A Nonword Repetition Task to Assess Bilingual Children's Phonology

Author(s): Christophe dos Santos and Sandrine Ferré

Source: *Language Acquisition* , January-March 2018, Vol. 25, No. 1, Special Issue: The Identification of Children with Specific Language Impairment (January-March 2018), pp. 58-71

Published by: Taylor & Francis, Ltd.

Stable URL: https://www.jstor.org/stable/10.2307/48541007

LANGUAGE ACQUISITION
2018, VOL. 25, NO. 1, 58–71
http://dx.doi.org/10.1080/10489223.2016.1243692

Routledge
Taylor & Francis Group

A Nonword Repetition Task to Assess Bilingual Children's Phonology

Christophe dos Santos and Sandrine Ferré

Université François-Rabelais de Tours

ABSTRACT

Children with specific language impairment (SLI) are particularly sensitive to phonological complexity in their language. Their performance drops when there are specific phonological structures or when complexity increases. A nonword repetition (NWR) test, which aims to assess the phonology of bilingual speakers with and without SLI, should include phonological properties that are independent of the language and phonological properties whose complexity is quantifiable. The methodology and constraints related to the creation of a NWR test named LITMUS-NWR-FRENCH, which combines these two objectives, are presented. This task was tested on a population of 67 children, 5½ to 8½ years old, bilingual and monolingual, with and without SLI, having in common French as L1 or L2. Results show that the LITMUS-NWR-FRENCH task differentiates between children with and without SLI in the context of bilingualism. It also shows the influence and importance of phonological complexity in children with SLI.

ARTICLE HISTORY
Received 18 April 2016
Accepted 28 September 2016

1. Introduction

During COST Action IS0804 (European Cooperation in Science and Technology), devoted to bilingualism and Specific Language Impairment (SLI), possible diagnostic tools and/or screening tools were discussed and designed. One of the goals was to have tools at our disposal that will be able to disentangle typically developing bilingual children (Bi-TD) from bilingual children with SLI (Bi-SLI).

Indeed, in the vast majority of languages, when standardized tests to detect SLI exist, the norms of these tests are based on a monolingual population; to our knowledge, the only exceptions are Lise-Daz for German (Schulz & Tracy 2011) and TAK for Dutch (Verhoeven & Vermeer 2002), both of them including bilingual populations in their standardization. Since usually only monolingual norms are available, SLI in bilingual children could be either underdiagnosed or overdiagnosed (Crutchley, Conti-Ramsden & Botting 1997; Grimm & Schulz 2014). Bilingual underdiagnosed children are often seen as children with difficulties because their L2 has not been mastered yet. These underdiagnosed children with SLI will probably not have access to regular clinical follow-up. On the other hand, the difficulties of overdiagnosed bilingual children are often wrongly analyzed as potentially caused by SLI. These typically developing children will be identified as requiring speech therapy even though they have no disorders.

As regards phonological assessment, in some batteries of tests aiming at assessing language fundamentals, some tools are often labeled as "phonological assessment tools." When they exist, these tools, which could range from phoneme production to word repetition, are again normed according to monolingual populations. Moreover, these tools often test not only phonology but also other areas—for example, the lexicon and/or the working memory. To cite one example, in France,

CONTACT Christophe dos Santos Christophe.dossantos@univ-tours.fr Département de Sciences du Langage, UFR Lettres & Langues, 3, Rue des Tanneurs BP 4103, 37041 Tours Cedex 1, France.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hlac.

the "Bilan Informatisé de Langage Oral" (BILO, Khomsi, Khomsi & Pasquet 2007) contains a word repetition test. The repetition is based on real L1 lexicon but also on the length of these words.

In order to assess mainly the phonology, we decided to develop a nonword repetition task, with nonwords limited to a maximum of a three-syllable length to reduce the effect of working memory. This type of task has a substantial theoretical advantage. This task can lean only weakly on L1 vocabulary knowledge. Bilingual children's performances should approximate monolingual ones. Nonword repetition tasks have other advantages: The time required for the task is only approximately five minutes, and adjustments for other languages are simplified compared to tasks using real words. The extensive list of these advantages are presented in Chiat (2015), which provides a synthesis of the COST Action IS0804 work, dealing especially with the relevance of such tasks for bilingual children's assessment.

2. Test development constraints

The nonword repetition test we wanted to develop aims at disentangling bilingualism from SLI. To reach this aim, we had one strong prerequisite: to test only phonology as much as possible. We wanted to limit two external influences often encountered on that type of tasks—working memory and lexical knowledge. For working memory, the longer the nonword, the stronger is the effect of working memory on repetition performance. This effect of length starts to appear in children with SLI when nonwords have more than two syllables (e.g., Poncelet & van der Linden 2003). In most nonword repetition tasks, nonwords can have up to five syllables or sometimes even more (e.g., Archibald & Gathercole 2006; Poncelet & van der Linden 2003). In these tests, working memory has a significant influence that we wanted to restrict. To limit this effect, we chose to have nonwords that will have no more than three syllables. Lexical knowledge plays also a role when nonwords sound like real words. This fact can be captured with a measure of wordlikeness (Wurm & Samuel 1997; Frish et al. 2000; Munson, Kurtz & Windsor 2005; among others). In order to limit the effect of lexical knowledge, we chose not to use existing L2 words, in our case French, to build our nonword test but rather to use elementary blocks (syllables and segments). These blocks would be combined to create nonwords. This choice allowed us to finely manipulate the blocks to combine and attune them to phonological assessment. We decided to mainly manipulate syllable structure, which seems to be a suitable marker for assessing phonological disorders (Ferré et al. 2012).

The flexibility given by block combinations should allow us to meet two goals that are sometimes conflicting: a scientific goal and a clinical goal. The first goal is to explore questions around phonological complexity in nonword repetition tasks. For this goal, we wanted to explore two assumptions. First, we wanted to check if complex nonwords more frequently failed. Second, we wanted to know if French phonological complexities, which are infrequent in other languages, are better mastered by French monolinguals compared to bilinguals. To investigate these two questions, creating enough nonwords is a requirement to have enough data to analyze the influence of complexity on such a task. The second goal, which is a clinical one, seeks to develop a screening tool for detecting children with SLI that can be used by a practitioner. In order to fulfill this goal, it is essential to choose the best nonwords for differentiating TD children from children with SLI, whether these children are bilingual or monolingual. This last point implies that the test should be effective in the greatest possible number of languages.

