languages

Article

Dynamic Assessment of Word Learning to Diagnose Developmental Language Disorder in French-Speaking Monolingual and Bilingual Children

Mélodie Matrat *, Hélène Delage and Margaret Kehoe

Faculty of Psychology and Educational Sciences, University of Geneva, 1205 Geneva, Switzerland; helene.delage@unige.ch (H.D.); margaret.winkler-kehoe@unige.ch (M.K.)
* Correspondence: melodie.matrat@unige.ch; Tel.: +41-22-379-0914

Abstract: Dynamic assessment (DA), which evaluates the learning process rather than static knowledge, has been found to be effective in diagnosing developmental language disorder (DLD) in English-speaking bilingual children. We present three studies that examine whether a French dynamic word learning task can distinguish caseload children with DLD from control children with typical development (TD). Forty-eight monolingual and 69 bilingual French-speaking children, aged four to nine, were required to learn three to six non-words and their semantic characteristics. DA consisted of three phases: (1) a teaching phase in which non-words were taught to the child; (2) an immediate test phase, with graduated prompts, in which children were required to identify and produce the target words and their semantic features; and (3) a delayed test phase. Global results indicated that there were no differences between monolingual and bilingual TD children on the DA whereas there were differences on the static assessment of vocabulary knowledge: bilinguals performed less well than monolinguals. In addition, DA differentiated control and caseload monolingual children. Further work is needed to develop a dynamic word learning task, administered in one session, which can even more accurately differentiate TD and DLD bilingual children.

Keywords: dynamic assessment; word learning; developmental language disorder; bilingualism; children

check for updates

Citation: Matrat, Mélodie, Hélène Delage, and Margaret Kehoe. 2022. Dynamic Assessment of Word Learning to Diagnose Developmental Language Disorder in French-Speaking Monolingual and Bilingual Children. *Languages* 7: 181. https://doi.org/10.3390/languages7030181

Academic Editors: Cornelia Hamann and Lina Abed Ibrahim

Received: 28 March 2022
Accepted: 5 July 2022
Published: 14 July 2022

1. Introduction

In Switzerland, as in other countries, the number of bilingual children has increased in recent years. The latest report on language practices in Switzerland (Office fédéral de la statistique 2021) indicates that, "68% of the population regularly use more than one language", a percentage which has risen since 2014. Consequently, speech-language therapists (SLTs) see more and more bilingual children and are frequently required to evaluate their speech, although they do not speak the children's other languages. They are also faced with a lack of tools to appropriately assess the children's language because standardized tasks contain many biases, which do not make them appropriate for use with bilingual children (De Lamo White and Jin 2011; Paradis et al. 2021). Dynamic assessment (DA) is one of the approaches that has been developed to address these shortcomings and it has already proven its worth in the assessment of English-speaking bilingual children (see Orellana et al. 2019 for a meta-analysis). However, there are no such DA tools in French. The goal of this research is to develop a French DA of word learning that can distinguish typically developing (TD) children from children with developmental language disorder (DLD), particularly bilinguals.

Bilingual children are often defined as children who "receive regular input in two or more languages during the most dynamic period of communication development" (Kohnert 2010, p. 457). In this study, we use the term bilingual to refer to children who speak two or more languages. More specifically, the bilingual children in this study include simultaneous bilinguals, who began acquiring both languages in infancy (before 3 years of age), and consecutive (or sequential) bilinguals, who receive exposure to their second language after infancy, often at school entry (after 3 years of age) (Kohnert 2010).

1.1. Developmental Language Disorder and the Assessment of Bilingual Children's Language

Children with DLD experience difficulties in understanding and producing language that impact their everyday life and social relations (Bishop et al. 2017). Bilingual children with DLD have difficulties in both of their languages (Kohnert 2010; Marini et al. 2019). According to Bishop et al.'s (2017) position statement, the prevalence of language problems in children is between 3 and 7%.

There are many challenges associated with assessing the language of bilingual children, some of which have been highlighted in recent surveys examining the practices of SLTs when working with bilinguals (Arias and Friberg 2017; Caesar and Kohler 2007). An improvement in the methods used to assess bilinguals has been found between surveys conducted in 2007 and 2017 with for example a greater focus on testing the child's native language, although only 60% of SLTs indicate that they "often" perform complete assessments in both the child's languages, despite the fact this has been shown to be essential (Nayeb et al. 2021). These findings, nonetheless, refer to English-speaking SLTs and we have little available data on the practices of French-speaking SLTs when assessing bilinguals (see, however, Stanford et al. Forthcoming).

Standardized tasks continue to be much employed by SLTs to diagnose language disorders (Arias and Friberg 2017). However, these tasks are not suitable for bilingual populations because they have multiple biases (De Lamo White and Jin 2011), including content bias (when the content of a task is more difficult for one group than for the other because of different opportunities and life experiences associated with cultural diversity), linguistic bias (when criteria of the assessment do not make allowances for linguistic or dialectal variation between languages), and normative bias (when tasks are not normed on the same type of population, that is, mostly normed on monolingual children). To remove these biases, De Lamo White and Jin (2011) propose content and format changes to standardized tasks (e.g., the rephrasing of instructions or the use of more familiar materials) or re-standardization of tasks with larger sample populations including minority groups (such as bilinguals). Nevertheless, these solutions take time and can be difficult to implement. The accumulation of these problems has led to an over-identification of bilingual children as having DLD or to an under-identification, which may have the adverse effect of proposing rehabilitation to a child who does not need it or leading to delayed intervention (Grimm and Schulz 2014). Recognition of these problems has led to the development of a set of tools for Language Impairment Testing in Multilingual Settings (LITMUS), an initiative that stems from the Cost Action project IS0804 (Armon-Lotem et al. 2015). While acknowledging the importance of the LITMUS tools, which include processing measures such as nonword and sentence repetition and assessment instruments for narratives, we consider another type of assessment, namely DA. It has been proposed as a complementary or alternative approach for reducing the cultural and linguistic biases when assessing bilinguals (De Lamo White and Jin 2011; Paradis et al. 2021; Peña et al. 2001).

1.2. Dynamic Assessment

DA measures the child's learning potential by evaluating how much s/he learns rather than what s/he knows (Kapantzoglou et al. 2012). It is based on Vygotsky's sociocultural theory (Vygotsky 1978), a central concept of this theory being the "zone of proximal development", which is the distance between the child's "actual developmental level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" (Vygotsky 1978, p. 86). DA is particularly relevant for bilingual children because it is not based on previous language experiences but rather on the children's current learning abilities. During DA, the examiner teaches the child how to do a task; support is provided to help the child to learn. DA has two main formats or approaches: test-teach-retest and graduated prompting. The test-teach-retest format consists of a pre-test, in which the child's existing level of knowledge is evaluated, a teaching phase, in which the examiner explains to the child how to succeed at the task and teaches him/her strategies to improve his/her performance, and a post-test, in which gains, maintenance, and transfer as a result of the learning are evaluated. Graduated prompting is a more structured approach in which prompts (such as phonological or semantic cues) are given during the assessment to help the child succeed at the task (Campione and Brown 1987). Prompts are provided in hierarchical order, so levels of support increase progressively. To complete this assessment, modifiability scales can be rated by the examiner. The two most common are the Learning Strategies Checklist (LSC, Lidz 1991; Peña 1993), which assesses the child's responsiveness (i.e., attention, planning, self-regulation, and motivation) and transfer of knowledge or skills, and the Modifiability Scale (MS, Lidz 1987, 1991), which indicates the effort the examiner made and the support the child needed during the DA.DA has been shown to differentiate bilingual children with no real language difficulties from those with language disorders (see Orellana et al. 2019 for a meta-analysis). Such diagnostic accuracy of DA has been found for narratives skills (Petersen et al. 2017), labeling skills (Peña et al. 2001), and receptive vocabulary (Camilleri and Law 2007). Other research has also suggested that DA is useful in predicting language development and in informing intervention (Hasson and Joffe 2007). Despite these promising findings, DA is not often used in the clinical setting: in a recent survey, 76% of SLTs indicated that they were not familiar with DA and 90% that they did not use it in their assessment practice (Delage et al. 2021).

Thus, there is a need to develop new DA approaches that can be easily used in clinical practice. This study focuses on DA of lexical skills because assessing the lexical domain in bilingual children is particularly relevant for DLD diagnosis (Marini et al. 2019).

