Fokus* (Berlin: Peter Lang), pp. 175–205.

Condon, S. and Régnard, C., 2010, Héritage et pratiques linguistiques des descendants d'immigrés en France. *Hommes et migrations*, **1288**, 44–56.

Conti-Ramsden, G., Botting, N. and Faragher, B., 2001, Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, **42**, 741–748.

De Almeida, L., Ferré, S., Morin, E., Prévost, P., Dos Santos, C., Tuller, L. and Barthez, M. A., 2017, Identification of bilingual children with specific language impairment in France. *Linguistic Approaches to Bilingualism*, **7**, 331–358.

De Lamo White, C. and Jin, L., 2011, Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language and Communication Disorders*, **46**, 613–627.

Dos Santos, C. and Ferré, S., 2018, A nonword-repetition task to assess bilingual children's phonology. *Language Acquisition*, **25**, 58–71.

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. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer and B. Kühnert (eds), *Consonant Clusters and Structural Complexity* (Berlin: De Gruyter), pp. 285–308.

Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E. and Zebib, R., 2018, How to identify SLI in bilingual children? A study on sentence repetition in French. *Language Acquisition*, **25**, 85–101.

Fox-Boyer, A., 2014, *Psycholinguistische Analyse kindlicher Aussprache- störungen (PLAKSS-II)* (Frankfurt am Main: Pearson).

Gallon, N., Arris, J. and Van Der Lely, H., 2007, Non-word repetition: an investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics and Phonetics*, **21**, 435–455.

Glück, C., 2011, *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6–10*, 2nd edn (Munich: Elsevier, Urban & Fischer).

Grimes, D. A. and Schulz, K. F., 2005, Refining clinical diagnosis with likelihood ratios. *Lancet*, **365**, 1500–1505.

Grimm, A. and Hübner, J., in press, Nonword repetition by bilingual learners of German: the role of language specific complexity. In C. dos Santos and L. de Almeida (eds), *Bilingualism and Specific Language Impairment: Selected Proceedings of Bi-SLI 2015* (Amsterdam: Benjamins).

Grimm, A. and Schulz, P., 2014, Specific language impairment and early second language acquisition: the risk of over- and underdiagnosis. *Child Indicators Research*, **7**, 821–841.

Hamann, C., 2012, Bilingual development and language assessment. In A. K. Biller, E. Y. Chung and A. E. Kimball (eds), *Proceedings of BUCLD 36* (Somerville MA: Cascadilla), pp. 1–28.

Hamann, C. and Abed Ibrahim, L., 2017, Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, **2**, 1–19. https://www.doi.org/10.3389/fcomm.2017.00016.

Hamann, C., Penner, Z. and Lindner, K., 1998, German impaired grammar: the clause structure revisited. *Language Acquisition*, **7**, 193–246.

Jakubowicz, C. and Tuller, L., 2008, Specific language impairment in French. In D. Ayoun (ed.), *Studies in French Applied Linguistics* (Amsterdam: John Benjamins), pp. 97–134.

Khomsi, A., Khomsi, J., Pasquet, F. and Parbeau-Gueno, A., 2007, *Bilan Informatisé de langage Oral au cycle III et au Collège (BILO3C)* (Paris: ECPA).

Maddieson, I., Flavier, S., Marsico, E. and Pellegrino, F., 2011, *LAPSyD: Lyon–Albuquerque Phonological Systems Databases, Version 1.0* (available at: http://www.lapsyd.ddl.ish-lyon.cnrs.fr/).

Marinis, T. and Armon-Lotem, S., 2015, Sentence repetition. In S. Armon-Lotem, J. de Jong and N. Meir (eds), *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment* (Bristol: Multilingual Matters), pp. 95–122.

Marinis, T., Armon-Lotem, S. and Pontikas, G., 2017, Language impairment in bilingual children: state of the art 2017. *Linguistic Approaches to Bilingualism*, **7**, 265–276.

Montrul, S., 2008, *Incomplete Acquisition in Bilinguals: Re-Examining the Age Factor* (Amsterdam: John Benjamins).

Paradis, J., 2010, The interface between bilingual development and specific language impairment. Keynote article for special issue with peer commentaries. *Applied Psycholinguistics*, **31**, 3–28.

Paradis, J., 2011, Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, **1**, 213–237.

Paradis, J., Emmerzael, K. and Sorenson Duncan, T., 2010, Assessment of English language learners: using parent report on first language development. *Journal of Communication Disorders*, **43**, 474–497.

Paradis, J., Tulpar, Y. and Arppe, A., 2016, Chinese L1 children's English L2 verb morphology over time: individual variation in long-term outcomes. *Journal of Child Language*, **43**, 553–580.

Polišenská, K., Chiat, S. and Roy, P., 2015, Sentence repetition: what does the task measure? *International Journal of Language and Communication Disorders*, **50**, 106–118.

Schulz, P. and Tracy, R., 2011, *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSE-DaZ)* (Göttingen: Hogrefe).

Sua-Kay, E. and Santos, M. E., 2014, *Grelha de Avaliação da Linguagem – nível escolar (GOL-E) 2ª edição revista* (Alcoitão: Escola Superior de Saúde do Alcoitão).

Swets, J. A., Dawes, R. and Monahan, J., 2000, Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, **1**, 1–26.

Thordardottir, E., 2015, Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong and N. Meir (eds), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (Bristol: Multilingual Matters), pp. 331–358.

Thordardottir, E. and 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.