3. The litmus nonword repetition test

As previously mentioned, the test should be able to assess specific complexity points (in particular, points that are present in French phonology) and should be able to be applied to a large number of different languages. In order to cover these prerequisites, we created two types of items: one labeled "language independent" (LI), and the other labeled "language dependent" (LD). The items labeled as "language independent" should have a phonological structure that is possible in most languages of

the world. This label was given for the sake of simplicity, since these items would be more properly characterized as "quasi-independent." Indeed, it is not possible to completely isolate nonwords from language (Chiat 2015). LD items were created by integrating two complexity points that are part of French phonology. These complexity points are also present in a few languages.

3.1. *LI NW*

This test was designed to study the effect of phonological complexity and measure the ability of this complexity to detect (combined with tests from other language areas) children with SLI. To create LI items, we selected complexity points that are bound to three phonological aspects.

The first aspect to take into account is the syllable structure. To create nonwords, the first step was to select the syllable types that the NW could have. Our starting point was the syllable structure considered to be the simplest one either from typology or acquisition. This syllable structure is exclusively composed of an onset followed by a nucleus. The simplest syllable then consists of a consonant followed by a vocalic element: CV. This syllable type is universal since all studied languages to date show this syllable type. Two other syllable types were selected as well: CCV and CVC#. These two other syllable types were selected for two reasons: because of their complexity compared to CV syllable and because these syllable types are found in a majority of the world's languages. According to Maddieson (2006), in a sample of 515 languages, 88% can have a branching onset and/or consonant after a vowel,[1] generally at the end of a word.

The second aspect is related to segmental complexity. Segmental complexity was taken into account for consonants only. As for the choice of syllable types, we selected consonants that are early acquired and available in a large number of phonological systems. We chose two stops [p] and [k], as this manner of articulation is the first to be acquired. Without taking into account voicing (see later for explanation), the great majority of phonological systems possess a labial and a dorsal stop (LAPSyD). As regards place of articulation, dorsal [k] is considered more complex than [p] (Paradis & Prunet 1991; Hayes & Steriade 2004; among others). We have, then, at our disposal a labial/dorsal contrast in order to create our items. The choice of a labial stop instead of a coronal stop—both have roughly the same degree of complexity—is bolstered by the fact that a labial, like a dorsal, allows more easily branching onsets, as [pl] and [kl]. Indeed, the branching onset [tl] with a coronal stop is rarely encountered in the world's languages (Goad & Rose 2004). For manner contrast, we chose the stop/fricative one. In this case, fricative is considered more complex than stops (e.g., Jakobson 1969). The fricative [f] was selected as a possible segment for the creation of the items. Finally, to be able to have items with branching onsets that are present in a large number of languages, we have selected the liquid [l]. In a first version of the test, the rothic [ʁ] was the segment selected instead of [l]. Its high variability among the world's languages and dialects (Walsh Dickey 1997; Wiese 2001) and the difficulty of its transcription, caused us eventually to consider [l] as a better choice. For vowels, we have selected the three most common vowels among the languages of the world, according to LAPSyD: [i], [a], and [u] (Maddieson et al. 2011).

The third aspect is sequenciality (i.e., syntagmatic axis). Two types of sequenciality could increase item complexity: consonant sequence and syllable sequence. As regards consonant sequence, when places and/or manner of articulation alternate, the sequence is viewed as more complex. Furthermore, items beginning with a labial—that is, a nonlingual consonant—are regarded as simpler. For syllable sequences, in three syllable items, the second syllable is more fragile in French (Braud 2003)—even more so when this syllable has one or more points of complexity (e.g., a branching onset). In other languages, when stress is controlled, it seems to be the case as well. For example, Marshall & van der Lely (2009) found that children with SLI and children with dyslexia produced more errors on word-medial clusters compared to word-initial clusters.

[1]These numbers are corroborated by the LAPSyD database consultation (Maddieson et al. 2011).

Table 1. Complexity aspects taken into account in the development of the LITMUS-NWR-FRENCH Test

	Complexity	–	+	Examples	References
Segmental	Labial	Dorsal	[p] vs. [k]	Jakobson 1969; Aicart-De-Falco & Vion 1987; Johnson & Reimers 2010; Yamaguchi 2012	
	Stop	Fricative	[p] vs. [f]		
Syllabic	CV	CCV and CVC#	[ka] vs. [kla]	Jakobson & Halle 1956; Levelt, Schiller & Levelt 2000; MacNeilage 2008	
Séquential	Segment	Same place	Different place and/or manner	[pupa] vs. [puka]	Kern & Davis 2009; Fikkert & Levelt 2008
		Lab-Dor	Dor-Lab	[puka] vs. [kapi]	Costa 2010; Gerlach 2010; MacNeilage 2008; Santos 2007
	CCV and CVC position in 3 σ NW	In 1st and 3rd position	In 2nd position	[kupifla] vs. [kuflapi]	Braud 2003; Demuth & Song 2012; Marshall & van der Lely 2009

All these aspects are synthesized in Table 1. Each mentioned aspect includes nonexhaustive references from which our classification is based. This LI part of the test has been adapted in Arabic and German (Grimm et al. 2014). It is now being adapted in European Portuguese and Canadian English.

3.2. *LD NW*

LD NW was created with the same elements as LI NW: CV, CCV, and CVC# syllable types and segments [p], [k], [f], [l], [i], [a], and [u]. We added the coronal fricative [s] to these elements, which were already used for the creation of LI NW. This additional consonant enabled us to create NW, which could have the following complex consonant clusters: #sCV, #sCCV, sC#, and Cs#. These complex consonant clusters are unusual in the world's languages (Goad & Rose 2004) but are allowed in French, and this is why we named these NW "LD NW." For the creation of LD items, internal codas were also included, which was not the case for LI NW. It follows from these additional possible clusters that the LD part is considered more complex than the LI part.