1.3. Vocabulary Development in Bilingual Children

It is well known that vocabulary knowledge is essential for everyday comprehension, academic achievement, and literacy acquisition (Morgan et al. 2015; Ouellette 2006). Countless studies indicate that bilinguals perform less well than monolinguals on vocabulary tasks, be they receptive or expressive (Hoff et al. 2014; Oller et al. 2007). For example, in an aggregate analysis (combining data from different studies) of performance on the Peabody Picture Vocabulary Test (PPVT-III; Dunn and Dunn 1997), which included 1738 children aged 3 to 10, Bialystok et al. (2010) reported that monolingual children had significantly higher scores than bilingual children at each age. More precisely, the distributions of the scores for both groups were in the normal range, but the overall mean of the monolinguals was higher than the mean of bilinguals. Even when the authors separated the bilinguals into those who spoke an East Asian language and those who did not, they still found that the two subgroups of bilinguals performed less well than monolinguals. Moreover, when word types were compared, monolinguals exhibited better performance than bilinguals on words frequently used in the home context but not on those used in school. Bialystok and colleagues concluded that vocabulary differences were principally due to bilingualism, that is, to lack of exposure to the language of the environment, and to the influence of context (e.g., home vs. school). Studies not only show that bilinguals as a group perform less well than monolinguals, they also show that consecutive bilinguals perform less well than simultaneous bilinguals. For example, Hammer et al. (2008) found that bilingual children who begin to learn English at the entry of Head Start[1] (i.e., consecutive bilinguals) had lower performance on English receptive vocabulary tasks, for at least the first two years, than bilingual children who were exposed to English since birth (i.e., simultaneous bilingual).

There are at least three reasons for the poor performance of bilinguals on vocabulary tasks: 1. Distributed nature of bilingual knowledge; 2. Reduced language exposure; and 3.

Socio-Economic Status (SES) effects, which may influence bilinguals in particular because they often belong to minority groups. First, bilingual children divide their time between two language environments, which leads to distributed vocabulary knowledge across their languages. For some concepts, children know the words in their two languages, that is, they possess "translation equivalents" or "doublets". For other concepts, however, children know the words only in one language, due to context-specific learning, that is, they possess only "singlets". As noted above, the bilingual children of Bialystok et al.'s (2010) study had acquired a rich English vocabulary in the context of school but not in the context of home. Thus, low scores in vocabulary assessment, if assessed in only one language, do not indicate that bilingual children are poor learners, but that their vocabulary is distributed across their languages (Oller et al. 2007). Second, language exposure or language input has a significant influence on rate of vocabulary learning (Gathercole et al. 2016; Hammer et al. 2008; Thordardottir 2011). Thordardottir (2011), for example, documented a strong relationship between amount of exposure and receptive and expressive vocabulary levels with the strongest effects for expressive vocabulary. This author noted that bilingual children must have at least 60% exposure to the test language in order to perform as well as monolinguals for expressive vocabulary. Third, it is important to note that SES also plays a role in vocabulary development. Children from low SES have smaller receptive vocabularies than children from high or mid SES (Gathercole et al. 2016). This effect is well documented in monolingual populations and is likely to play a role in bilingual populations given the large proportion of bilinguals who are migrants or who belong to minority groups.

Because vocabulary is distributed across languages and is influenced by language exposure and SES, it is particularly difficult to assess the lexical knowledge of bilinguals by common vocabulary tasks. Some researchers have investigated other ways to test lexical knowledge in bilinguals, in particular by examining their abilities to learn new words.

1.4. Word Learning

Word learning is a much-studied area of language development (see for example meta-analyses of Flack et al. 2018; Kan and Windsor 2010). Different theories have been developed to explain the complex and challenging process of learning a word. Carey and Bartlett (1978) first introduced the term "fast mapping" to describe young children's ability to hear a word and create, with very few exposures, an initial representation of its phonological and semantic information. In this initial process, children form preliminary links between words and referents. This is followed by the process of "slow mapping" during which, with repeated exposures, children develop more robust phonological and semantic representations of new words and thereby strengthen the links between these representations. According to He and Arunachalam (2017), learning a word occurs in situations with rich cues like observational, social, or linguistic contexts. A learner has to develop strategies and capacities such as to rely on cognitive biases, language parsing, and social-pragmatic abilities to create the correct representation of a new word. Despite the fact that word learning is a challenging task, children with TD are able to understand and produce new words at an early age.

Children with DLD, however, typically have difficulties learning new words (see Gray 2004 and meta-analysis by Kan and Windsor 2010). It has been suggested that this poor word learning performance may be related to difficulties in creating and storing phonological and semantic representations of new words and/or building links between them (Alt and Plante 2006). Children with DLD require more exposure than TD children to reach the same level of learning (Gray 2004). Moreover, learning the word's form (phonological representation) appears to be more difficult than learning the word's meaning (semantic representation), especially for children with DLD (Gray 2004).

Only a few studies have evaluated in a dynamic manner the word learning capacities of bilingual children (e.g., Hasson et al. 2013; Kapantzoglou et al. 2012). One of them is Hasson et al. (2013) who developed the DAPPLE (Dynamic Assessment of Preschoolers'

Proficiency in Learning English), a dynamic screening of language skills, which contains a subtest devoted to word learning. The task follows a test-teach-retest design, based on a procedure previously developed by Camilleri and Law (2007). The pre-test consists of administration of a receptive vocabulary task in which six unfamiliar words are selected to be targeted. The teaching phase takes the form of a posting game in which the child has to post the correct cards (picture of the unknown target item alongside two known distractors) into a post box. Prompting is given if the child is unable to select the correct item. After teaching three items, a post-test evaluates expressive skills: the child has to name the pictures s/he has posted. This test is repeated for a second evaluation of expressive retention (i.e., delayed post-test). Results indicated that bilingual caseload children (i.e., bilinguals receiving speech-language therapy), aged 3–5 years, required more assistance than those in the control group to identify the target item in the teaching phase. No significant differences were found between the two groups in the first expressive task but, in the second expressive task (i.e., the delayed post-test), the caseload group performed less well than the control group.

Another study of dynamic word learning is Kapantzoglou et al.'s (2012), which assessed Spanish-English bilingual children, with TD and DLD, aged 4–5 years, in a test-teach-retest format. Children had to learn three novel words. Words were presented 9, 18, or 27 times. Children with DLD obtained significantly poorer results than children with TD for word identification after having received nine exposures. They did not differ from TD children for word identification after repeated exposures (i.e., 18 or 27 times) nor did they differ for word production. Despite these relatively modest results, Kapantzoglou et al. (2012) argue that their DA task holds potential for accurately differentiating children with TD from children with DLD.

Apart from studies conducted within the DA framework, some authors have compared monolingual and bilingual children's word learning in order to determine whether bilingual children display different processes than those of monolinguals. Alt et al. (2013), for example, found that school-aged bilingual children had nearly identical performance to monolingual peers: the bilingual children showed equivalent performance on word identification but were less accurate in word production. In a more recent study, Alt et al. (2019) proposed various word learning situations to monolingual and bilingual children. The bilinguals differed from the monolinguals only in some conditions. For example, the bilingual children were less able to detect mispronunciations of novel words, which may relate to their greater exposure to non-native speech and variable input forms.

To sum up, the assessment of word learning skills using DA seems to be a more promising method for the identification of DLD in bilingual children than the assessment of vocabulary knowledge using vocabulary tasks. The use of DA, which is not sensitive to the child's language experience and cultural background (Gutiérrez-Clellen and Peña 2001), should minimize biases and, thus, enhance the identification of children with DLD. Nonetheless, no study has investigated the diagnostic potential of a dynamic word learning task for French-speaking children.

1.5. The Current Study

In a series of three studies, we aim to develop a DA of word learning and to examine its feasibility and validity in French-speaking monolingual and bilingual children. Our main question is: Can DA of word learning distinguish TD monolingual and bilingual children from those with DLD?

The first study compared TD monolingual and bilingual children on both static and dynamic measures to determine whether DA was more appropriate than static tasks. The second study compared the DA results of bilingual children who were receiving speech-language therapy (caseload group) with those who were not (control group). The last study combined features of the first two by comparing performance of monolingual and bilingual—control and caseload—children on static and dynamic measures. These studies were approved by the University Commission for Ethical Research in Geneva and all parents gave their informed and written consent for the participation of their children.

