First and Second Language Sentence Repetition: A Screening Measure for Dual Language Learners?

May-Britt Monsrud, Veslemøy Rydland, Esther Geva & Solveig-Alma Halaas Lyster

Abstract

The sentence repetition (SR) test is considered as a promising diagnostic tool for detecting language proficiency in monolingual learners, but less is known about its potential to identify dual language learners' (DLLs) linguistic proficiency. Considering that challenges with language learning, such as developmental language disorders (DLDs), is evident in both first and second language (L1 and L2), it is important to use tools that can assess language functions in both languages whenever feasible.

We found that students' L1 and L2 SR performance levels draw upon their phonological short-term memory capacity, vocabulary and grammar skills in the language being assessed. We also found a positive relationship between students' L1 phonological memory and L2 SR scores, suggesting that DLLs' phonological short-term memory capacity assessed in their L1 indicates individual differences in short-term verbal memory that support SR performance, as assessed in L2.

The results highlight a challenge in both research and clinical settings which often use monolingual assessment tools and associated norms to identify DLLs who might be at risk of developing DLDs or other learning difficulties.

Keywords: Assessment; Dual Language Learners; Developmental Language Disorders; Sentence Repetition

First and Second Language Sentence Repetition: A Screening Measure for Dual Language Learners?

Introduction

In many countries, children are learning more than one language, and their first and home languages often are not the societal language. The large individual heterogeneity of these children's language skills has been documented (Hammer et al., 2014). We do not expect a greater prevalence of developmental language disorders (DLDs) among dual language learners (DLLs) as compared with monolingual language learners, but many countries face a situation in which DLLs are both over- and underrepresented in special needs assessment services (Solari et al., 2014; Sullivan & Bal, 2013). A major concern is the lack of assessment measures to assess first language (L1) and second language (L2) competencies (Bedore & Pena, 2008). It is essential to increase our knowledge of DLLs L1 and L2 competence levels and shed light on how assessment procedures can investigate language proficiency among (Pratt et al., 2021).

L1 assessments are important in identifying potential DLDs (American-Speech-Language-Hearing Association [ASHA], 2021). However, few assessment tools are available in the language combinations spoken by the multitude of DLLs in Norway and other European countries. Also, in situations in which assessments are available in both languages, the target language's norms may not be reliable for DLLs who have experienced multifaceted exposure to L1 and L2 in their language environments.

In this study we examined L1 and L2 language skills through a sample of 546 DLLs (ages 6–13 years) with no known special needs and a sample of 14 DLLs with DLDs (ages 8-13 years). The home languages are Albanian, Somali, Tamil, Turkish, Urdu/Panjabi and Vietnamese.

By employing a widely used screening measure, the sentence repetition (SR) test, the present study seeks to investigate whether and how assessment in both L1 and L2 increases understanding of language proficiency and DLLs' potential language difficulties beyond an assessment of L2 skills only.

Developmental Language Disorders (DLDs)

Approximately 10% of all children start school with language difficulties that cannot be diagnosed (Paul et al., 2018). The term *specific language impairment* (SLI) has been used in many contexts to define language impairments in children whose cognitive skills fall within normal limits, with no identifiable reason for the impairment (Reilly et al., 2014). Recently, a consensus has been reached suggesting that the somewhat broader term *developmental language disability* (DLD) better conveys language difficulties with unknown causes (Gallinat & Spaulding, 2014; Paul et al., 2018). Identification of DLDs should involve use of multiple assessments, should take into account whether the difficulties encountered impact everyday life and academic attainment. However, no clear cut-off point distinguishes DLDs from the lower end of normal variations in language abilities (Bishop et al., 2017; Paul et al., 2018) and so cognitive cut- offs are not recommended.

Identifying DLDs in DLLs is challenging because grammatical structures commonly found in second language (L2) acquisition overlap to some extent with the errors made by monolingual children with DLDs. Moreover, capturing vocabulary knowledge distributed between L1 and L2 is often not possible because of a lack of available tests (Bedore & Pena, 2008; Paradis, 2016; Peña et al., 2016; Tuller et al., 2018).

DLDs among DLLs need to be identified in both L1 and L2; therefore, whenever feasible, it is important to use procedures that assess language functions effectively in both (Paradis, 2016; Thordardottir & Brandeker, 2013). A study of Swedish-Arabic preschool children with and without DLDs found that DLLs' without DLDs demonstrated strong language development in at least one language (Hakansson et al., 2003). Considering that DLLs may have unequal exposure to L1 and L2 it is difficult to differentiate between language variety caused by a lack of exposure and language challenges caused by DLDs (Hammer et al., 2014). This is a major concern in assessment procedures for detecting possible DLDs in DLLs (Geva & Farnia, 2017).

The systematic group differences between many DLLs and monolingual learners evident in measures of vocabulary (e.g. August et al., 2005) and grammar (e.g. Hakansson et al., 2003) were less pronounced in assessments of phonological short-term memory, such as non-word repetition, recall of digit span and recall of words (word span) (Calvo & Bialystok, 2014; de Abreu et al., 2013; Peña et al., 2018). However, non-word repetition tests are not viewed as entirely independent of language experience, as they may be influenced by phonological proximity to real words (Estes et al., 2007). Word span tests and forward digit span tests require children to recall lists of words or numbers, but word span tests are viewed as relying more on linguistic knowledge than digit span tests (Buac et al., 2016; Melby-Lervåg & Hulme, 2010). Altogether, phonological short-term memory tests are viewed as more distinct and less influenced by language skills such as tests of grammar and vocabulary.

Using the Sentence Repetition Test (SR) to Detect Potential DLDs in DLLs

SR is considered as a clinical marker for DLDs and is included in most diagnostic batteries to identify DLD among monolingual and DLL (Hesketh & Conti-Ramsden, 2013; Riches, 2012; Tuller et al., 2018; Zebib et al., 2019). Even though SR is viewed as a promising

diagnostic tool, the skills needed to perform such tasks are not fully understood (Riches, 2012). According to Riches, two main factors predict SR in Baddeley et al.'s (1998) multicomponent model of working memory: predictors that implicate memory and those that implicate language (Baddeley et al., 1998; Riches, 2012). In this model, SR taps the episodic buffer's capacity. Research with monolingual learners has demonstrated that SR performance draws on a broad set of related skills, including vocabulary, grammar and phonological short-term memory (Frizelle et al., 2017; Klem et al., 2015). A study on monolingual learners provided an examination of SR's potential underlying mechanisms – such as short-term memory, syntactic knowledge and working memory – and demonstrated that the DLD group relied more on short-term memory than students without DLDs (Riches, 2012). This also was the case with DLL (English/Spanish speaking) samples with and without DLDs when the contribution of short-term memory, vocabulary and language exposure in SR was investigated (Pratt et al., 2021). The results indicated that SR relied on skills related vocabulary, grammar and short-term memory in both groups, but that DLLs with DLDs rely more heavily on short-term memory than DLLs without DLDs.