Whether it was for the LI or LD part, we controlled the frequency of occurrences of each element previously mentioned. Indeed, [p] is as frequent as [k], or [i] is as frequent as [u]. We also have collected information about the wordlikeness of our NW with French words (on a 5-point scale).

Lastly, all NWs (LI and LD) were randomly ordered. The last step was to reorder consecutive NW that were too close phonetically. All the LI and LD NW are presented in Table 2.

We hypothesized that the TD bilingual children, whose first language do not have elements added in the LD part, will have more difficulties than monolingual TD for the LD part but will perform roughly the same as monolingual TD in the LI part. Children with SLI should show bad performance for both parts of the test, notably in the LD part, as this part includes complex French structures.

4. Method

4.1. *Data collection and procedure*

NW were included in a PowerPoint presentation. In order to make the test attractive, an alien named "Zubilu" appears on the screen. We told the child that Zubilu comes from another planet and would like to teach its language to him/her. The only thing the child has to do to learn that language is to repeat what the alien says. To hear the NW well, the child is equipped with CS100 Koss headphones. The child's productions are recorded digitally with a Zoom H4. Recordings allow transcribers to

Table 2. Nonword list by complexity

Control Items			
CCV	[kla], [fli], [plu]	CVC	[kip], [paf], [fuk], [kis], [fal]
sCV	[spu]	CV.CV	[faku], [pilu], [kapi], [lafi]
Low complexity items			
CV.CCV	[paklu], [fupli]	CCV.CV	[plifu], [fluka]
CV.CVC	[pukif], [kafip]	CV.CVs	[kifus], [fapus]
CV.CVL	[fapul], [kufal]	CV.CV.CV	[kifapu], [pufaki]
Medium complexity items			
CCV.CVC	[flukif], [klifak]	CCV.CCV	[flaplu], [plaklu]
CCV.CV.CV	[flipuka], [klipafu]	CV.CV.CCV	[kupifla], [fikupla]
CV.CCV.CV	[kuflapi], [piklafu]	CV.CV.CVC	[kapufik], [pifakup]
CV.CV.CVs	[kifapus], [pifukas]	CV.CV.CVL	[fikapul], [pakifal]
CCVC	[klaf], [fluk]	CCVL	[plal], [klil]
CVCs	[fips], [piks]	CVsC	[pusk], [kusp]
CCVs	[flis], [klis]	sCV.CV	[skafu], [spiku]
sCCV	[skla], [spli]	sCVC	[skap], [spaf]
CVL.CV	[pilfu], [filpa]	CVs.CV	[kuspa], [fiska]
High complexity items			
CCVCs	[pliks], [klups]	CCVsC	[klisp], [plusk]
CV.CVL.CV	[kufalpi], [kupalfi]	CV.CVs.CV	[pafuski], [fikuspa]
sCV.CV.CV	[skapufi], [spakifu]		

transcribe and check after the event. It also enables the experimenter to stay engaged with the child, as transcription will be done after the test session.

For the transcription, we used an Excel file where children's codes (to make them anonymous) and basic information were inserted. In this Excel sheet, NW pronounced incorrectly were transcribed. One sample of each transcription was also blind checked by a third party. A NW was considered as incorrect when at least one segment was deleted, added, or replaced. The only exception was voicing substitution (for example [p] repeated as [b]). In a number of phonological systems only one consonant of the voiceless/voiced pair was available. For example, in modern standard Arabic we only found the voiced labial stop and not the voiceless one (Amayreh & Dyson 2000; Ryding 2005). This test is intended for a bilingual population; therefore, it is important to code errors that are not due to differences coming from different phonological systems. Moreover, when voice and voiceless categories are present in the system, in general they do not share the same voice onset time (VOT) values (Lisker & Abramson 1964; Cho & Ladefoged 1999). Thus, voiced stops in stressed syllables have a small positive value, which can be interpreted as voiceless stops by Parisian French monolinguals. Finally, it appears that while some children could phonetically produce a voicing distinction, this distinction might stay unnoticed by transcribers (Scobbie 1998); for all these reasons, we decided not to report voicing errors. Whenever a doubt remains about child production, this doubt always benefits the child. For example, if transcribers are uncertain whether the vowel produced is an [a] or an [ɛ], no error will be counted. As the aim of the task is to assess phonology and not articulation, we also did not count as an error the following case: any sound that is systematically substituted by another sound (for example, [t] systematically substituted by [k]).

Once the transcription and the check are done, each error is coded according to its position in the NW and its type (deletion, substitution, metathesis, etc.). From this coding, we computed one score based on exact repetition: NW will be coded as 1 when it is correctly repeated and 0 when it was repeated with at least one error.

4.2. *Participants*

Sixty-seven children, from 5½ to 8½ years old, participated in the study. They were divided into four groups: monolingual TD, bilingual TD, monolingual SLI, and bilingual SLI. They were recruited

within the context of the BiLaD project (ANR-12-FRAL-0014-01), whose main goal is to study bilingual language development. All the children under study are bilinguals, sharing French as one of their languages and controlled for their language dominance. Hearing problems were a criteria of exclusion. For children in speech language therapy, only the children with SLI but no reported hearing problems were recruited. For TD children, in France, a screening is systematically done at schools at age 5. Children were excluded if doubt persisted.

4.3. *Bilingual child assessment*

We used a parental questionnaire, the PaBiQ (Tuller 2015), to assess the bilingualism level of each child. This questionnaire gives us a portrait of the child's linguistic situation and also gives us a rough risk factor for SLI. Some differences between contexts of migration (born in France or arrived later) exist between our participants. Nevertheless, there is no significant difference in terms of length of exposure or age of contact between groups. In addition, we did not find any significant correlation between the test results and the PaBiQ variables.

As we have previously noted, bilingual child assessment with monolingual norms may result in misdiagnosis and inappropriate treatment (Crutchley, Conti-Ramsden & Botting 1997; Grimm & Schulz 2014). This is the reason why we tested our bilingual population with a battery of standardized tests targeting their L1: for English CELF4-UK (Semel, Wiig & Secord 2006) and for Arabic ELO-L[2] (Zebib et al. In press). We also tested French with the help of N-EEL and BILO (N-EEL, Chevrie-Muller & Plaza 2001; BILO-2, Khomsi, Khomsi & Pasquet 2007), which are standardized batteries of tests assessing French. None of these batteries has computed norms for bilingual populations. Therefore, we followed the guidelines proposed by Thordardottir (2012) for the identification of SLI in a bilingual population with standardized monolingual tests as reported in Table 3.