2. Study 1

The first study investigated whether a dynamic word learning task better assesses the lexical skills of TD bilingual children than a static standardized vocabulary task. We predict that bilingual children will perform less well than monolingual children on the static task, but that they will perform similarly to monolinguals on the DA.

2.1. Method

2.1.1. Participants

Thirty TD children (23 girls, 7 boys) participated in the study (see Table 1 for a description). They were divided into three groups depending on their language background: 10 monolingual (ML) French-speaking children, 9 simultaneous bilingual (SBL) children, and 11 consecutive bilingual (CBL) children. All children were between the ages of 4;1 and 6;8 years, with a mean of 5;8 (SD = 8 months). The mean age did not differ across groups (H (2,30) = 3.99, p = 0.14). In all studies, bilingual status was determined based on a parent questionnaire in which the parents indicated whether their child spoke another language at least 30% of the time in addition to French. They were required to indicate which language the child spoke at home, and at which age the child had acquired French. In this study, the other languages spoken by the children were Portuguese (n = 5), Spanish (*3*), Italian (*3*), Arabic (*2*), English (*1*), Albanian (*1*), Japanese (*1*), Czech (*1*), Russian (*1*), Polish (*1*), and Uzbek (*1*). We excluded children with hearing or visual impairments, neurological disorders, DLD or other disorders (e.g., autism spectrum disorder, attention deficit hyperactivity disorder, intellectual disability).

Table 1. Participant information—Study 1.

Groups	Monolinguals (ML)	Simultaneous Bilinguals (SBL)	Consecutive Bilinguals (CBL)
N	10	9	11
Age range (years; months)	4;1–6;5	5;3–6;5	4;10–6;8
Mean age SD (months)	5;3 9.3	5;10 5.3	5;10 7.3
Gender	8 G, 2 B	7 G, 2 B	8 G, 3 B
Raven–Mean standard scores (SD)	0.62 (1.14)	0.08 (0.997)	−0.11 (1.08)

G = Girl; B = Boy.

2.1.2. General Procedure

Children were recruited from public schools in Geneva and were tested individually in a single session of 45 minutes in their school. The following tasks were proposed in this order: (a) static receptive and expressive vocabulary tasks using the standardized N-EEL subtests ("*Nouvelles Epreuves pour l'Evaluation du Langage*", a French language assessment battery, Chevrier-Muller and Plaza 2001); (b) a teaching phase of DA; (c) an immediate post-test of DA; (d) Raven's matrices (Raven 1998); and (e) a delayed post-test of DA. A master's student in speech-language therapy at the University of Geneva administered the different tasks.

2.1.3. Tasks

- Standardized vocabulary tasks

The static tasks included receptive and expressive vocabulary subtests of the N-EEL (Chevrier-Muller and Plaza 2001). In the expressive task, the child had to name 36 colored pictures. Two points were awarded for a correct answer, resulting in a maximum production score of 72 points. In the receptive task, the child was presented with a set of eight pictures and had to point to the picture named by the examiner. One point was awarded if the child pointed to the correct picture, resulting in a maximum receptive score of 36. Standard scores were calculated in order to compare the results between groups.

- Dynamic assessment

Our DA task was based on Kapantzoglou et al. (2012) and Hasson et al.'s (2013) procedures and combined a test-teach-retest as well as a graduated prompting format. The experimenter presented a puppet who came from another planet. This puppet had a suitcase with objects that it wanted to show to the child. The objects consisted of five real objects (plastic banana, plush cat, plastic sheep, pencil, and ball) and three novel objects that were associated with three non-words (see Appendix A for more details). The non-words—/nefa/, /gopi/ and /tibak/—contained simple and common syllable structure (i.e., CV or CVC, see Maddieson et al. 2014), early acquired consonants (MacLeod et al. 2011), and low neighborhood density (i.e., few real words differed from the non-word by the addition, omission, or modification of a phoneme), as determined by the French lexical database "Lexique" (New et al. 2004).

Six objects (three real and three novel) were presented to the child. Because we used (novel) non-words and novel objects, we could not perform a pre-test. We started directly with the teaching phase. The experimenter named each novel object and provided information on its perceptual and functional properties. As an example, the experimenter said: "*Look, it's a /tibak/. I like the /tibak/ because I use it to talk to my friends on other planets. Here, look at the /tibak/.*" During teaching, the names of each novel object were presented nine times and the child had to repeat the name once. During post-tests, the experimenter asked the child to name each target object, which provided a score of novel word production. If the child could not name the target word or made an error, prompts were given (see below). After each target word was named, the puppet asked the child to help him tidy up his things. Each object was named by the experimenter and the child was asked to give the object to the puppet, thus providing a score of novel word comprehension. If the child was unable to identify the target object, prompts were also given. A delayed post-test took place after administration of the Raven's matrices, that is, after a delay of about 10 min.

- Prompts and scoring

Prompts in the production tests were phonological cues. Scores ranged from 0 to 3 for each target. If the child could say the word without cues, s/he would receive 3 points; if s/he could say the word with the first prompt (i.e., the initial sound), s/he would receive 2 points; if s/he could say the word with the second prompt (i.e., the initial syllable), s/he would receive 1 point, and if the child could not provide the target word, s/he would receive no points. For the comprehension tests, prompts were based on the contextual cues of the DAPPLE, which involved the child using a process of elimination to determine the correct item (Hasson et al. 2013). Scores ranged from 0 to 4 for each target. If the child could point to the object without cues, s/he would receive 4 points; if s/he pointed to a non-target object, the object was named and removed from view, and one point was deducted. This was repeated until the correct object was identified or until only the target object remained, the score being then 0 points.

- Raven's matrices

Raven's matrices are a non-verbal reasoning task consisting of 36 sheets of matrices of different patterns and colors. The participant must correctly select the missing part. We included this measure to ensure that there were no differences in non-verbal reasoning between the three groups. Indeed, the Kruskal–Wallis test, conducted on the standard scores, indicated that there were no between-group differences (H (2,30) = 2.48, $p = 0.29$).

2.1.4. Data Analysis

Analyses were performed using R statistical software (R Core Team 2020). The independent variable was Group (ML, SBL, or CBL) and the dependent variables were scores on the static tasks (expression and reception) and DA (prompt scores in production and in comprehension on immediate and delayed post-tests). Due to the small number of participants in this study, non-parametric (Kruskal–Wallis) tests were used to compare results of the three groups. In the case of significant differences, pairwise Wilcoxon rank sum tests with Bonferonni correction for adjusting *p* values were conducted to compare groups.

2.2. Results

Table 2 presents the results on static and dynamic tasks for the three groups.

Table 2. Means (SD) of static and dynamic scores for the three groups—Study 1.

Groups			ML	SBL	CBL
N			10	9	11
Static vocabulary tasks (standard scores)		**Expression**	0.68 (0.85)	−0.70 (1.11)	−1.87 (1.25)
		Reception	0.31 (0.67)	−2.06 (1.09)	−3.29 (3.18)
DA: Prompt scores	**Production (/9)**	**Immediate**	1.20 (1.40)	1.56 (1.51)	0.91 (1.92)
		Delayed	1.20 (2.10)	0.75 (1.16)	1.50 (2.92)
	Comprehension (/12)	**Immediate**	11.70 (0.48)	11.56 (0.73)	11.27 (1.19)
		Delayed	11.50 (0.67)	11.63 (0.74)	11.20 (0.92)

ML = Monolingual; SBL = Simultaneous Bilingual; CBL = Consecutive Bilingual.

2.2.1. Static Tasks

For the static tasks of vocabulary knowledge (N-EEL subtests), Kruskal–Wallis tests indicated significant differences between groups for the expressive ($H\ (2,29) = 15.90$, $p < 0.001$) and receptive tasks ($H\ (2,30) = 18.58$, $p < 0.001$). Pairwise comparison tests revealed that the two bilingual groups had lower standard scores than the ML group: for expression: $U = 76$, $p = 0.038$ between ML and SBL; $U = 97.5$, $p = 0.001$ between ML and CBL; for reception: $U = 9$, $p = 0.001$ between ML and SBL; $U = 10$, $p < 0.001$ between ML and CBL. There were no differences between the two bilingual groups.