Tomblin, J. B., Records, N. L. and Zhang, X., 1996, A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, **39(6)**, 1284–1294.

Topbaş, S. and Güven, S., 2008, *Turkce Erken Dil Gelisim Testi* [Test of Early Language Development—TELD3. Turkish Adaptation] (Yönerge Kitabı. Ankara, Türkiye: Detay Yayıncılık).

Tucci, I. and Groh-Samberg, O., 2008, *Das enttäuschte Versprechen der Migration: Migrantennachkommen in Frankreich und Deutschland.* Schweizerische Zeitschrift für Soziologie **34(2)**, 307–334.

Tuller, L., 2015, Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong and N. Meir (eds), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (Bristol: Multilingual Matters), pp. 299–328.

Tuller, L., Abboud, L., Ferré, S., Fleckstein, A., Prévost, P., Dos Santos, C., Scheidnes, M. and Zebib, R., 2015, Specific language impairment and bilingualism: assembling the pieces. In C. Hamann and E. Ruigendijk (eds), *Language Acquisition and Development: Proceedings of GALA 2013* (Newcastle: Cambridge Scholars), pp. 533–567.

Zebib, R., Henry, G., Khomsi, A., Messarra, C. and Hreich, E., 2017, *Evaluation du langage oral chez l'enfant libanais: ELO-L* (Antelias: Liban Tests Editions).

Appendix A

Table A1. Standardized tests used to assess language skills in Arabic, French, German, Portuguese and Turkish

		Language skill tested						
Language	Test	Phonology	Vocabulary reception	Vocabulary expression	Morphosyntax comprehension	Morphosyntax production	Scoring system	Norm group
Arabic	ELO-L[a]	Word repetition	Picture selection	Picture naming	Picture–sentence matching	Sentence completion	Individual subtest scores and global score	3;0–7;11
French	N-EEL[b]	–	Picture selection	Picture naming	Picture–sentence matching	Sentence completion	Individual subtest scores	5;7–8;6
	BILO[c]	Word repetition	–	–	–	–	Individual subtest scores	5;0–15;0
German	WWT 6–10[d]	–	Picture selection	Picture naming	–	–	Individual subtest scores	5;6–10;11
	LiSe-DaZ[e]	–	–	–	Picture–sentence matching, truth value judgment	Story, sentence completion, lead-in questions	Individual subtest scores	3;0–6;11 (monolinguals), 3;0–7;11 (bilinguals)
	PLAKSS-II[f]	Picture naming	–	–	–	–	Individual subtest scores	2;6–7;11
Portuguese	PALPA-P[g]	Non-word repetition	Picture selection	Picture naming	Picture selection	Sentence repetition	Individual subtest scores	5;0–9;11 (some missing norms for all tasks)
	GOL-E[h]	–	Word definition	Antonym naming	–	Complex S from two simple S's	Individual subtest scores and global score	5;07–10;00
Turkish	TEDIL[i]	–	Picture selection	Picture naming	Picture selection, truth value judgment	Sentence repetition/completion/construction	2 composite scores: production & comprehension	2;0–7;11

Sources: [a]Zebib *et al.* (2017); [b]Chevrie-Muller and Plaza (2001); [c]Khomsi *et al.* (2007); [d]Glück (2011); [e]Schulz and Tracy (2011); [f]Fox-Boyer (2014); [g]Castro *et al.* (2007); [h]Sua-Kay and Santos (2014); [i]Topbaş and Güven (2008).

Appendix B

Table B1. Bivariate Pearson correlation: *r*- and *p*-values

	% NWR (global score)		% SR (identical rep.)	
	Germany	France	Germany	France
% SR (identical repitition)	**.689*****	**.714*****		
	.000	.000		
Age	**.408****	.205*	**.354****	.294**
	.002	.047	.007	.004
Mother's Education	.259	.259*	**.482*****	**.445*****
	.064	.014	.000	.000
Father's Education	.154	.255*	**.391****	.211*
	.262	.014	.003	.041
Positive Early Development	**.371****	**.493*****	**.407****	**.439*****
	.005	.000	.002	.000
No Risk Index	**.367****	**.467*****	**.395****	**.444*****
	.005	.000	.003	.000
Early L1 Exposure	–.009	–.098	–.205	–.083
	.948	.347	.130	.422
Early L2 Exposure	**–.336***	.095	–.212	.125
	.011	.364	.117	.229
AoO L2	.146	.044	.041	.013
	.282	.676	.767	.901
LoE L2	.125	.071	.180	.156
	.357	.494	.183	.130
Current L1 Use at Home[a]	–.226	–.203*	–.272*	–.203*
	.094	.049	.042	.048
Current L2 Use at Home	.115	.185	.119	.142
	.400	.074	.382	.170
Current L1 Richness	–.165	. 085	–.250	.120
	.225	.413	.063	.247
Current L2 Richness	.285*	.057	**.387****	.145
	.033	.584	.003	.160
L2 Language Dominance	.122	.123	**.317***	.096
	.371	.237	.017	.357

Notes: Correlation coefficients > .300 are shown in bold. $^{***}p < .000$, $^{**}p < .01$, $^{*}p < .05$. The Bonferroni multiple correlation adjusted *p*-value = .0033.

[a] Current use at home, in the Parents of Bilingual Children Questionnaire (PABIQ), is the sum of frequency of use of a language (0 = never, 1 = rarely, 2 = sometimes, 3 = usually, 4 = very often/always) between the child and the following persons: mother, father, other adults (grandparent, babysitter), and siblings (Tuller 2015).