Table 3. Guidelines for the identification of SLI in a bilingual population (Thordardottir 2012)

Assessment		Cutoff in Two Different Language Areas[4]
Monolingual		−1.25 *SD*
Bilingual	In the dominant language	−1.5 *SD*
	Balanced Bilingualism	−1.75 *SD*
	In the nondominant language	−2.25 *SD*

Nonverbal cognitive level of bilingual children was assessed through Raven's Progressive matrices (Raven, Court & Raven 1986). Children below the 9th percentile were excluded, which could be translated into a nonverbal level of 80 or below in Weschler's classification.[3]

For a complete overview of the recruitment and the classification of the bilingual children with or without SLI in our project, see Tuller et al. (2015). As the results of this first step, we ended up with the four groups of children mentioned previously: monolingual TD, bilingual TD, monolingual with SLI, and bilingual with SLI. These four groups of children are presented in Table 4.

(i) The group of Bilingual SLI children (Bi-SLI) is composed of 13 Bi-SLI children according to the criteria proposed by Thordardottir (2012). Ten out of 13 have Arabic as an L1 and 3 have English as L1. These children were recruited in speech-language pathologists' (SLPs)

[2]The ELO-L (Zebib et al. In press) is, to the best of our knowledge, the only standardized tests available for Arabic. ELO-L is based on Lebanese and normed on Lebanese children. Adaptations for Algerian, Libyan, Moroccan, and Tunisian were recorded.
[3]However, children with a score below the 9th percentile but with a normal WISC PIQ or a normal score on another nonverbal test were included.
[4]From Thordardottir (2013).

offices of the Region "Centre" in France. The preliminary criterion was that children were involved in speech therapy for oral language disorders.

(ii) The group of typically developing bilingual children (Bi-TD) is composed of 30 children (18 with Arabic as L1 and 12 with English as L1). These children were recruited in ordinary schools or associations.

(iii) The group of monolingual French-speaking children with SLI (Mo-SLI) is composed of 10 children. They were recruited from a reference center for language impairment and learning disorders (Centre de Référence des Troubles du Langage et des Apprentissages —CRTLA) of the Tours hospital. To be sure about the specificity of their linguistic deficit, all children were checked for normal hearing with an audiogram, and their nonverbal IQ was above 85.

(iv) The group of typically developing monolingual French-speaking children (Mo-TD) is composed of 14 children without any speech therapy or other difficulties. These children were recruited in ordinary schools in Tours and the surrounding region.

Table 4. Participants

	L1	N	Total	Mean Age (SD)	Age Min.–Max.	Age of Contact	Length of Exposure (in Mo.)
Bi-SLI	English	3	13	6;06 (0;08)	5;04–8;01	0–3;00	62–83
	Arabic	10					
Bi-TD	English	12	30	6;03 (0;09)	5;04–8;02	0–6;04	12–81
	Arabic	18					
Mo-SLI	French	10	10	7;05 (0;07)	6;02–8;05		
Mo-TD	French	14	14	5;08 (0;03)	5;04–6;03		

On average, there is a difference of one year between Mo-SLI and Bi-SLI children due to recruitment differences (CRTLA or SLP's office). However, even if Mo-SLI children are older in average, it does not impact our analysis, as they performed lower than our Bi-SLI children.

Considering our reduced sample sizes, group results were compared using a Mann-Whitney nonparametric test for independent samples. The Wilcoxon signed-rank nonparametric test was used for within-group analyses.

5. Results

5.1. *Does LITMUS-NWR-FRENCH allow us to disentangle children with SLI from TD children?*

Figure 1 shows the results of the exact repetition task for monolingual TD, bilingual TD, monolingual with SLI, and bilingual children with SLI with percentages and standard deviations.

These results show a statistically significant difference between the performance of Bi-SLI and Bi-TD children (Bi-SLI/Bi-TD: $U = 74.5$; $p < .005$; Mo-SLI/Mo-TD: $U = 4.0$; $p < .001$). Furthermore, we didn't find any difference between the score of Bi-TD and Mo-TD children (Bi-TD/Mo-TD: $U = 158.5$; $p = .26$). However, Mo-SLI children performed significantly lower than Bi-SLI children (Mo-SLI/Bi-SLI: $U = 21.5$; $p < .05$). This difference could be due to a difference in recruitment places. As noted previously, Mo-SLI children were recruited at the hospital, whereas Bi-SLI children were recruited in SLPs' offices. We can assume that more severe phonologically impaired children are more likely to be referred to the hospital. If this is the case, this bias could explain the difference between Mo-SLI and Bi-SLI results. There was almost no overlap between SLI and TD groups, in any case. This first measure suggests that the LITMUS-NWR-FRENCH is a good indicator of SLI in bilingual children.

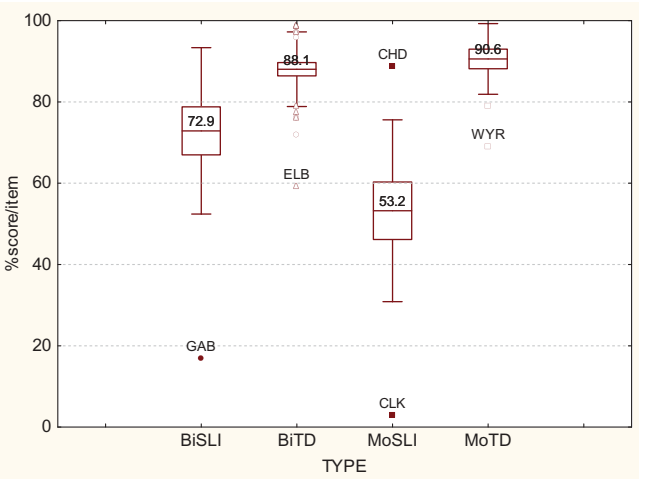

Figure 1. Mean percentages of exact repetition on the LITMUS-NWR-FRENCH for each group of children.