2.2.2. Dynamic Assessment

To remind the reader, the DA yielded scores for the number of prompts needed in the immediate and delayed post-tests of the production and comprehension post-tests. In the production post-tests, no significant differences were found across groups, neither for the immediate ($H\ (2,30) = 2.19$, $p = 0.34$), nor for the delayed post-test ($H\ (2,28) = 0.05$, $p = 0.97$). In other words, all children, monolingual and bilingual alike, needed the same number of prompts to produce the novel words in the two phases. Nevertheless, mean scores (see Table 2) show that few children were able to produce the target nonwords in the two post-tests, namely, scores were very low (floor effect). In the comprehension subtest, however, results revealed a ceiling effect, since all children succeeded without prompts or with very few prompts. Thus, we did not analyze the data further in this condition.

2.3. Discussion

This first study compared the performance of monolingual and bilingual children on static and dynamic tasks. We found that monolingual children obtained better results than bilingual children on the receptive and expressive subtests of a standardized vocabulary task. Thus, as previously shown (e.g., Camilleri and Law 2007; Hoff et al. 2014), static vocabulary tasks seem to penalize bilingual children who lack exposure to the language being assessed. In contrast, we did not find any differences between groups on the production scores of the DA task; this suggests that a dynamic word learning task indeed evaluates the learning process, without penalizing bilinguals (Camilleri and Law 2007; Peña et al. 2001).

Nonetheless, in the comprehension subtest, we observed a clear ceiling effect. All children succeeded in identifying the target objects. Furthermore, in the production post-tests, the scores were low, indicating that all children had difficulty learning the phonological form of the non-words. Thus, it seems that, for children aged four to six, a phonological representation associated with a novel object (referent) begins to be constructed after a few exposures to the word. This representation is available to the child to enable him/her to identify the correct object when the new word is named (i.e., retrieving the referent from the phonological form) but it remains too fragile for the child to access it when s/he has to retrieve the phonological form from the referent (i.e., the new object). This floor effect needs to be removed in order to effectively distinguish between TD and DLD. Reasons for the difficulty in retrieving the phonological form of the word can be of different kinds: low exposure to the word, low repetition/production by the child, and/or the teaching method (i.e., the context and manner in which non-words are taught to children). In the next study, we attempt to modify these elements to improve children's performance in word production. Moreover, if the phonological representation remains fragile after a few exposures to the new word, we investigate if the semantic features of the word can be more easily recalled by the child. We, therefore, add a test of semantic feature recall. According to Gray et al. (2020), both phonological and semantic components are involved in the initial stages of word learning.

3. Study 2

The second study evaluated the validity of the dynamic task with TD and DLD bilingual children. The main question was: Does a DA of word learning discriminate TD and DLD bilingual children? We hypothesized that TD bilingual children (control group) would have had better performance on our DA than bilingual children with DLD (caseload group).

3.1. Method

3.1.1. Participants

Thirty consecutive bilingual children (i.e., exposed to French after three years of age) participated in the study. They formed two groups: a "caseload" and a "control" group. We use the terms caseload and control instead of DLD and TD because of the risks of misdiagnosis in bilingual individuals (Tuller et al. 2013). According to parent and SLT reports, "caseload" participants were children who were receiving intervention for DLD. Three children in the control group and two in the caseload group were excluded because there was missing information on the parental questionnaire ($n = 3$) or because they did not complete all the tasks ($n = 2$). Thus, data from twenty-five children (11 girls, 24 boys) were analyzed: 15 children in the control group (CBL-CTRL) and 10 in the caseload (CBL-CSL) group (see Table 3 for a description). All children were between the ages of 4;7 and 8;4 years, with a mean age of 6;4 (SD = 10 months). The mean age did not differ across groups (U = 55.5, $p = 0.29$). The languages spoken by the participants, in addition to French, were English ($n = 6$), Arabic (*5*), Italian (*2*), Turkish (*2*), German (*2*), Kurdish (*2*), Portuguese (*2*), Korean (*1*), Russian (*1*), Spanish (*1*), and Tamil (*1*). There were no significant differences between duration of exposure to French between the CBL-CTRL and CBL-CSL groups (U = 80.5, $p = 0.78$). As in Study 1, we excluded children with hearing or visual impairment, neurological disorders, or other neuro-developmental disorders.

3.1.2. General Procedure

Children in the control group were recruited in public schools in Geneva and in the canton of Valais. Children in the caseload group were recruited by contacting SLTs. They were tested individually in a single session (of 30 to 45 min duration) either at school or in the speech-language therapy clinic. The following tasks were administered: (a) a teaching phase of DA, which included a repetition task of the non-words and a memory game; (b) an immediate post-test of DA; (c) Raven's matrices (Raven 1998); and (d) a delayed post-test of DA. For the same reasons as in study 1, there was no pre-test. Two master's students in speech-language therapy at the University of Geneva administered the different tasks.

Table 3. Participant information—Study 2.

Groups	Control Group (CBL-CTRL)	Caseload Group (CBL-CSL)
N	15	10
Age range (years; months)	4;11–6;10	4;7–8;4
Mean age SD (months)	6;2 5.6	6;6 14.4
Gender	8 G, 7 B	3 G, 7 B
Duration of exposure (months)	40.4 (10.54)	40.3 (19.44)
Raven–Mean standard scores (SD)	−0.79 (1.14)	−0.43 (1.16)

G = Girl; B = Boy.

3.1.3. Tasks

- Dynamic assessment

The assessment was based on the same procedure as in Study 1 with some modifications. The experimenter still presented a puppet to the child. The puppet introduced eight objects to the child (two real objects, pen and deer, and six novel objects) by naming them and describing their functions, as in this example: *"This is a /klipu/. Do you know what it is? Are there any on your planet too? [answer] I like /klipus/ because they allow us to see at night"*. Six non-words were associated with the six unfamiliar objects (see Appendix A). The non-words were /blodavo/; /zilobɛf/; /pitapu/; /gopim/; /klipu/ and /nefa/. Two versions of novel objects and non-word correspondences were created to control for name or object preferences. As in study 1, the non-words contained simple and common syllable structure (CV or CVC), early acquired consonants, and low neighborhood density. To increase the likelihood that the child would establish a phonological representation of the word, increased exposure to the target words was provided by including a repetition task and a memory game. Children had to repeat the object's name after the model on two separate occasions, thus, providing a repetition score. During the learning phase, children played a memory game in which they had to search for pairs of the target non-words (depicted on colored cards). Each time, the child turned over a card, s/he had to name the object, thus, ensuring multiple productions of the non-words. During testing phases, the puppet asked the child to name three objects that they wanted to play with, as well as their functions. In contrast to Study 1, children could choose three target words out of the six learned. We reasoned that children may have more success in word production if they chose the objects they liked, thus, reducing the chances of floor effects as observed in Study 1. Prompts were provided if the child could not produce the target word (see below). Apart from naming the target word, which yielded a phonological production score, we also tested recall of semantic features, which yielded a semantic production score. Finally, the puppet announced its departure and asked the child to help it put away the objects: each target object was named, and the child had to give the puppet the named object, thus providing a comprehension score. A delayed post-test took place after administration of the Raven's, that is, after a delay of about 10 minutes.

- Prompts and scoring

For the phonological production score, we used graduated prompting to help the child produce the target word: the first prompt was the initial sound, the second prompt was the initial syllable, and the third prompt was the first and second syllable in the case of a trisyllabic non-word. Each non-word was allocated 4 points: if the non-word was correctly produced, the score was 4. Depending on the type of error (distortions, additions, and deletions of phonemes or syllables, confusions, perseverations, or non-responses), the non-word could receive partial points: 1, 2, or 3 points. If the child needed prompting, the score was 2 or 3 depending on the number of phonemes or syllables produced. The maximum phonological production score was 12. For the semantic production score, we coded children on the recall of key semantic features (see Appendix A). If the child recalled semantic features correctly, s/he received 2 points. If s/he used a gesture or provided a semantic characteristic related to the function but not the exact function, s/he obtained 1 point. The maximum semantic production score was 6. For the comprehension score, if the child identified the correct target object, s/he obtained 1 point. If not, s/he obtained 0 points. The maximal score for comprehension was 6 points since all nonwords were included in the comprehension task. For the repetition score, each correct repetition was allocated 1 point, i.e., 2 points per non-word for a total of 12 points.

- Raven's matrices

This task was already presented in Study 1. The Mann–Whitney test indicated that there was no difference in non-verbal ability on the standard scores of the Raven's matrices between the two groups (U = 60.5, p = 0.44).