Present Study

Considering that SR is used widely in screening for potential DLDs, we wanted to look more closely at what skills SR provides in both L1 and L2. We examined language components that predict L1 and L2 SR performance in DLLs. Even though SR differentiates between DLLs with and without DLDs, the results from prior research indicated large variability in DLL than results from monolingual learners (Chiat et al., 2013). However, more studies are needed, including investigations of non-English-speaking populations with diverse L1 backgrounds.

In the present cross-sectional study, we investigated to what extent short-term memory. grammar and vocabulary predict SR in L1 and L2 in a large sample of DLLs without DLDs. We also wanted to explore whether L1 and L2 SR can identify DLLs at risk and compare L1 and L2

SR, L1 and L2 word span, L1 and L2 grammar and L1 and L2 vocabulary results from DLLs at risk to DLLs with identified DLDs.

The following research questions (RQs) guided this study:

RQ1. To what extent were students' L1 and L2 sentence repetition scores predicted by their L1 and L2 word span, L1 and L2 vocabulary, and L1 and L2 grammar?

RQ2. How do students identified as being at risk for DLDs based on their L1 and L2 sentence repetition perform relative to a sample of DLLs identified as having DLDs?

Norwegian Context

Recent statistics indicate that in Norway's population (5 million inhabitants), 220 L1s are used among immigrant families. The languages represented in the present study—Albanian, Somali, Tamil, Turkish, Urdu/Panjabi and Vietnamese—are among the main immigrant languages. Parents of the participants in this study arrived in Norway in the 1970s (Turkish and Urdu/Panjabi) or as refugees between 1970 and 2000 (Albanian, Somali, Tamil and Vietnamese). The families in our study predominately lived in in the greater Oslo area. The immigrant populations in this area generally have lower socioeconomic status (SES) compared to majority of the population (Statistics, 2016). Among elementary school students, 14% exclusively speak a language other than Norwegian at home. However, in some districts in the greater Oslo area, DLLs comprise 50–90% of the student population. The main language of instruction is Norwegian, and local schools do not provide L1 educational support on a regular basis. Teacher education covers bilingualism issues to varying degrees.

Approximately 90% of DLLs attend preschool by the time they reach 5 years old (Norwegian Directorate for Education and Training, 2016). I

Method and Design

Participants

All the students in the present study attended schools in multi-ethnic neighbourhoods in the greater Oslo area. The first sample of participants comprised 546 students without DLDs (girls = 294, boys = 252; age range = 6.0–12.11). These students were recruited for the study by teachers who distributed written information about the project's bilingual approach in L1 and L2 (Norwegian) to parents who used a home language other than Norwegian. The teachers were instructed to include only students with no special education needs and who did not receive any special needs education.

We recruited students whose first languages were Albanian (n = 62), Somali (n = 66), Tamil (n = 89), Turkish (n = 116), Urdu (n = 136) and Vietnamese (n = 77). A questionnaire submitted to parents sought information about each student's country of birth, home language background, language used at home and preschool attendance.[1] The participating schools distributed the questionnaires (in Norwegian and in the respective L1s), and 60% of the parents (n = 325) returned completed questionnaires. The questionnaire data from the respondents revealed that 87% of the students were born in Norway (n = 291), and 13% (n = 42)[2] were born in other countries, while 85% (n = 276) attended preschool. The questionnaire data revealed that N = 324 parents reported that students use their L1 at home.

Table 1 displays the distribution of participants by gender and age group.

---------- *Table 1 approximately here* ----------

The second sample comprised DLLs whom special needs services (i.e., school psychologists or speech therapists) identified as having DLDs (n = 14; girls = 5, boys = 9; age

[1] Children in Norway generally begin nursery/preschool as 1-year-olds until they start school at age 6, but preschool attendance is not compulsory.
[2] Albania = 4, Somalia = 24, Sri Lanka (Tamil) = 1, Turkey = 5, Pakistan (Urdu) = 5 and Vietnam = 3.

range = 8.0–12.11)[3]. The participants with DLDs were recruited from the same neighbourhoods and schools as the DLLs without DLDs. They were all born in Norway, representing the following home language groups: Tamil = 5; Turkish = 4; Somali = 1; Urdu = 3; and Vietnamese = 1. The parents of the students in the DLD group reported that their children used their L1 to communicate with family members, but used their L2 (i.e., Norwegian) to interact with siblings and friends. Students in this sample attended preschool for one to three years. We were unable to recruit students who had been referred to special need services in the youngest age groups. Given the small sample, the participants are merged into two age groups. Table 2 displays information about the DLLs with DLDs by gender and age group (9.0-10.11 and 11.0-12.11).

---------- *Table 2 approximately here* ----------

Measures[4] [5]

The measures are standardised tests tailored to the Norwegian context and normed on monolingual Norwegian students. These tests were translated and adopted to the six home languages.

Translation and Adaptation of Measures

To identify students' abilities in their L1 and L2, we translated and adapted the Norwegian tests into the students' first languages in line with the principles of translation and test adaptation (American Educational Research Association et al., 2014; Peña, 2007). For each language, two qualified translators with native competence in the language independently

[3] Given the N=14 only, the participants are merged into two age-groups

[4] Permission for translation for research purposes in this project was provided by the study authors, E. Ottem and J. Frost, for sentence repetition, word span and expressive vocabulary by the GL Assessment for the British Picture Vocabulary Scale (BPVS-II) and by Pearson for TROG-2 Test for the reception of grammar.

[5] Additional information is avilaiable in Appendix 1.

translated all the test items. The translators had high levels of L1 and L2 competence and relevant professional/academic backgrounds as linguists, interpreters, teachers or researchers, along with expertise in the participants' bilingual and bicultural backgrounds. The two translators and a Norwegian project group comprising professionals and research experts in the field of language development discussed each item's translations. Once the draft list of translated items was complete, six independent translators who were not involved in the previous phase (with the same relevant qualifications) back-translated the home-language versions into Norwegian to determine whether the test's original form was retained.

It was important to evaluate whether the words, utterances, grammatical syntax and complexity were appropriate for the students in their specific bicultural and bilingual contexts (American Educational Research Association et al., 2014; Peña, 2007).

Given that translation is not the best practice for language assessments, we also calculated the percentage of participants who correctly responded to each item on the tests. We noted a steady decline in the percentage of correctly identified items as the tests progressed, and we observed this trend in both L1 and L2 assessment procedures.

Sentence Repetition. The way SR tasks are constructed vary as to sentence length, syntactic and grammatical complexity, number of items and the scoring process, correct repetition of the target constructions and number of target words repeated correctly. This SR test comprises 16 sentences of increasing length, syllables, grammatical complexity and information units (Ottem & Frost, 2005).