5.2. *Complexity effect on children with SLI*

As mentioned in the test description, the present test is composed of two parts, allowing us to assess language impairment in a bilingual context. Structural variables shared by a high number of languages are included in the LI part, which can be considered as the less complex part of the test. More complex structural variables are included in the LD part. These more complex variables should be challenging for bilingual children but above all to children with SLI, regardless of the bilingualism context.

The results in Figure 2 show that only Mo-SLI children showed a significant difference between the LI and LD parts ($z = 2.49$, $p < .05$). This result can be explained, here again, by a bias in recruitment locations between Mo-SLI and Bi-SLI children, as already mentioned in the previous section. No difference was found for the three other groups of children (Bi-SLI: $z = 1.25$, $p = .21$; Bi-TD: $z = 0$, $p = 1$; Mo-TD: $z = 0.31$, $p = .75$). For the TD children, this result can be explained by the fact that these children perform at ceiling; no difference can be found between the two parts of the

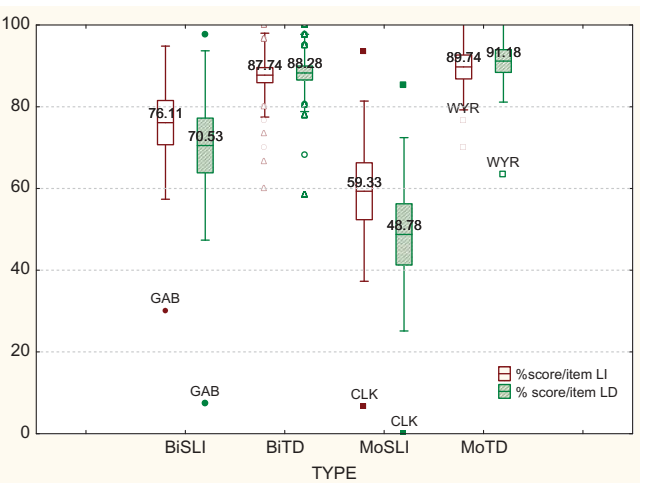

Figure 2. Mean percentages of exact repetition on LI and LD parts of LITMUS-NWR-FRENCH for each group of children.

test in that specific case. Finally, for Bi-TD children, this absence of difference could be due to their bilingualism, as it is balanced in most of the cases. However, we found the same intergroup differences for both parts of the test with even higher values for the LD part for the bilingual children (Bi-SLI/Bi-TD: U = 73.0; $p < .005$). The more complex LD part seems to be then more difficult for children with SLI.

5.2.1. *Complexity effect: Syllable number*

The number of syllables is in general considered as having a great impact on repetition tasks. From lengths of three syllables and more, performances decrease, especially for children with SLI. A working memory deficit was hypothesized to explain these lower performances of children with SLI (Archibald & Gathercole 2006; Tamburelli & Jones 2012; Thordardottir & Brandeker 2013). Our data presented in Figure 3 confirm the results of these previous studies. All our groups show a significant decrease in performance between dissyllabic items and trisyllabic items, with the exception of Mo-TD children, who perform at ceiling (Bi-SLI $z = 2.31$, $p = .02$; Bi-TD $z = 2.03$, $p = .04$; Mo-SLI $z = 2.29$, $p = .02$; Mo-TD $z = 1.33$, $p = .18$). No decrease or increase was found between monosyllabic and disyllabic nonwords.

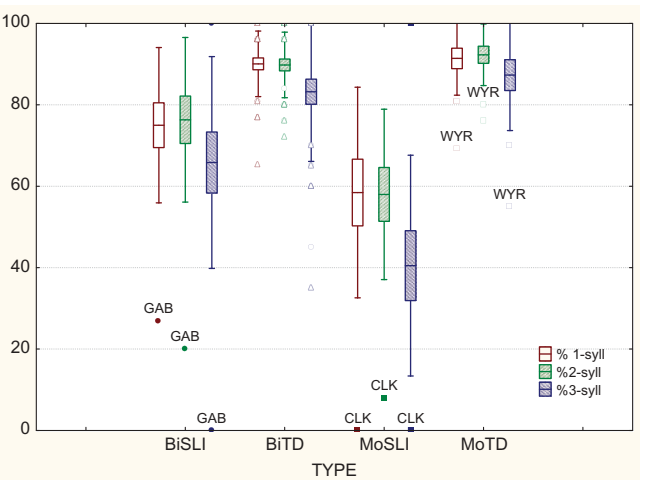

Figure 3. Mean percentages of exact repetition on the LITMUS-NWR-FRENCH depending on the number of syllables.

Intergroup differences follow the previous results for nearly all nonword lengths: A statistically significant difference between Bi-SLI and Bi-TD was found, and similar performances were observed for Bi-TD and Mo-TD children.

5.2.2. *Complexity effect: Consonant clusters (CC)*

When two consonant clusters are present within nonwords, there is a large fall in performances for Mo-SLI and Bi-SLI children as shown in Figure 4 (from 69.4% to 25%, $z = 2.66$ $p < .01$, and from 52.9% to 25%, $z = 1.88$ $p < .05$ respectively). Bi-TD productions were also impacted by the presence of consonant clusters, but because we found a great variability of performance in this group, no difference was found between nonwords with one consonant cluster and nonwords with two consonant clusters ($z = 1.4$, $p = .15$). Again, no difference was found for Mo-TD, as they already perform nearly at ceiling in all the cases. For children with SLI, the large difference produced by the presence of a second consonant allows us to assume that phonological structure is more important than length for impairment detection, at least in our test where nonwords were no longer than three