3.1.4. Data Analysis

Analyses were performed using R statistical software (R Core Team 2020). The independent variable was Group (CBL-CTRL or CBL-CSL), and the dependent variables were scores on nonword repetition, phonological production, semantic production, and comprehension on immediate and delayed post-tests. Due to the small number of participants in this study, a non-parametric (Mann–Whitney) test with Bonferonni correction for adjusting p values was employed to compare inter-group performance.

3.2. Results

Table 4 presents scores on all measures for the two groups and the results of the Mann–Whitney tests. For the repetition score, the Mann–Whitney test indicated a significant difference between groups: the CBL-CSL group performed less well than the CBL-CTRL group when they had to repeat the non-words. For dynamic scores, there were no significant differences across phonological production, semantic production, and comprehension, although, at a qualitative level, the results went in the right direction, at least for phonological and semantic subtests, with superior scores for the control versus caseload group. Marginal effects were found for the phonological score in the delayed post-test and for the semantic score in the immediate post-test.

Table 4. Means (SD) of all dynamic scores for the two groups—Study 2.

Groups		CBL-CTRL	CBL-CSL	Mann–Whitney Tests
N		15	10	
Repetition (/12)		10.60 (1.64)	7.00 (2.91)	U = 129.5, p = 0.002 *
Phonological production (/12)	**Immediate**	3.87 (3.68)	2.80 (2.66)	U = 85, p = 0.59
	Delayed	3.47 (3.36)	1.30 (1.49)	U = 107, p = 0.07 (*)
Semantic production (/6)	**Immediate**	4.73 (1.75)	3.40 (2.00)	U = 107.5, p = 0.07 (*)
	Delayed	4.53 (1.51)	3.60 (2.07)	U= 95, p = 0.27
Comprehension (/6)	**Immediate**	4.60 (0.99)	4.10 (1.37)	U = 88.5, p = 0.45
	Delayed	4.47 (0.99)	4.40 (1.26)	U= 77, p = 0.93

* CBL-CTRL = Consecutive Bilinguals Control; CBL-CSL = Consecutive Bilinguals Caseload; * indicate a significant difference; (*) indicate a marginal difference.

3.3. Discussion

This second study evaluated the validity of a dynamic word learning task with caseload and control bilingual children. We hypothesized that the control group would have better scores on DA than the caseload group. Our hypotheses were not supported: no significant differences were found between groups on all components of the DA. Hence, this task did not appear to strongly distinguish between control and caseload consecutive bilingual children. Nonetheless, the control group obtained better scores than the caseload group on all tasks, with two marginal effects for the production scores. As in Study 1, we observed a floor effect for phonological production: all children experienced difficulty recalling the name of a recently learned object. They experienced less difficulty recalling the semantic features of a novel object; however, this subtest did not significantly discriminate the two groups. Moreover, our findings on the repetition component of the task confirm the results of numerous studies that non-word repetition is a sensitive diagnostic measure of DLD (see Schwob et al. 2021 for a meta-analysis).

As for limitations, the small number of participants, particularly in the caseload group, may have reduced statistical power, thus, preventing significant differences to emerge. In addition, there were some methodological factors which may have compromised results. Although we had hypothesized that letting children have the choice of recalling three words in the production task may improve recall, it did not do so. Moreover, the learning procedure took place during a memory game, which may explain the floor effects in phonological production: since this game involves memorization abilities, it draws on cognitive resources that may have been directed at the memory task and not at the learning of new words. This learning method also does not allow for good control of the number of exposures to the non-words since this will vary between children depending upon their success in the game. Due to all of these limitations, we conducted a third study enlarging the number of participants and adjusting the procedure (e.g., controlling for the number of repetitions, etc.) to once again examine the efficacy of a word learning DA to differentiate TD and DLD children. In this study, we included both monolingual and bilingual children.

4. Study 3

The third study examined the validity of a modified version of our DA task with monolingual and bilingual TD and DLD children. The main question was: Does a dynamic word learning task distinguish a group of TD monolingual and bilingual children from a group of peers with DLD? Once again, we assumed that DLD children would have poorer results on DA than TD children.

4.1. Method

4.1.1. Participants

Seventy-seven children were originally tested. Because of missing data ($n = 13$, i.e., incomplete data and questionnaire information, recording problems) and unforeseen circumstances that did not permit us to recruit sufficient numbers of bilingual CSL children to form a fourth group ($n = 2$), the data of sixty-two children (28 girls, 34 boys) were used for further analyses (see Table 5 for a description). Twenty-three children were in the monolingual TD control group (ML-CTRL), 24 in the bilingual TD control group (BL-CTRL)—with 12 simultaneous (SBL-CTRL) and 12 consecutives (CBL-CTRL)—and 15 in the monolingual caseload group (ML-CSL). All children were between the ages of 4;5 and 9;3 years, with a mean age of 6;10 (SD = 17 month). The mean age did not differ across groups (H (2,62) = 0.65, $p = 0.72$). The languages spoken by the bilingual participants, in addition to French, were English ($n = 4$), Italian (*3*), Spanish (*3*), Turkish (*3*), Portuguese (*3*), Albanian (*1*), Tagalog (*1*), Hungarian (*1*), German (*1*), Malayalam (*1*), Tamazight (*1*), Dutch (*1*), Persian (*1*), and Arabic (*1*). As in previous studies, we excluded children with additional diagnoses.

Table 5. Participant information—Study 3.

Groups	Monolingual Control (ML-CTRL)	Bilingual Control (BL-CTRL)	Monolingual Caseload (ML-CSL)
N	23	24	15
Age range (years; months)	4;5–9;2	4;10–8;9	5;0–9;3
Mean age SD (months)	6;9 16.9	6;10 17.7	7;1 17.1
Gender	8 G, 15 B	14 G, 10 B	6 G, 9 B
Raven–Mean standard scores (SD)	−0.05 (0.69)	0.02 (0.69)	−0.85 (0.84)

G = Girl; B = Boy.

4.1.2. General Procedure

Children were recruited from schools in Geneva, and in neighboring France. The study was conducted between April 2020 and June 2021 during which there was COVID-19 lockdown in Geneva and France. Thus, some sessions ($n = 6$) also took place on video call with the Zoom software (Zoom n.d.). However, for the majority of sessions, children were tested individually in a single session (of 30 to 50 min duration) at their schools or in their homes. The following tasks were proposed in this order: (a) Raven's matrices (Raven 1998); (b) a teaching phase of DA of word learning; (c) an immediate post-test of DA; (d) static expressive vocabulary task using the standardized subtests of EVALO 2-6 (Coquet et al. 2009) or BILO 3C ("*Bilan Informatisé de Langage Oral pour le cycle 3 et le Collège*", Khomsi et al. 2007), two French language assessment batteries; and (e) a delayed post-test of DA. Three trained experimenters (two master's students and one doctoral student) administered the protocol.

4.1.3. Tasks

- Standardized vocabulary tasks

The vocabulary tasks included the expressive vocabulary subtests of the EVALO 2-6 (Coquet et al. 2009) and the BILO 3C (Khomsi et al. 2007). It was necessary to administer two different vocabulary tasks because of the wide age range of the children. For children under 6 years of age, the "*naming*" task (short version) of the EVALO 2-6 was used. For those over 6 years, the BILO 3C "*lexicon in production*" task was employed. For the EVALO 2-6 task, children had to name 40 pictures: four body parts, 28 objects, and eight actions. Two points were accorded if the children produced the word, resulting in a maximum score of 80. For the BILO 3C task, children had to name 39 pictures: 24 objects and 19 actions. One point was attributed to each target word correct, resulting in a maximum score of 39 points. Standard scores were calculated in order to compare the results between groups.

- Dynamic assessment

The assessment was based on the same procedure as in Studies 1 and 2 with some modifications: the number of presentations by the examiner and repetitions by the child were increased and better controlled. Six objects and associated functions were created to correspond to the six target non-words (see Appendix A). Twelve additional common objects were also used as distractor items: duck, pen, glasses, snail, tomato, fish, scissors, fork, butterfly, mushroom, car, balloon. Each object was represented by a card with its picture. For the game, two puppets and a letter box were also used. The non-words were /nefa/, /gopim/, /klipu/, /blodavo/, /mitapu/ and /zilobɛf/, the same words as in study 2, except /mitapu/ was employed instead of /pitapu/ due to the fact that /pitapu/ was often pronounced as /patapuf/ (a French word).