The test administrator read the sentences to each student one at a time, and the student was asked to repeat each sentence verbatim. Those who correctly repeated the test item received 1 point per item. The administrator terminated the test after three consecutive incorrect answers. Here is

an example of a sentence repetition task: 'The girl kicked the soccer ball over the roof' ('Jenta sparket ballen over hustaket').[6]

Word Span (Serial Word Recall). The word span test comprised a series of three, four or five monosyllabic words, with four series at each length, for a total of 12 test items (Ottem & Frost, 2005). The words in each item were read to the student, who was asked to repeat them verbatim and in the same order. Like the SR test, each item was scored as 1 or 0, with testing terminated after three consecutive incorrect answers. Here is an example of the word span test: 'green, can, cut, wind' ('grønn, kan, kutt, vind').

Grammar. The Test for Reception of Grammar (TROG) assessed students' grammatical comprehension (Bishop, 1989; Lyster & Horn, 2009). The TROG-2 is a picture test in which the student was asked to identify one out of four pictures that is consistent with the sentence read. The test comprises 80 items divided into 20 blocks of four sentences with similar structures. Because the syntactic structures are not equally difficult in each language, the complete test (80 items) was administered to children in their L1 and L2.

Expressive Vocabulary. The test of expressive vocabulary comprises antonyms (14 items) and word definitions (12 items) (Ottem & Frost, 2005). In the antonym subtests, the researcher asked the child to name the opposite of a given word (e.g., 'What is the opposite of *cold*?'). In the definition section, the researcher asked the student to define words (e.g., 'What is a *bed*?'). Each item received a score of correct or incorrect (1 or 0). In each section, testing is terminated after three consecutive incorrect responses.

Receptive Vocabulary. The Norwegian version of the British Picture Vocabulary Scale (BPVS-II) contains 144 items (12 blocks) of increasing difficulty (Dunn et al., 1997; Lyster et al.,

2010). The student was asked to identify one out of four drawings that corresponded to a word that the test administrator spoke. In this study, each student started with the first item given in both languages because of the large variation that exists in vocabulary knowledge in general in DLLs. We used the BPVS-II manual's termination criterion, i.e., the test ended when eight or more of the 12 items in a block were answered incorrectly. Each correct answer was awarded 1 point.

Procedure

All the children were assessed individually in their L1 and L2 at their respective schools. The assessors had professional backgrounds as speech and language therapists, school psychologists or trained graduate students. Assessors from respective L1 backgrounds assessed L1, and native Norwegian assessors evaluated L2. The L1 and L2 assessment order was counterbalanced, i.e., half the DLL sample in each age group was assessed in their L1 first, then tested in Norwegian (their L2) later. The other half was assessed in Norwegian first and later in their L1.

All the measures were scored on the spot. There was some attrition due to school absences. There was also some attrition due to availability of the L1 test material of the Turkish version; however, the Turkish-speaking sample did not differ from the sample size or age groups of the other language groups.

Analysis

To answer RQ1, we examined the extent to which phonological memory, vocabulary and grammar predicted L1 and L2 SR scores in the larger sample of students (N = 546). We investigated this in two separate models because we expected that the same-language predictors would account for the variation in students' L1 and L2 SR scores. However, we also explored whether phonological memory, vocabulary and grammar assessed in one language predicted SR outcomes in the other language. Preliminary analyses revealed that the children's L1 and L2 SR

scores were not related significantly once the predictor variables were included in the analyses. We used the Mplus software program (Muthén & Muthén, 1998–2017) to perform analyses to achieve two goals: First, as students were nested within schools, the Mplus program's Complex option allowed us to account for the cluster effect on estimated standard errors. Second, because of the relatively large variability in students' ages, we included a covariate (age in months) for all the variables to increase the estimates' precision. The predictor variables were allowed to co-vary in both predictive models.

To predict the students' scores on L2 SR, we first regressed L2 SR on L2 vocabulary (expressive), L2 grammar and L2 word span. As a second analytic step, we entered the equivalent L1 measures as predictors. Only the significant predictor variables are included in the figures presented here. It should be noted that due to the high correlation between expressive and receptive vocabulary, we only included expressive vocabulary because it also represented receptive vocabulary when it came to building a model that predicted the SR results (Appendix 2). We employed the same procedures to predict students' L1 SR, which was first regressed on L1 vocabulary, L1 grammar and L1 word span before we entered L2 predictors.

We evaluated model fit against the following guidelines (see discussion in Hu & Bentler, 1999). The root mean square error of approximation (RMSEA) estimate should be lower than .06, and the standard root mean squared residual (SRMR) should be lower than .08.

In response to RQ2's more exploratory approach, we needed to identify students who might be at risk of DLDs in the sample without DLDs. We identified students who might be at risk of DLDs by identifying those in this sample with low SR scores in both their L1 and L2 (in the 9.0-10.11 and 11.0-12.11 age range). This group was designated an *at-risk* group.

To investigate the percentage of students in the potential at-risk group when comparing DLLs to monolingual norms, we investigated the percentage of students who had scores at or below -1.25 SD in L2 SR, compared with monolingual norms (n = 900; see Appendix 3) in the selected age groups (6.0-8.11, 9.0-10.11, 11.0-12.11).

Results

The descriptive statistics of the typically achieving sample, in terms of the language task battery in L1 and L2 (by age group), are summarised in Table 3, with Cronbach's alphas included.

----------*Table 3 approximately here*----------

Table 3 shows the expected age-related increases in raw scores for all the variables. At the group level, the students' L1 test scores for SR, word span and vocabulary (receptive and expressive) were stronger than their scores on equivalent L2 assessments. However, the difference between the students' L1 and L2 scores on the grammar assessment was minor.

Predicting L2 and L1 Sentence Repetition

The prediction of L2 SR had good model fit (RMSEA = 0.57; CFI = .990; TLI = .962; SRMR = .010) and is displayed in Figure 1. It is evident that L2 expressive vocabulary had a relatively strong standardised partial regression on L2 SR ($\beta = 0.30$, $t = 5.57$, $p < 0.001$). L2 SR also was predicted by L2 grammar ($\beta = 0.20$, $t = 4.33$, $p < 0.001$), L2 word span ($\beta = 0.25$, $t = 5.69$, $p < 0.001$) and L1 word span ($\beta = 0.13$, $t = 3.33$, $p < 0.001$). Notably, L1 word span explained the significant variance in L2 SR beyond same-language predictors. Moreover, as Figure 1 shows, relatively strong intercorrelations were observed between the predictor variables, particularly between L2 expressive vocabulary and L2 grammar ($r = 0.49$, $t = 14.20$, $p < 0.001$), and between L1 and L2 word span ($r = 0.42$, $t = 13.06$, $p < 0.001$).