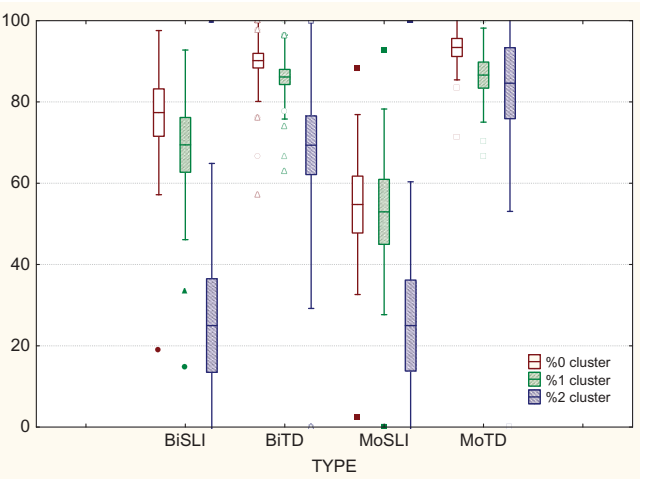

Figure 4. Mean percentages of exact repetition on the LITMUS-NWR-FRENCH depending on the number of clusters.

syllables. This hypothesis is reinforced by the fact that the difference between the performances of Mo-SLI and Bi-SLI no longer exists when we take into account consonant cluster number (Bi-TSL/Mo-TSL 1-cluster: U = 32; p = .06; 2-clusters: U = 58; p = .87). This factor seems to be more resilient to different degrees of severity.

5.3. *Sensitivity and specificity*

A test that aims to screen or assess language impairment needs to fulfill high standards. It needs to avoid as much as possible false positives (a TD child who will be diagnosed with SLI) and false negatives (a child with SLI who will be considered as TD). In order to do so, we computed sensitivity and specificity, which are statistical measures of the performance of a medical testing (in other words, its diagnostic power). Sensitivity measures the proportion of actual positives that are correctly identified as such (true positive: in our case, SLI children diagnosed as SLI), and specificity measures the proportion of negatives that are correctly identified as such (true negative: in our case, TD children who were considered as TD). A cutoff point needs to be set to determine the sensitivity and specificity of a test. A large number of studies follow the cutoff proposed by Vance & Plante (1994). For them, a sensitivity or specificity level of 90% or higher could be considered indicative of good discriminant accuracy, whereas levels between 80% and 89% are considered fair, and levels below 80% are considered too low. Likelihood ratios are calculated for a positive decision (LR+) as well as for a negative decision (LR–). The positive likelihood ratio measures the probability to be diagnosed as impaired if impaired and is calculated with the following formula: LR+ = sensitivity/(1 – specificity). The negative likelihood ratio measures the probability to be diagnosed as not impaired if not impaired and is calculated with the following formula: LR– = (1 – sensitivity)/specificity. According to Jekel, Elmore & Katz (1996), a test should be considered useful if (LR+/LR–) is above 50. Sensitivity and specificity is calculated from mean and standard deviation (*SD*) of the control group (in our case, TD children). This *z*-score allows us to manipulate different cutoffs: –1 *SD*, –1.28 *SD*, and –2 *SD* (see Thordardottir et al. 2011) for the explanation of these cutoffs choice).

Our test aims to discriminate SLI from TD children in bilingual or monolingual contexts. Our population is composed of two types of children. In a first step, if we only consider the monolingual population, it turns out that our test has an excellent discriminatory power: With a –2 *SD* cutoff, we reach a sensitivity of 90% and a specificity of 92%. LR+ is equal to 11.7 and LH– is equal to 0.11, which give us a ratio of 106. Therefore, with a cutoff at –2 SD, the test can be used as screening tool

for SLI. It should be noted that these values are also found for both parts of the test (namely, LI and LD) independently from a cutoff of –1.28 *SD*

If we now consider all groups including bilinguals, the test is never at the same time highly specific and highly sensitive. If we set the cutoff at –1 *SD*, sensitivity reaches 73% and specificity 82%. Likelihood ratios are relatively poor: LR+ = 4 and LR– = 0.33. With a cutoff at –1.28 SD, specificity rises to 89% but sensitivity decreases to 68%. Positive likelihood ratio is equal to 6. When –2 SD is chosen, the specificity is even higher (95%), but as a trade-off, sensitivity falls to 59%, which is far from acceptability. Likelihood ratios are slightly better but still not acceptable, at least for LR– (LR+ = 13 and LR– = 0.43). If we take a look only at the LI part, the test seems to be more specific than sensitive, particularly with a standard deviation of –1.65. At this threshold, specificity has a value of 91%, while sensitivity is only 59%, again far from acceptable levels. Likelihood ratios are not better than before, with a LR+ of 6.5 and a LR– of 0.45. The LD part shows the best sensitivity/specificity ratio. Indeed, for a threshold at –1 *SD*, sensitivity reaches 82% and specificity 86%. However, likelihood ratios are still not acceptable (LR+ = 6 and LR– = 0.21). The best balance for the LD part is found with a cutoff of –1.28 *SD*. Specificity reaches 93%, while sensitivity has a value of 77%. Positive and negative likelihood ratios are 11.33 and 0.24 respectively.

Thus, if Bi-SLI children are always significantly different from Bi-TD children, it seems, however, that difference in severity level due to a possible bias of recruitment location (Mo-SLI at the hospital, Bi-SLI at speech language therapist's offices) causes a too significant drop in sensitivity but allows to keep a good specificity at –2 *SD*. Furthermore, we have seen that children with SLI performances seem to be more impacted than TD children by some variables like structural complexity. Our first trial to downsize the test, keeping in mind structural complexity, shows that sensitivity and specificity should be reinforced. Our next step will be to study in depth the LITMUS-NWR-French reduced version.

6. Conclusion

LITMUS-NWR-FRENCH was developed in order to discriminate between children with and without SLI in multilingual and monolingual contexts. This tool seeks to test phonological skills and includes items varying in phonological structure complexity. For the creation of this task, we narrowed down some of the characteristics of already existing nonword repetition tasks (e.g., for English: Archibald & Gathercole 2006 and Dollaghan & Campbell 1998; for Spanish: Girbau & Schwartz 2007; for French: Poncelet & van der Linden 2003 and Thordardottir & Brandeker 2013). In particular, we limited working memory effects by having short items (of at most three syllables). We also limited the effect of lexical knowledge by choosing and combining elementary blocks (syllables and segments) in order to have building blocks that are shared by a large number of languages in the world system (Maddieson et al. 2011), which therefore were not based on a specific phonological system (at least for the language independent part of the task). These choices were always made according to our goal: to create a tool to screen quickly and easily bilingual children on phonology, other variables being minimized or controlled (for a review on nonword repetition tasks, see Chiat 2015). As a result, our items are shorter compared to most nonword repetition tasks and are neither a sequence of CV syllables (e.g., Thordardottir & Brandeker 2013), nor limited to a specific phonological system (e.g., Archibald & Gathercole 2006), which makes it unique in shape and applicable to any multilingual learning contexts.