This task was based on the DAPPLE posting game (Hasson et al. 2013). The child had to help a stuffed animal send gifts to its friends on another planet. The first step was a familiarization phase in which three novel objects were named three times and repeated by the child one time. The semantic features of the objects were also described, as in this example: "*Here, it's a /gopim/. Can you repeat this? Yes, a /gopim/. The /gopim/ is a sweet fruit that grows on my planet*". During the teaching phase, a picture of each novel object was presented to the child along with a picture of two known objects. The stuffed animal asked the child to identify the target object (the novel object) that it wanted to send to its friends. During this phase, the name of the target object was provided about 12 times by the examiner and repeated three times by the child. This step was performed four times: once for the three novel objects and once for a known object. Two different orders of presentation for the non-words were proposed to control for primacy and recency effects during the test phase. During post-tests, the examiner presented the three pictures representing the novel objects that had been sent by the stuffed animal to its friends and asked the child to name these objects (=phonological production), and to describe their function (=semantic production). Following the phonological production post-test, for each incorrectly named object, the examiner gave the correct name and asked the child to point to the correct object (comprehension). The whole procedure was then repeated for the three other novel objects. Delayed post-tests were conducted after the vocabulary task: the sister of the puppet wanted to know which gifts had been sent. The six novel objects were presented, and the child had to name them and provide their function, yielding production scores for the delayed post-test. In the same way, the delayed comprehension score was obtained after the naming of the unnamed objects.

- Scoring

For the phonological scores, each correctly produced non-word scored 3 points; an incorrect production, in which one phoneme was omitted, substituted, or added, scored 2 points; an incorrect production in which two phonemes or a syllable were changed or omitted, scored 1 point. The total phonological production score was 18. For the semantic scores, key features were associated with each function (see Appendix A). If the key feature (or an acceptable synonym) was included in the child's description, one point was awarded. Three non-words were associated with three key features and three were associated with two key features yielding a total semantic score of 15. For the comprehension score, two points were awarded when the non-word was correctly produced (since we assumed that if the child could produce the word, s/he could also identify it), or when it was correctly identified after naming. These scores were calculated for immediate and delayed post-tests.

- Raven's matrices

Even if this task was proposed to control for the non-verbal abilities of the children and to verify that there were no differences between groups, the Kruskal–Wallis test indicated significant difference between the three groups (H (2,62) = 10.87, p = 0.004). Differences were found between ML-CSL and ML-CTRL (U = 270, p = 0.01) and between ML-CSL and BL-CTRL (U = 284, p = 0.009). The caseload group had lower standard scores than the two control groups.

4.1.4. Data Analysis

Analyses were performed using R statistical software (R Core Team 2020). The independent variable was Group (ML-CTRL, BL-CTRL, and ML-CSL) and the dependent variables were static and DA scores (phonological production, semantic production, and comprehension scores on immediate and delayed post-tests). Because of the lack of a fourth group of bilingual caseload children, we compared the groups two by two. To do this, we performed non-parametric Mann–Whitney tests with Bonferonni correction for adjusting p values to compare the results of the monolingual groups (CTRL vs CSL) and then the control groups (ML and BL).

4.2. Results

Table 6 displays results on the static and dynamic tasks for the three groups.

Table 6. Means (SD) of static and dynamic scores for the three groups—Study 3.

Groups			ML-CTRL	BL-CTRL	ML-CSL
N			23	24	15
Static expressive vocabulary tasks (standard scores)			0.14 (0.74)	−1.07 (1.43)	−0.34 (1.12)
DA	Phonological production (/18)	Immediate	5.78 (2.83)	5.29 (3.17)	3.93 (2.31)
		Delayed	1.78 (2.41)	1.17 (1.90)	0.07 (0.26)
	Semantic production (/15)	Immediate	7.57 (2.21)	6.38 (2.52)	4.47 (2.85)
		Delayed	7.65 (2.25)	6.25 (2.54)	4.67 (3.11)
	Comprehension (/12)	Immediate	9.30 (2.53)	8.92 (2.57)	8.13 (3.34)
		Delayed	6.87 (3.76)	5.42 (3.56)	4.53 (2.33)

ML-CTRL = Monolingual Control; BL-CTRL = Bilingual Control; ML-CSL = Monolingual Caseload.

4.2.1. Monolingual Control and Caseload Groups

For the standard scores on the expressive vocabulary tasks (EVALO 2-6 and BILO 3C subtests), the Wilcox–Mann Whitney test indicated no significant difference between monolingual CTRL and CSL groups (U = 124.5, $p = 0.156$).

For the phonological production subtest of the DA, there was a significant difference between groups, in the immediate (U = 244, $p = 0.03$) and delayed post-tests (U = 241, $p = 0.01$). The same results were obtained for the semantic production subtest, with a significant difference between ML-CTRL and ML-CSL in the immediate (U = 278, $p = 0.002$) and the delayed post-tests (U = 274.5, $p = 0.002$). In both cases, caseload monolingual performed less well than control monolingual children. Nevertheless, in the comprehension subtest, there was no difference between the two groups (U = 207.5, $p = 0.29$ in immediate and U = 229, $p = 0.09$ in delayed post-tests).

4.2.2. Monolingual and Bilingual Control Groups

For the standard scores on all the expressive vocabulary tasks (EVALO 2-6 and BILO 3C subtests), the Wilcox–Mann Whitney test indicated a significant difference between monolingual and bilingual CTRL groups (U = 123.5, $p = 0.001$), with bilinguals performing less well than monolinguals.

For the phonological production subtests, there was no significant difference between groups, in the immediate (U = 296, $p = 0.67$) and in the delayed post-tests (U = 307, $p = 0.46$). For the semantic production subtests, however, a significant difference was found between ML-CTRL and BL-CTRL in the delayed (U = 373, $p = 0.038$) but not in the immediate post-test (U = 355.5, $p = 0.09$). For the comprehension subtests, there were no differences between the two control groups (in the immediate, U = 292, $p = 0.72$, and in the delayed post-tests, U = 336.5, $p = 0.19$).

Since our analyses indicated some differences between the monolingual and bilingual controls, we investigated whether type of bilingualism (i.e., simultaneous or consecutive) played a role. We, therefore, performed multiple Wilcoxon comparisons between these three groups. For the static tasks, consecutive bilinguals performed less well than the two other groups (U = 12.5, $p < 0.001$ with ML-CTRL; U = 3.5, $p < 0.001$ with SBL-CTRL). For semantic production in the delayed post-test, Figure 1 compares the results of the three control groups alongside the caseload group. Wilcoxon pairwise comparison indicated a significant difference between consecutive bilingual children and the two other control groups (U = 43.5, $p = 0.003$ with ML-CTRL; U = 27, $p = 0.03$ with SBL-CTRL), with consecutive bilinguals performing less well.

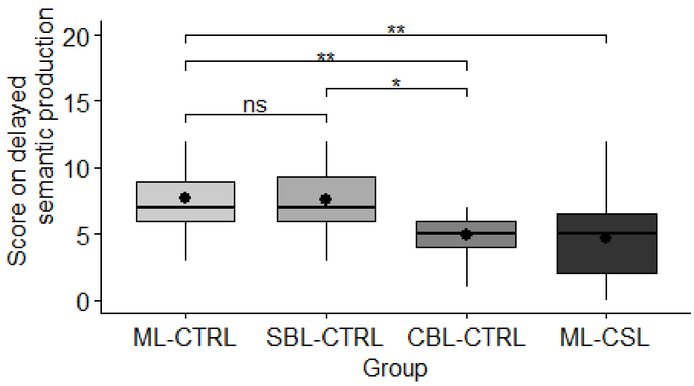


Figure 1. Study 3: Boxplot of scores on the delayed semantic production test for the four groups. ML-CTRL = Monolingual Control; SBL-CTRL = Simultaneous Bilingual Control; CBL-CTRL = Consecutive Bilingual Control; ML-CSL = Monolingual Caseload; Points represent the mean of each group; *: $p < 0.05$; **: $p < 0.01$; ns = non-significant.