As expected, the predictor variables were related to the students' age in months at the time of the assessments. This age effect was particularly strong for the L2 expressive vocabulary (β = 0.66, t = 24.77, $p < 0.001$) and L2 grammar (β = 0.59, t = 21.88, $p < 0.001$), but it also was evident in the students' scores on L2 word span (β = 0.37, t = 11.99, $p < 0.001$) and L1 word span (β = 0.32, t = 5.76, $p < 0.001$). Students' age in months did not predict additional variance in their L2 SR scores once the age effects on the predictor variables were accounted for (β = 0.00, t = 0.03, p = 0.978). This model explained 48% of the variance in L2 SR.

----------*Figure 1 approximately here*----------

Figure 2 presents the model predicting L1 SR. This model fit the data well (RMSA=.000; CFI=1.00; TLI=1.005; SRMR=.0.12). As Figure 2 shows, L1 expressive vocabulary had a strong standardised partial regression on L1 SR (β = 0.52, t = 11.01, $p < 0.001$), followed by L1 word span (β = 0.23, t = 4.39, $p < 0.001$) and grammar (β = 0.14, t = 2.32, $p < 0.001$). L2 word span did not have a standardised partial regression on L1 SR (β = −0.02, t = −0.35, p = 0.725). Similar to the L2 model, we found strong intercorrelations between L1 expressive vocabulary and L1 grammar (r = 0.45, t = 6.38, $p < 0.001$), and between L1 word span and L2 word span (r = 0.42, t = 13.76, $p < 0.001$). Notably, L1 expressive vocabulary and grammar also correlated significantly with students' L2 word span scores (r = 0.20, t = 4.43, $p < 0.001$ and r = 0.23, t = 3.75, $p < 0.001$, respectively).

As one might expect, all the predictor variables were related to students' age in months: L1 expressive vocabulary (β = 0.47, t = 9.16, $p < 0.001$); L1 grammar (β = 0.44, t = 9.72, $p < 0.001$); L1 word span (β =. 30, t = 5.22, $p < 0.001$); and L2 word span (β = 0.36, t = 10.88, $p < 0.001$). Similar to what we found in the L2 model, students' age in months did not predict additional

variance in students' L1 SR score once the predictor variables were included ($\beta = -0.06$, $t = -1.16$, $p = 0.245$). This model explained 49% of the variance in L1 SR.

---------- *Figure 2 approximately here* ----------

Comparing the At-Risk Group to the DLD Group

When identifying at-risk students within the DLL sample using a cutoff of -1.25 SD in both L1 and L2, a few students (1%; n = 5) scored that low. However, some of the students scored -1.25 on L1 SR (n = 56), and others (n = 58) scored that low on L2 SR.[7] Because we were unable to detect a sizable at-risk group using the -1.25 SD cut-off, we applied a -1.00 SD cut-off in both L1 and L2. Altogether, 26 participants had a score that was -1.00 SD or lower than the mean on L1 and L2 SR tests.

Table 4 presents descriptive statistics of the at-risk group (n = 26) and the identified DLD group (n = 14) on L1 and L2 SR, L1 and L2 word span, L1 and L2 grammar, L1 and L2 expressive vocabulary and L1 and L2 receptive vocabulary. We excluded the youngest age group among the at-risk students because no students in this age group were in the identified DLD group.

----------*Table 4 approximately here* ----------

The comparison between the at-risk group and identified DLD group found small differences between the groups. However, the L2 expressive and receptive vocabulary was stronger in the DLD group than in the at-risk group.

[7] Students scoring low on L1: Albanian = 5; Somali = 4; Tamil = 9; Turkish = 0; Urdu = 5; and Vietnamese = 22. Students scoring low on L2: Albanian = 2; Somali = 10; Tamil = 2; Turkish = 6; Urdu = 15; and Vietnamese = 6. Students scoring low on both L1 and L2 sentence repetition: Somali = 1; Tamil = 2; and Vietnamese = 2.

Identifying Students Who Might Be at Risk Using Monolingual Norms

Next, we used monolingual norms for SR assessment in three age groups to uncover possible at-risk students in the DLL group (n = 900; see Appendix 3). Only the students with no previously known special needs (n = 546) who scored at least 1.25 SD below the mean on L2 SR were considered. This resulted in a considerable overrepresentation of DLLs who might be viewed as at risk of DLDs. The percentages of students scoring below the cut-off point were 45%, 58% and 40% in the DLL group, compared with 9%, 12% and 9% in the same three age groups of monolingual students, respectively.

Discussion

This study offered an investigation of the language components that predict L1 and L2 SR performance in DLLs. We explored whether low scores on L1 and L2 SR indicated a group of DLLs at risk and also compared a group of DLLs at risk to a group with DLDs.

First, we found that students' L1 and L2 SR performance levels draw upon their language (vocabulary and grammar) and memory capacities (phonological short-term memory) in the language being assessed. The SR performance assessed in L2 was predicted by students' L2 vocabulary, grammar and word-span scores, but not by their L1 vocabulary and grammar skills. However, students' L1 word span scores explained significant additional variance in their L2 SR scores, suggesting that DLLs' phonological short-term memory capacity assessed in their home language relates to the students' performance on SR when assessed in L2. However, the reverse was not found, as the relation between students' L2 word span and L1 SR was not significant. Although the relation between L1 word span and L2 SR should be interpreted with caution, these findings might suggest that a specific cross-language effect may occur, encompassing the language that children are exposed to from birth and the societal language that children typically encounter later in life. Furthermore, the fact that we detected cross-language correlations between

students' word span scores in one language and vocabulary and grammar scores in the other language also indicates that word span is an important marker of language aptitude among DLLs. The present study builds on a concurrent correlational design with limited ability to address questions of transfer between students' L1 and L2 over time. Nevertheless, the relationship between students' L1 word span and L2 performance suggests that cognitive processing impacts language skills. Future longitudinal studies should investigate the relationship between DLLs' word span and SR scores over time.

Although the patterns of relationships were relatively similar in our predictive models of L1 and L2 SR, it is notable that vocabulary explained most of the variance in L1 SR. L1 vocabulary was a stronger predictor of students' L1 SR scores than their L1 grammar. The large standard deviations in the students' L1 vocabulary scores across age groups indicate that some of the students in the sample did not have a strong command of L1.

Second, a surprising finding when we used -1.00 SD on L1 and L2 SR to identify an at-risk group for DLLs without DLDs was those minor differences between the DLD and at-risk groups were detected. However, the low scores on SR and word span indicate that undetected language difficulties may exist among DLLs without DLDs. This finding suggests that SR in both L1 and L2 is a promising clinical marker for identifying language difficulties, and that the test may be a good steppingstone gauge for further decision makers to detect language difficulties or other learning disabilities.