In this first study, our results show a significant difference between children with and without SLI despite differences in recruitment locations, language pairs, age, or length of exposure. However, including some complexity points that are present in French but not frequent in other languages does not impact the performance of bilingual children. The question arises whether we need to keep these specificities in future adaptations of this test, such as downsizing it for clinical purposes; therefore, we wonder whether we would obtain the same results with these adaptations and in other languages.

Finally, it has been shown that the effect of working memory in our test seems to be much weaker than the effect produced by phonological complexity. Performances of children with SLI fall when it

moves from dissyllabic items to trisyllabic items, but this fall is much bigger when it concerns the number of consonant clusters.

This test is an integral part of the battery of tests designed for the population assessed in the BiLaD project we are working in. Under this context, bilingual child recruitment is being pursued and additional data will allow us to strengthen these first results. Another part of the study, currently undergoing, is a reduction of the test in order to increase its discriminating power and to improve its clinical use. Adaptations of at least one part of the test were already developed or are also ongoing in Arabic, German, and Portuguese, for example. First results seem to confirm the contribution of this test for oral language assessment of bilingual children, especially when used in conjunction with tests assessing other language domains.

Acknowledgments

We thank the anonymous reviewers for helpful comments and all those who assisted with various aspects of the study: Laurie Tuller, Philippe Prévost, Rasha Zebib, Lætitia de Almeida, Éléonore Morin, Angela Grimm, the student assistants, and finally the clinicians and the families who participated in the study.

Funding

This study was supported by the French National Research Agency (ANR) Grant - ANR-12-FRAL-0014-01.

ORCID

Christophe dos Santos http://orcid.org/0000-0003-3520-0064
Sandrine Ferré http://orcid.org/0000-0002-4925-0006

References

Archibald, Lisa M. D. & Susan E. Gathercole. 2006. Short-term and working memory in specific language impairment. *International Journal of Language and Communication Disorders* 41. 675–693.

Aicart-De-Falco, Suzy & Monique Vion. 1987. La mise en place du système phonologique du français chez les enfants entre 3 et 6 ans: une étude de la production. *Cahiers de Psychologie Cognitive* 7(3). 247–266.

Amayreh, Mousa & Alice Dyson. 2000. Phonetic inventories of young Arabic speaking children. *Clinical Linguistics & Phonetics* 14(3). 193–215.

Braud, Virginie. 2003. *Acquisition de la prosodie chez des enfants francophones: les phénomènes de troncations*. Nantes, France: University of Nantes dissertation.

Chevrie-Muller, Claude & Monique Plaza. 2001. *N-EEL Nouvelles épreuves pour l'examen du langage*. Paris: ECPA.

Chiat, Shula. 2015. Nonword repetition. In Sharon Armon-Lotem, Jan de Jong & Natalia Meir (eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment*, pp. 125–150. Bristol, UK: Multilingual Matters.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27(2). 207–229.

Costa, Teresa. 2010. *The acquisition of the consonantal system in European Portuguese: Focus on place and manner features*. Lisbon, Portugal: University of Lisbon dissertation.

Crutchley, Alison, Gina Conti-Ramsden & Nicola Botting. 1997. Bilingual children with Specific Language Impairment and standardized assessments: Preliminary findings from a study of children in language units. *International Journal of Bilingualism* 1. 117–134.

Demuth, Katherine & Jae Yung Song. 2012. How phonological representations develop during first language acquisition. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford handbook of laboratory phonology*, 397–406. Oxford: Oxford University Press.

Dollaghan, Chris & Thomas F. Campbell. 1998. Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research* 41(5). 1136–1146.

dos Santos, Christophe (see Santos)

Ferré, Sandrine, Laurie Tuller, Eva Sizaret & Marie-Anne Barthez. 2012. Acquiring and avoiding phonological complexity in SLI vs. typical development of French: The case of consonant clusters. In Christine Mooshammer,