4.3. Discussion

The goal of the third study was to make additional modifications to our dynamic word learning task and to test its validity with monolingual and bilingual control and caseload children. We assumed that the caseload group would have poorer results on DA than the control group. Unfortunately, due to unforeseen circumstances we were not able to recruit a bilingual caseload group. We found that monolingual caseload children performed less well than monolingual controls both for phonological and semantic production, which is consistent with our hypothesis. Moreover, we did not find any differences between monolingual and bilingual control children in phonological production. However, we did find differences in delayed semantic production. As indicated in Figure 1, this difference resulted from the poorer performance by the consecutive bilingual children, while the simultaneous bilingual control children patterned similarly to the monolingual controls. As is the case with a standardized vocabulary task, the recall of semantic features in the DA appeared to penalize at least some of the bilingual children. We hypothesize that for consecutive bilinguals, it is more difficult to understand, store, and retrieve the definition of the word because they have had less exposure to the language. Indeed, we found that delayed semantic production was correlated with duration of exposure to French[2]. Finally, we did not find any significant differences between groups for comprehension.

Although the phonological production task served its purposes in that it distinguished between monolingual control and caseload children and it did not distinguish between monolingual and bilingual control groups, it was nevertheless subject to floor effects, particularly in the delayed post-test. Even in the immediate post-test, children were able to produce only 0 to 4 words, with a mean of less than 2 words. In this study, we proposed two different orders of presentation of non-words in the learning phases. We looked at whether this influenced the immediate phonological production of words and we found that when the word appeared last, it was better recalled than when it appeared first or second[3]. Phonological production is, therefore, difficult and shows a recency effect, i.e., the last learned element is better recalled. This effect was not found in delayed post-tests whereby the mean number of words recalled for each group decreased sharply showing a significant "forgetting" effect. These effects did not appear for semantic production: key features were not better recalled for the last learned words, and the mean score was maintained between the immediate and delayed post-test for each group.

Regarding the difference found on the standard scores of the Raven's between the caseload and the two other control groups, we found that three caseload children and one monolingual control child had a standard score below −1.6. We conducted an analysis without these children and still found a significant difference between the ML-CSL and the ML-CTRL groups[4]. Some studies have shown that children with DLD have poorer performance on nonverbal tasks than age-matched TD peers (Leonard 2014) and, in the actual definition of DLD (Bishop et al. 2017), nonverbal delays are no longer an exclusionary criterion for a DLD diagnosis.

In sum, we found that, for monolingual children, phonological and semantic production of novel words discriminate control and caseload groups, which were encouraging results. The main limitation of this third study was the lack of a bilingual caseload group, which could be compared to the bilingual control children. Nevertheless, we found no differences between monolingual and bilingual control groups on phonological production. We did, however, find differences between consecutive bilingual and other control children on delayed semantic production, indicating that some features of our semantic task penalized at least some of the bilingual children. Other authors have reported word learning differences between bilinguals and monolinguals (Alt et al. 2013, 2019).

5. General Discussion

We designed this series of studies with the aim of creating a dynamic word learning task that could distinguish TD and DLD—monolingual and bilingual—children. The ultimate goal was to develop a diagnostic tool that SLTs could use with bilingual children in a single session. We created three dynamic tasks in which children had to learn the phonological form of non-words and their semantic features and to recall them on immediate and delayed post-tests. These studies investigated whether phonological production, semantic production, and comprehension could accurately distinguish caseload children, i.e., children who were receiving speech-language therapy for DLD, and their control peers. The DA tasks employed in the three studies were based on methodologies adopted by Hasson et al. (2013) and Kapantzoglou et al. (2012) and included graduated prompting and test-teach-retest formats.

The first study compared TD monolingual and bilingual children on both static and dynamic measures. We found significant differences between monolinguals and bilinguals on the static measures but no differences between the two groups on the dynamic measures, consistent with previous research that indicates that standardized vocabulary tasks penalize bilingual children, and that DA is a promising way to assess bilingual children (Camilleri and Law 2007; Peña et al. 2001). The second study compared the DA results of bilingual children who were receiving speech-language therapy for DLD (caseload group) with those who were not (control group). We observed some trends in the data that suggested that the caseload children performed less well than the control children; nevertheless, the results did not reach significance. The third study combined features of the first two by comparing the performance of monolingual and bilingual—control and caseload—children on both static and dynamic measures. The static measures did not distinguish between monolingual caseload and control groups and, furthermore, penalized the bilingual children. In contrast, the dynamic measure was effective in distinguishing control and caseload monolingual groups (we did not test bilingual caseload children) and did not discriminate between monolingual and bilingual control groups with the exception of the semantic production task, a point which we will discuss below.

If we consider the component of word learning requiring the child to produce the phonological form of the word, we found that monolingual children in the control group performed better than those in the caseload group. This result agrees with findings that DLD children require more exposures to each word to reach the same learning criterion level as their peers (Alt and Plante 2006; Gray 2004; Gray et al. 2012). In our studies, all children were exposed to similar degrees of exposure to each word, which explains why caseload children had lower scores than control children. Nevertheless, even our control children obtained poor overall results on production scores. Learning three to six new words in a single session is, thus, a difficult task even for TD children. Our task results indeed indicate floor effects in the phonological production subtest and ceiling effects in the comprehension subtest. Such findings are well known and other researchers have reported that production is more challenging than comprehension (Gray 2004). Gray et al.

(2020) suggest that the differences between production and comprehension are due to task demands related to the robustness of underlying representations. For comprehension, the phonological representation of the word is provided, and the child only has to activate the link with the semantic representation. However, for production, the child has to recall the phonological representation to produce the form of the words. Representations and links must, therefore, be stronger to succeed in the production task.

If we focus on the component of word learning involving recall of semantic features, we found that monolingual control children also performed better than caseload children. Other researchers have reported that DLD children struggle not only with the learning of the phonological form but also with learning of semantic features (Alt and Plante 2006). However, our study showed that consecutive bilingual control children also struggle with the learning of semantic features. In Study 2, no difference was found between control and caseload consecutive bilingual children, and in Study 3, consecutive bilingual children had lower scores than monolingual and simultaneous bilingual children on the delayed semantic production post-test. We interpret the low scores on semantic feature recall by the CBL children to reflect the same sort of difficulty they experience on standardized vocabulary tasks. Due to reduced exposure to the L2, they have less robust semantic representations than their monolingual and SBL control peers. Their less developed semantic knowledge penalizes them on standardized vocabulary tasks but also on short duration dynamic measures, such as the ones developed in our studies. We still consider recall of semantic features to be an important component of word learning; however, future revisions to our DA will need to carefully select semantic features so as not to disadvantage the CBL children. We acknowledge that our findings to date have been relatively modest concerning the ability of our word learning tasks to discriminate between TD and DLD children, a finding that has been noted in other studies (Gray et al. 2012; Kapantzoglou et al. 2012). The main explanation for the non-significant difference between caseload and control groups on word learning tasks is the large overlap in scores: some caseload children perform well in word learning and some control children do not. Moreover, the cognitive demands of the word learning task are highly sensitive to methodological variations. Indeed, many factors influence word learning and it is important to control all of these factors. Word characteristics (i.e. phonotactic probability, neighborhood density, etc.), referent characteristics (i.e. concrete or not, familiar or not, etc.), learner characteristics (age, SES, bilingualism), the learning situation (implicit or explicit, etc.), the number of words to be learned, the number of distractors present during learning trials, and the balance between teaching and testing all impact on word learning (Gray et al. 2012). This study took into consideration many of these factors, but not all of them. Furthermore, some of the factors were different between our three studies (e.g., the nature of the referent, the learning situation, the number of words), which may have led to different results. Modifications to the task and a larger sample size are, thus, needed to examine more accurately group differences between TD and DLD children. Finally, another limitation was the wide age ranges of children across studies. Study 1 was characterized by children aged 4 to 6 years, Study 2 by children aged 4 to 8 years, and Study 3 by children aged 4 to 9 years. So, age range increased across studies. We ensured that the mean age of groups was comparable within a study; however, we did not consider the age range differences across studies: indeed, word learning at age 4 may involve different processes than word learning at age 8–9. In future studies, we will focus our efforts on a narrower age range of children so as to delineate the word learning processes specific to that group of children. Moreover, within studies, although the mean age did not differ between groups, the age range was sometimes quite different, and these subtle differences may have influenced the results, particularly in Studies 1 and 2. For example, in Study 2, there were older children in the caseload than in the control group. So, it is possible that caseload children had higher scores, which could have contributed to the lack of difference in DA scores between the two groups.