The results from comparing L1 and L2 SR among DLLs without DLDs using a -1.25 SD cut-off point criterion suggests that some students are L1-dominant, while others are L2-dominant. In other words, L1 and L2 language proficiency levels in the sample had large variations. This finding also may indicate that DLLs receive language exposure in two different

languages, and that L1 and L2 SR detects variability in their relative language proficiency, but not DLDs.

The current results also underscore the systematic bias related to DLLs having potential language difficulties generated by monolingual norms. These results underscore a challenge in both research and clinical settings, often using monolingual assessment tools and associated norms to investigate DLLs' language skills to identify DLLs who might be at risk for DLDs or other learning difficulties.

Implications for Assessment and Practice

Given that SR uncovers a broad range of language competencies in both L1 and L2, our study indicates that SR also may identify language difficulties. To avoid over-identification of DLLs at risk for DLDs, it is important to compare students' performance in relation to their linguistic reference group and not to monolingual norms. Moreover, both L1 and L2 assessments are needed to uncover potential DLDs.

What is new from this study is that L1 and L2 SR was predicted by variations in same-language competencies to a large extent. Low scores on L1 and L2 SR may identify language difficulties, and an assessment in L2 using only monolingual norms is a source of error. The results from this study also suggest that vocabulary—in addition to memory and grammar—is important to include in the analysis of what SR measures.

This study examined a group of students not identified with DLDs and not referred to special need services by their teachers even though some test scores were similar to those of the identified DLD group. The distinction between low-performing L2 skills among some DLLs when compared with monolingual norms and identification of DLDs is, to a large extent, arbitrarily understood by special needs services and teachers. This study's results suggest that

low scores in L2 SR may identify students who need access to systematic and qualified language assessment that provides a platform for developing interventions.

Limitations

Because of limited background information, we could not determine with certainty whether the students who scored below the cutoff on both L1 and L2 SR tests were at risk for DLDs, or whether their scores reflected insufficient exposure to L1 and/or L2 language. Furthermore, the DLD group was small, which might have resulted in a biased result. However, bearing in mind that this group was identified by language experts and school psychologists, we should trust the assessment and deem this a valid group to use for comparison. Although the translations of the measures accounted for the participants' bilingual and bicultural background, and assessors assessed L1 using relevant L1 backgrounds, translation is not the best practice for language assessments. This is another potential limitation of the study.

References

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

August, D., Carlo, M., Dressler, C. & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice, 20*(1), 50-57.

Baddeley, A., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review, 105*(1), 158.

Bedore, L. M. & Pena, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*(1), 1-29.

Bishop, D. (1989). Test for reception of grammar. *London: Medical Research Council.*

Bishop, D., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Consortium, C., Adams, C., Archibald, L., Baird, G., Bauer, A. & Bellair, J. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary

Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068-1080.

Buac, M., Gross, M. & Kaushanskaya, M. (2016). Predictors of processing-based task performance in bilingual and monolingual children. *Journal of Communication Disorders*, *62*, 12-29.

Calvo, A. & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278-288.

Chiat, S., SArmon-Lotem, S., Marinis. Theo, Polinska, K., Roy, P. & Seeff-Gabriel, B. (2013). The Potensial of Sentence Imitation Tasks for Assessment of Language.

de Abreu, P. M. E., Baldassi, M., Puglisi, M. L. & Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, *56*(2), 630-642.

Dunn, L. M., Dunn, L. M., Whetton, C. & Burley, J. (1997). *The British Picture Vocabulary Scale* (2nd. utg.). nferNelson.

Estes, K. G., Evans, J. L. & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*.

Frizelle, P., O' & Bishop, D. V. (2017). Assessing understanding of relative clauses: A comparison of multiple-choice comprehension versus sentence repetition. *Journal of Child Language*, *44*(6), 1435-1457.

Gallinat, E. & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1363-1382.

Geva, E. & Farnia, F. (2017). Issue Editor Foreword. *Topics in Language Disorders*, *37*(2), 109-113.

Hakansson, G., Salameh, E.-K. & Nettelbladt, U. (2003). Measuring language development in bilingual children: Swedish-Arabic children with and without language impairment. *Linguistics*, *41*(2; ISSU 384), 255-288.

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C. & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*(4), 715-733. https://doi.org/10.1016/j.ecresq.2014.05.008

Henriksen, K. (2010). *Levekår og kjønnsforskjeller blant innvandrere fra ti land (Rapporter 6/2010) [Living conditions and gender differences among immigrants]*. Statistics Norway.

Hesketh, A. & Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of specific language impairment. *PloS one*, *8*(2), e56314.

Hu, L. t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E. & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental science*, *18*(1), 146-154.

Lyster, S. & Horn, E. (2009). Test for Reception og Grammar (TROG-2). Norsk Versjon. I. Stockholm: Pearson Assessment.

Lyster, S., Horn, E. & Rygvold, A. (2010). Ordforråd og ordforrådsutvikling hos norske barn og unge. Resultater fra en utprøving av British picture vocabulary scale, (BPVS II). *Spesialpedagogikk*, *9*, 35-43.

Melby-Lervåg, M. & Hulme, C. (2010). Serial and free recall in children can be improved by training: Evidence for the importance of phonological and semantic representations in immediate memory tasks. *Psychological science*, *21*(11), 1694-1700.

Norwegian Directorate for Education and Training. (2016). *Utdanningsspeilet 2016 : tall og analyse av grunnopplæringen i Norge* Utdanningsdirektoratet.

Ottem, E. & Frost, J. (2005). Språk 6-16; screening test: manual. *Bredtvet kompetansesenter.*

Paradis, J. (2016). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research, 59*(1), 171-182.

Paul, R., Norbury, C. & Gosse, C. (2018). Language disorders from infancy through adolescence: Listening, Speaking. *Reading, Writing, and Communicating, 4*, 756.

Peña, E. D. (2007). Lost in Translation: Methodological Considerations in Cross-Cultural Research. *Child development, 78*(4), 1255-1264.

Peña, E. D., Bedore, L. M. & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders, 51*(2), 192-202.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A. & Bedore, L. M. (2018). Bilingual English Spanish Assessment (BESA). *Baltimore, MD: Brookes.*

Pratt, A. S., Peña, E. D. & Bedore, L. M. (2021). Sentence repetition with bilinguals with and without DLD: differential effects of memory, vocabulary, and exposure. *Bilingualism: Language and Cognition, 24*(2), 305-318.

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders, 47*(5), 499-510.

Solari, E. J., Aceves, T. C., Higareda, I., Richards-Tutor, C., Filippini, A. L., Gerber, M. M. & Leafstedt, J. (2014). Longitudinal prediction of 1st and 2nd grade English oral reading fluency in English language learners: Which early reading and language skills are better predictors? *Psychology in the Schools, 51*(2), 126-142.