Marianne Pouplier, Philip Hoole & Barbara Kühnert (eds.), *Consonant clusters and structural complexity*, 285–308. Berlin: Mouton de Gruyter.
Fikkert, Paula & Claartje Levelt. 2008. How does place fall into place? The lexicon and emergent constraints in children's developing grammars. In Peter Avery, B. Elan Dresher & Keren Rice (eds.), *Contrast in phonology: Theory, perception, acquisition*, 231–268. Berlin: Mouton de Gruyter.
Frisch, Stefan A., Nathan R. Large & David B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on processing of nonwords. *Journal of Memory and Language* 42(4). 481–496.
Gerlach, Sharon Ruth. 2010. *The acquisition of consonant feature sequences: Harmony, metathesis and deletion patterns in phonological development*. Minneapolis, MN: University of Minnesota dissertation.
Girbau, Dolors & Richard G. Schwartz. 2007. Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders* 42(1). 59–75.
Goad, Heather & Yvan Rose. 2004. Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. In René Kager, Joe Pater & Wim Zonneveld (eds.), *Constraints in phonological acquisition*, 109–157. Cambridge: Cambridge University Press.
Grimm, Angela, Sandrine Ferré, Christophe dos Santos & Shula Chiat. 2014. Can nonwords be language-independent? Cross-linguistic evidence from monolingual and bilingual acquisition of French, German, and Lebanese. Paper presented at the 13th International Congress for the Study of Child Language (IASCL), July 14–18, Amsterdam.
Grimm, Angela & Petra Schulz. 2014. Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research, Special Issue: Children at Risk for Adverse Educational Achievement* 7(4). 821–841.
Hayes, Bruce & Donca Steriade. 2004. Introduction: The phonetic basis of phonological markedness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 1–33. Cambridge: Cambridge University Press.
Jakobson, Roman. 1969. *Langage enfantin et aphasie*. Paris: Éditions de Minuit.
Jakobson, Roman & Morris Halle. 1956. *Fundamentals of language*. The Hague: Mouton.
Jekel, James F., Joann G. Elmore & David L. Katz. 1996. *Epidemiology, biostatistics, and preventative medicine*. Amsterdam: W. B. Saunders.
Johnson, Wyn & Paula Reimers. 2010. *Patterns in child phonology*. Edinburgh: Edinburgh University Press.
Kern, Sophie & Barbara Davis. 2009. Emergent complexity in early vocal acquisition: Cross-linguistic comparisons of canonical babbling. In François Pellegrino, Egidio Marsico, Iona Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 353–376. Berlin: Mouton de Gruyter.
Khomsi, Abdelhamid, Jabril Khomsi & Frédérique Pasquet. 2007. *Bilan informatisé de langage oral au cycle 2 (BILO-2)*. Paris: Éditions du Centre de Psychologie Appliquée.
Levelt, Claartje, Niels Schiller & Willem Levelt. 2000. The acquisition of syllable types. *Language Acquisition* 8(3). 237–264.
Lisker, Leigh & Arthur Abramson. 1964. Cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.
MacNeilage, Peter F. 2008. *The origin of speech*. Oxford: Oxford University Press.
Maddieson, Ian. 2006. Correlating phonological complexity: Data and validation. *Linguistic Typology* 10(1). 106–123.
Maddieson, Ian, Sébastien Flavier, Egidio Marsico & François Pellegrino. 2011. LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0. http://www.lapsyd.ddl.ish-lyon.cnrs.fr/.
Marshall, Chloe R. & Heather K. J. van der Lely. 2009. Effects of word position and stress on onset cluster production: Evidence from typical development, SLI and dyslexia. *Language* 85(1). 39–57.
Munson, Benjamin, Beth A. Kurtz & Jennifer Windsor. 2005. The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research* 48(5). 1033–1047.
Paradis, Carole & Jean-François Prunet. 1991. Introduction: Asymmetry and visibility in consonant articulations. In Carole Paradis & Jean-François Prunet (eds.), *The special status of coronals: Internal and external evidence*, Vol. 2, 1–28. New York: Academic Press.
Poncelet, Martine & Martial van der Linden. 2003. L'évaluation du stock phonologique de la mémoire de travail: élaboration d'une épreuve de répétition de non-mots pour population francophone. *Revue de Neuropsychologie* 13 (3). 375–405.
Raven, John C., John Court & John Raven. 1986. *Raven's coloured matrices*. London: H. K. Lewis.
Ryding, Karin. 2005. *A reference grammar of standard modern Arabic*. Cambridge: Cambridge University Press.
Santos, Christophe dos. 2007. *Développement phonologique en français langue maternelle: une étude de cas*. Lyon: Université Lumière Lyon 2 dissertation. http://www.ddl.ish-lyon.cnrs.fr/fulltext/santos/dos Santos_2007.pdf.
Schulz, Petra & Rosemarie Tracy. 2011. *Linguistische Sprachstandserhebung—Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe Verlag.
Scobbie, James M. 1998. Interactions between the acquisition of phonetics and phonology. In M. Catherine Gruber, Derrick Higgins, Kenneth Olson & Tamra Wysocki (eds.), *Papers from the 34th Annual Regional Meeting of the Chicago Linguistic Society, Volume II: The Panels*, 343–358. Chicago: Chicago Linguistics Society.

Semel, Eleanor, Elisabeth H. Wiig & Wayne A. Secord. 2006. *Clinical Evaluation of Language Fundamentals (CELF4-UK)*, 4th edn. London: Harcourt Assessment.
Tamburelli, Marco & Gary Jones. 2012. Investigating the relationship between nonword repetition performance and syllabic structure in typical and atypical language development. *Journal of Speech, Language, and Hearing Research* 56(2). 708–720.
Thordardottir, Elin. 2012. Proposed diagnostic procedures and criteria for Bi-SLI COST action studies. COST Action IS080. http://www.bi-sli.org.
Thordardottir, Elin & Myrto Brandeker. 2013 The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders* 46(1). 1–16.
Thordardottir, Elin, Eva Kehayia, Barbara Mazer, Nicole Lessard, Annette Majnemer, Ann Sutton, Natacha Trudeau & Gevorg Chilingaryan. 2011. Sensitivity and specificity of French language and processing measures for the identification of Primary Language Impairment at age 5. *Journal of Speech, Language, and Hearing Research* 54 (2). 580–597.
Tuller, Laurie. 2015. Clinical use of parental questionnaires in multilingual contexts. In Sharon Armon-Lotem & Jan de Jong (eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment*, pp. 301–330. Bristol, UK: Multilingual Matters.
Tuller, Laurie, Layal Abboud, Sandrine Ferré, Alice Fleckstein, Philippe Prévost, Christophe dos Santos, Maureen Scheidnes & Rasha Zebib. 2015. Specific Language Impairment and bilingualism: Assembling the pieces. In Cornelia Hamann & Esther Ruigendijk (eds.), *Language acquisition and development: Proceedings of GALA 2013*. Newcastle: UK: Cambridge Scholars Publishing.
Vance, Rebecca & Elena Plante. 1994. Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools* 25. 15–24.
Verhoeven, Ludo & Anne Vermeer. 2002. *Taaltoets Alle Kinderen (TAK)*. Tilburg: Zwijsen.
Walsh Dickey, Laura. 1997. *The phonology of liquids*. Amherst, MA: University of Massachusetts dissertation.
Wiese, Richard. 2001. The phonology of /r/. In T. Alan Hall (eds.), *Distinctive feature theory*, 335–368. Berlin: de Gruyter.
Wurm, Lee H. & Arthur G. Samuel. 1997. Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language* 36(2). 165–187.
Yamaguchi, Naomi. 2012. *Parcours d'acquisition des sons du langage chez deux enfants francophones*. Paris: Université Sorbonne Nouvelle Paris 3 dissertation.
Zebib, Rasha, Abdelhamid Khomsi, Guillemette Henri, Camille Messarra, Edith Hreich, Lili Dahdah & Z. Attieh. In press. *Batterie d'Évaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Paris: ECPA.