Statistics, N. (2016). *Immigrants and Norwegian-born to immigrant parents, 1 January 2016.* Hentet 09.06.2016 fra https://www.ssb.no/en/befolkning/statistikker/innvbef/aar/2016-03-03

Sullivan, A. L. & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children, 79*(4), 475-494.

Thordardottir, E. & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders, 46*(1), 1-16.

Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., Dos Santos, C., Abed Ibrahim, L. & Zebib, R. (2018). Identifying language impairment in bilingual children in France and in Germany. *International Journal of Language & Communication Disorders, 53*(4), 888-904.

Zebib, R., Tuller, L., Hamann, C., Abed Ibrahim, L. & Prévost, P. (2019). Syntactic complexity and verbal working memory in bilingual children with and without Developmental Language Disorder. *First Language*, 0142723719888372.

Table 1. Typical developing dual language learners (DLLs) by age group and gender

Age	Girls	Boys	TOTAL
6.0–6.11	28	20	48
7.0–7.11	42	44	86
8.0–8.11	47	37	84
9.0–9.11	40	53	93
10.0–10.11	66	45	111
11.0–11.11	37	23	60
12.0–12.11	34	30	64
TOTAL	294	252	546

Table 2. DLLs with DLDs (N = 14) by age group and gender

Age	Girls	Boys	TOTAL
9.0-10.11	3	3	6
11.0-12.11	2	6	8
TOTAL	5	9	14

Table 3. Means, Standard Deviations and Cronbach's Alpha Values for the Variables of L1 and L2 Test Scores of Typical Developing DLLs (*N* = 546)

	Sentence Repetition		**Word Span**		**Grammar**		**Expressive Vocabulary**		**Receptive Vocabulary**	
Age	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2
	α .87	α .80	α .79	α .73	α .91	α .91	α .90	α .91	α .99	α .93
6.0–6.11	4.42 (*SD* = 26) *n* = 24	3.23 (*SD* = 1.7) *n* = 48	7.21 (*SD* = 2.4) *n* =14	4.05 (*SD* = 2.0) *n* = 39	52.09 (*SD* = 15.6) *n* = 22	52.81 (*SD* = 11.2) *n* = 47	6.23 (*SD* = 3 5 *n* = 13	4.27 (*SD* = 1.8) *n* = 40	61.33 (*SD* = 28.6) *n* = 24	48.04 (*SD* = 16.9) *n* = 48
7.0-711	4.55 (SD = 2.9) n=74	4.02 (*SD* = 1.9) *n* = 86	6.67 *SD* = 2.4) *n* =66	5.03 (*SD* = 2.1) *n* = 78	57.97 (*SD* = 11.2) *n* = 72	58.40 (*SD* = 10.7) *n* = 86	6.70 (*SD* = 3.6) *n* = 67	5.57 (*SD* = 2.3) *n* = 79	69.00 (*SD* = 29.8) *n* = 69	59.49 (*SD* = 18.4) *n* = 86
8.0-8.11	5.13 (*SD* = 2.9) *n* = 89	4.70 (*SD* = 2.2) *n* = 84	7.25 (*SD* = 2.5) *n* = 80	5.61 (*SD* = 2.0) *n* = 79	62.36 (*SD* = 10.1) *n* = 85	62.74 (*SD* = 9.7) *n* =84	8.08 (*SD* = 3.9) *n* = 79	7.49 (*SD* = 3.7) *n* = 80	76.03 (*SD* = 30.7) *n* = 86	69.71 (*SD* = 17.0) *n* = 84
9.0-9.11	5.61 (*SD* =3.2) *n* = 97	4.56 (*SD* = 1.8) *n* = 93	7.86 (*SD* = 2.5) *n* = 91	5.67 (*SD* = 2.1) *n* = 88	64.44 (*SD* = 9.9) *n* = 93	67.01 (*SD* = 7.0) *n* =92	9.47 (*SD* = 5.4) *n* = 91	8.45 (*SD* = 3.6) *n* = 88	83.76 (*SD* = 28.8) *n* = 97	74.60 (*SD* = 17.6) *n* = 93
10.0-10.11	6.86 (*SD* = 3. 8) *n* = 65	5.90 (*SD* = 2.3) *n* = 111	8.01 (*SD* = 2.2) *n* = 90	6.43 (*SD* = 1.8) *n* = 106	67.49 (*SD* = 9.6) *n* = 100	69.95 (*SD* = 6.8) *n* = 110	10.88 (*SD* = 4.8) *n* = 89	10.70 (*SD* = 3.5) *n* = 107	92.79 (*SD* = 27.7) *n* = 100	85.02 (*SD* = 18.1) *n* = 110
11.0-11.11	7.76	6.42	8.86	6.47	72.222 (*SD* = 5.6)	72.53	12.56	12.85	109.98	96.12

	(*SD* = 3.3) *n* = 68	(*SD* = 2.8) *n* = 59	(*SD* = 2.2) *n* = 64	(*SD* = 1.9) *n* = 59	*n* = 64	(*SD* = 5.0) *n* = 60	(*SD* = 4.87) *n* = 64	(*SD* = 4.5) *n* = 59	(*SD* = 16.6) *n* = 63	(*SD* = 17.2) *n* = 60
12.0-12.11	5.13 (*SD* = 2.9) *n* = 89	6.86 (*SD* = 2.8) *n* = 64	8.93 (SD = 2.1) N = 70	6.76 (*SD* = 1.9) *n* = 62	70.96 (*SD* = 8.5) *n* = 69	72.14 (*SD* = 6.3) *n* = 64	13.75 (*SD* = 5.1) *n* = 69	13.52 (*SD* = 3.9) *n* = 62	103.33 (*SD* = 22.0) *n* = 70	97.75 (*SD* = 16.1) *n* = 64

Table 4. Means, Standard Deviations and *N* for Variables of L1 and L2 Test Scores for DLLs without (*N* = 26) and with (*N* = 14) DLDs

		Sentence Repetition		**Word Span**		**Grammar**		**Expressive Vocabulary**		**Receptive Vocabulary**	
	Age	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2
At-risk	**9.0–10.11 months 110.50 (*SD* = 2.83)**	2.21 (*SD* = 2.1) *N* = 14	2.57 (*SD* = 1.2) *N* = 14	5.71 (*SD* = 1.8) *N* = 14	4.14 (*SD* = 2.3) *N* = 14	58.46 (*SD* = 10.8) *N* = 14	57.64 (*SD* = 12.5) *N* = 14	4.86 (*SD* = 2.9) *N* = 14	5.79 (*SD* = 3.7) *N* = 14	76.00 (*SD* = 29.3) *N* = 14	65.57 (*SD* = 19.2) *N* = 14
DLDs	**9.0–10.11 Months 116.33 (*SD* = 9.56)**	3.33 (*SD* = 2.9) *N* = 6	3.67 (*SD* = 0.8) *N* = 6	5.83 (*SD* = 2.2) *N* = 6	5.33 (*SD* = 2.5) *N* = 6	60.17 (*SD* = 10.5) *N* = 6	59.00 (*SD* = 14.2) *N* = 6	5.33 (*SD* = 4.2) *N* = 6	4.83 (*SD* = 2.2) *N* = 6	66.50 (*SD* = 27.0) *N* = 6	66.33 (*SD* = 12.8) *N* = 6
At-risk	**11.0–12.11 months 133.50 (*SD* = 1.00)**	2.36 (*SD* = 1.8) *N* = 12	3.64 (*SD* = 1.5) *N* = 12	6.36 (*SD* = 2.2) *N* = 12	4.73 (*SD* = 1.7) *N* = 12	65.08 (*SD* = 9.2) *N* = 12	68.36 (*SD* = 4.2) *N* = 12	7.0 (*SD* = 3.0) *N* = 12	7.83 (*SD* = 6.0) *N* = 12	82.3 (*SD* = 21.3) *N* = 12	80.5 (*SD* = 15.3) *N* = 12
DLDs	**11.0–12.11 months 145.50 (*SD* = 4.75)**	2.71 (*SD* = 2.1) *N* = 8	4.14 (*SD* = 1.1) *N* = 8	6.00 (*SD* = 2.5) *N* = 8	4.57 (*SD* = 1.7) *N* = 8	62.88 (*SD* = 10.4) *N* = 8	67.13 (*SD* = 4.2) *N* = 8	6.25 (*SD* = 2.6) *N* = 8	4.88 (*SD* = 4.1) *N* = 8	74.0 (*SD* = 20.8) *N* = 8	73.38 (*SD* = 7.5) *N* = 8

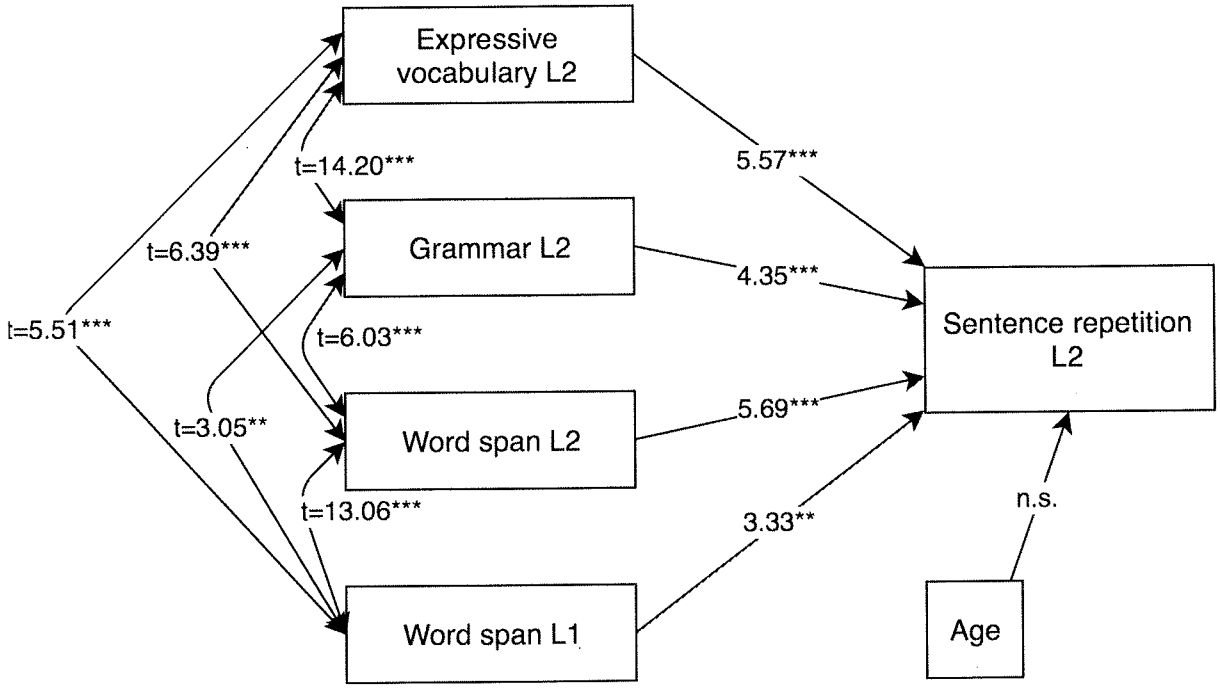

Figure 1 Model of L2 sentence repetition and linguistic measures (expressive vocabulary and grammar) and phonological short- term memory (word span)

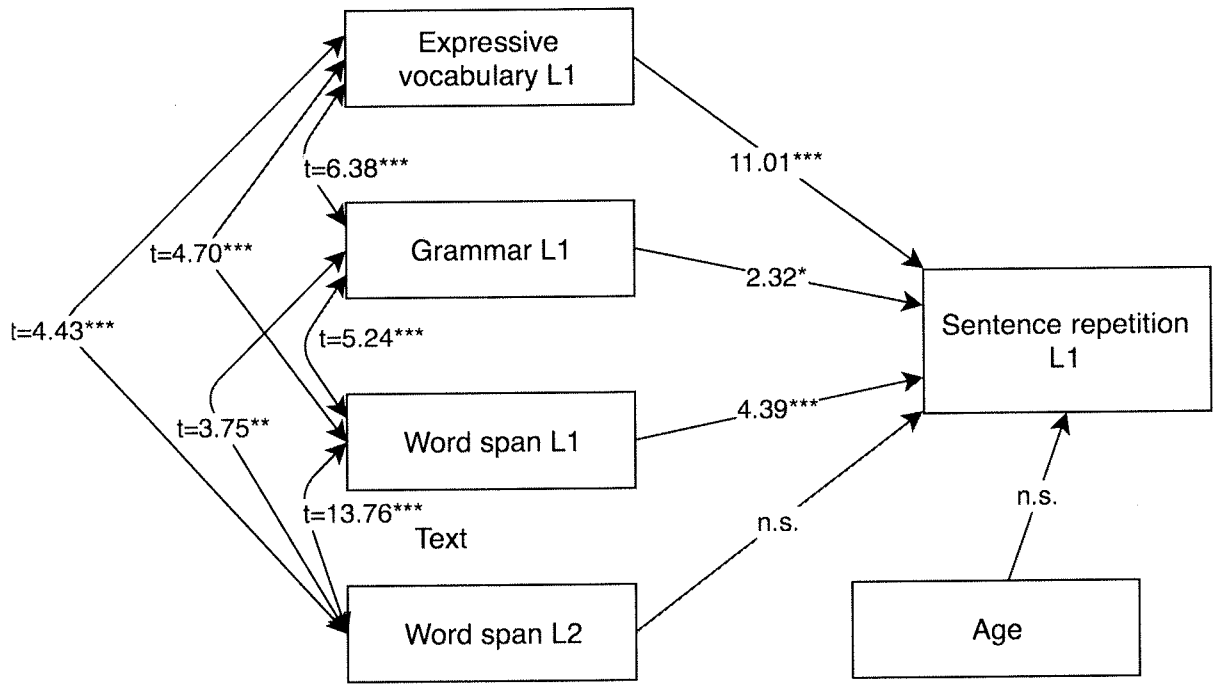

Figure 2 Model of L1 sentence repetition and linguistic measures (expressive vocabulary and grammar) and phonological short- term memory (word span)

Appendix 1. Additional Information: Measures

Sentence repetition, word span and expressive vocabulary are subtests from an individual administered language screening test: *Language 6-16 (Språk 6-16 (Ottem & Frost, 2005)*. The screening test is a nonprofit test distributed by Statped: a national service for special needs education, Department of Speech and Language Disorders, in Norway.

Special needs services and schools' speech therapists, school psychologists and special needs teachers frequently use *Language 6-16* to assess students with suspected DLDs. The norming data encompass 1,214 monolingual students with no suspected special needs from all over Norway.

As noted in the article, qualified translators with native competencies in L1 and the students' specific bilingual and bicultural context translated all the items in the included measures.

Sentence repetition (Language 6-16)
The content, grammar and syntax complexity of Norwegian versions of the sentences were changed in the L1 versions to preserve the sentences' difficulty and complexity. The sentences' length, number of syllables, syntactic and grammatical complexity, and information units in Norwegian were translated and adapted to L1 versions of the sentences. The structures tested in each language should reflect their relative difficulty level in that language, not the relative difficulty of the tested structures in Norwegian.

These adaptations were needed to preserve the relative difficulty in each L1 version. Features that were culturally biased in the Norwegian sentences were changed, and personal names were chosen independently for each language.[1]

Item No.	English translation	Norwegian wording
1	He ran out again.	Han løp ut igjen
4	The girl kicked the football over the roof.	Jenta sparket fotballen over hustaket
6	Pears in my garden are better than those in the shops.	Pærene i hagen min er bedre enn de i butikken
9	She is so good at playing guitar that she can play in a rock band.	Hun er så flink til å spille gitar at hun kan bli med i et rockeband
13	The boy who skipped training was not allowed to play on the team for one week	Gutten som skulka treningen, fikk ikke lov å spille på laget før det var gått en uke

We calculated the percentage of students who responded correctly to each item and noted a steady decrease in the percentage of correct responses in L1 and L2 as the complexity increased. One item in the L1 and L2 test was more difficult than the rest;[2] however, it did not lead to a situation in which students were forced to stop the test.

Word span (serial recall by words): The words in this test were monosyllabic, and high-frequency words in the age group were included, similar to the Norwegian version of the test. We calculated the percentage of students who correctly identified the items from each series and noted a steady decline in the percentage of correctly identified series as the test progressed. The researchers observed this trend in both L1 and L2 assessment procedures.

Grammar (TROG-2, Test for Reception of Grammar): The complexity of the 20 structures, (four items in each block; 80 items) increases as the test progresses. For this study, the researchers translated the sentences into the six (L1) home languages with corresponding syntactic structures found in all languages. It was not problematic to translate the different structures into the six home languages, but in some of the languages, the structure was more common than in others, and in some languages, one structure was more difficult than other structures. Accordingly, the sample with and the sample without DLDs were assessed using all 20 structures, and 80 items in L1 and L2.

Expressive vocabulary (Language 6-16): This test's translated versions contained words that were as close as possible to the originals in terms of degree of difficulty. Translating vocabulary tests can pose challenges because words tend to differ in concreteness and frequency across languages. The expertise of the translators' native competence in the languages and the participants' specific bilingual and bicultural expertise were crucial. We calculated the percentage of students who responded correctly to each item on the antonym and definition tests. A reasonable decrease in the percentage of correct items for L1 and L2 vocabulary items was observed. However, two items in the antonym subtest did not show a steady decrease.[3] Nevertheless, no students stopped the test. The definition section showed a steady decrease in the percentage of correct items.

[1] E.g., 'Ali' in Urdu, 'Arani' in Tamil and 'Chau' in Vietnamese.

[2] Item 5

[3] Item 3 was more difficult than Items 4 and 5, and Item 6 was more difficult than Items 7 and 8 in the Norwegian version and for all L1 language groups.

Receptive vocabulary (British Picture Vocabulary Test II): As noted in the paper, each student started with the first item on the test (Item 1, Block 1) provided in both languages because of the wide variations in vocabulary knowledge in DLLs generally. We calculated the percentage of students who passed each block. Researchers noted a steady decrease in the percentage of students who progressed to the end of the 12 items in the 12 blocks in each of the six languages assessed. The researchers observed this trend in both the L1 and L2 assessments. The findings indicate that it is unlikely for a student to master a subsequent block if he/she failed the previous one.

Appendix 2. Correlations Between L1 and L2 Sentence Repetition, Word Span, Grammar and Expressive and Receptive Vocabulary

	L1 Sentence Repetition	**L2 Sentence Repetition**	**L1 Word Span**	**L2 Word Span**	**L1 Grammar**	**L2 Grammar**	**L1 Expressive Vocabulary**	**L2 Expressive Vocabulary**	**L1 Receptive Vocabulary**
L1 Sentence Repetition									
L2 Sentence Repetition	.279**								
L1 Word Span	.464**	.412**							
L2 Word Span	.326**	.544**	.494**						
L1 Grammar	.400**	.285**	.292**	.314**					
L2 Grammar	.239**	.552**	.281**	.446**	.335**				
L1 Expressive Vocabulary	.649**	.301**	.392**	.350**	.523**	.347**			
L2 Expressive Vocabulary	.328**	.603**	.368**	.472**	.415**	.618**	.468**		
L1 Receptive Vocabulary	.383**	.250**	.263**	.321**	.538**	.393**	.523**	.435**	
L2 Receptive Vocabulary	.158**	.538**	.301**	.447**	.342**	.691**	.310**	.748**	.379**

Appendix 3. Monolingual Norms ($N = 900$) by Age Group, Sentence Repetition, Cronbach's Alpha, Raw Scores and SD

Age Group	Sentence Repetition α 81
6.0–6.11	6.45 ($SD = 2.26$) $n = 432$
9.0–10.11	8.28 ($SD = 2.52$) $n = 223$
11.0–12.11	9.43 ($SD = 2.86$) $n = 245$
TOTAL	$N = 900$