6. Conclusions

Even if our results to date are relatively modest concerning the ability of our word learning tasks to differentiate TD and DLD bilingual children, we are convinced of the importance of developing alternative approaches to assess the lexical skills of bilingual children. Indeed, the drawback of standardized vocabulary tasks for use with bilingual children were clearly evident in the current study. Presently, we are working on developing tasks with other word learning methodologies such as shared storybook reading (Burton and Watkins 2007; Flack et al. 2018) and word retrieval during the learning process (Leonard et al. 2021) to increase the number of words learned and recalled by children. Our ultimate aim is to create a dynamic word learning task that can be administered in a single session, and that can aid in the diagnosis of children with DLD. This type of task could then be employed by French-speaking SLTs who still lack effective tools to diagnosis DLD in bilingual children (Delage et al. 2021; Stanford et al. Forthcoming).

Author Contributions: Conceptualization, M.M., M.K. and H.D.; methodology, M.K. and H.D.; validation, M.M., M.K. and H.D.; formal analysis, M.M.; investigation, M.K.; resources, M.M. and M.K.; data curation, M.M. and M.K.; writing—original draft preparation, M.M.; writing—review and editing, M.M., M.K. and H.D.; visualization, M.M.; supervision, M.K.; project administration, H.D.; funding acquisition, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Swiss National Science Foundation Grant grant number 1001F_197983 (awarded to Hélène Delage).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the University Commission for Ethical Research in Geneva (CUREG) (FNS 1001F_197983, 23 September 2020).

Informed Consent Statement: Written informed consent was obtained from all the participants' parents for their children's involvement in the study and for any scientific publications that may ensue from this research.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical reasons.

Acknowledgments: We wish to thank O. Gubser (study 1), L. Navarra and V. Maselli (study 2) and S. Di Sciullo (study 3) for their contributions to the creation of the dynamic tasks and the testing of children.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Characteristics of the Non-Words and Novel Objects for the Three Studies

Table A1. Characteristics of the Non-Words (Phonetic Form and Syllabic Structure) and Novel Objects (Description and Semantic Features) for the Three Studies.

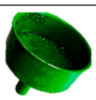
Novel Object	Non-Word	Syllabic Structure	Object Description	Semantic Features
Study 1				
	/nefa/	CV CV	Ball of fabric in the shape of a stone.	To wash.
	/tibak/	CV CVC	Small green varnished funnel.	To communicate/talk.
	/gopi/	CV CV	Toe spreader with wool rod.	To style your hair.

Table A1. *Cont.*

Novel Object	Non-Word	Syllabic Structure	Object Description	Semantic Features
Study 2				
	A: /blodavo/ B: */nefa/*	CCV CV CV *CV CV*	Two taped toothbrushes, with eyeglass stems and a flashing red fountain pen.	Flying transport.
	A: /zilobɛf/ B: */gopim/*	CV CV CVC *CV CVC*	Customized massager with silver stars and a black ribbon.	Teleporter to move to the desired location very quickly.
	A: /pitapu/ B: */klipu/*	CV CV CV *CCV CV*	Modified stuffed rabbit with sewn-on ears and a cotton wing added to the back.	Pet made of clouds.
	A: /gopim/ B: */pitapu/*	CV CVC *CV CV CV*	Plastic box decorated with flames and colored pendants.	Heats food in a second.
	A: /klipu/ B: */zilobɛf/*	CCV CV *CV CV CVC*	Binoculars with yellow glasses and earplug antennas.	Allows someone to see at night.
	A: /nefa/ B: */blodavo/*	CV CV *CCV CV CV*	Hairbrush without the spikes.	To style your hair.
Study 3				
	/blodavo/	CCV CV CV	Black tray containing a blue ball topped by a cylinder, a blue stone, and small balls of modeling clay.	Machine that turns pebbles into candy.
	/zilobɛf/	CV CV CVC	Soap pusher, with dough balls and a fringed key ring.	Tickling machine.
	/mitapu/	CV CV CV	Key chain with rings and dough.	Musical instrument to play at parties.
	/gopim/	CV CVC	Yellow playdough with clay balls and pink modeling clay.	Sweet fruit that grows on the stuffed toy's planet.
	/klipu/	CCV CV	"Alien" on a stick with modeling clay assembly.	Small animal that lives on the stuffed toy's planet.
	/nefa/	CV CV	Two wooden logs assembled by a ring of modeling clay, overhung by a lid.	To take photos at night.

The underlined words are the target words expected on the semantic production post-tests.

Notes

1. "Head Start is a federally-funded program in the United States that provides preschool services to children from families living in poverty." (Hammer et al. 2008).
2. Spearman's rank correlation between delayed semantic production and duration of exposition: r (60) = 0.43, $p < 0.001$.

[3] Results of Mann–Whitney between versions of order presentation for /*mitapu*/: U = 285.5, p = 0.002284; /*blodavo*/: U = 815, $p <$ 0.001; /*gopim*/: U = 182.5, $p < 0.001$; and /*klipu*/: U = 121 840, $p < 0.001$.

[4] Results of the Mann–Whitney test for RAVEN' standard scores between ML-CSL and ML-CTRL groups with the exclusion of children with standard scores below 1.6: U = 202, p = 0.039.

References

Alt, Mary, and Elena Plante. 2006. Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 49: 941–54. [CrossRef]

Alt, Mary, Christina Meyers, and Cecilia Figueroa. 2013. Factors that influence fast mapping in children exposed to Spanish and English. *Journal of Speech, Language, and Hearing Research* 56: 1237–48. [CrossRef]

Alt, Mary, Genesis D. Arizmendi, Shelley Gray, Tiffany P. Hogan, Samuel Green, and Nelson Cowan. 2019. Novel word learning in children who are bilingual: Comparison to monolingual peers. *Journal of Speech, Language, and Hearing Research* 62: 2332–60. [CrossRef] [PubMed]

Arias, Graciela, and Jennifer Friberg. 2017. Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools* 48: 1–15. [CrossRef] [PubMed]

Armon-Lotem, Sharon, Jan de Jong, and Natalia Meir. 2015. *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol and Blue Ridge Summit: Multilingual Matters. [CrossRef]

Bialystok, Ellen, Gigi Luk, Kathleen F. Peets, and Sujin Yang. 2010. Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition* 13: 525–31. [CrossRef]

Bishop, Dorothy V. M., Margaret J. Snowling, Paul A. Thompson, Trisha Greenhalgh, and the CATALISE-2 Consortium. 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: 1068–80. [CrossRef]

Burton, Vera Joanna, and Ruth V. Watkins. 2007. Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders* 40: 335–56. [CrossRef]

Caesar, Lena G., and Paula D. Kohler. 2007. The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools* 38: 190–200. [CrossRef]

Camilleri, Bernard, and James Law. 2007. Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology* 9: 312–22. [CrossRef]

Campione, Joseph C., and Ann L. Brown. 1987. Linking dynamic assessment with school achievement. In *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential*. Edited by Carol S. Lidz. New York: Guilford Press, pp. 82–115.

Carey, Susan, and Elsa J. Bartlett. 1978. Acquiring a single new word. *Proceedings of the Stanford Child Language Conference* 15: 17–29.

Chevrier-Muller, Claude, and Monique Plaza. 2001. *Nouvelles Epreuves pour L'examen du Langage (N-EEL)*. Paris: ECPA, Les Editions du Centre de psychologie appliquée.

Coquet, Françoise, Pierre Ferrand, and Jacques Roustit. 2009. *EVALO 2-6*. Isbergues: Ortho édition.

De Lamo White, Caroline, and Lixian Jin. 2011. Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language and Communication Disorders* 46: 613–27. [CrossRef] [PubMed]

Delage, Hélène, Pauline Prat, and Margaret Kehoe. 2021. Évaluation dynamique en orthophonie/logopédie. *Glossa* 131: 121–42.

Dunn, Lloyd M., and Leota M. Dunn. 1997. *Peabody Picture Vocabulary Test*, 3rd ed. Bloomington: Pearson Assessments.

Flack, Zoe M., Andy P. Field, and Jessica S. Horst. 2018. The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology* 54: 1334–46. [CrossRef] [PubMed]

Gathercole, Virginia C. Mueller, Ivan Kennedy, and Enlli Môn Thomas. 2016. Socioeconomic level and bilinguals' performance on language and cognitive measures. *Bilingualism: Language and Cognition* 19: 1057–78. [CrossRef]

Gray, Shelley. 2004. Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research* 47: 1117–32. [CrossRef]

Gray, Shelley, Shara Brinkley, and Dubravka Svetina. 2012. Word learning by preschoolers with SLI: Effect of phonotactic probability and object familiarity. *Journal of Speech, Language, and Hearing Research* 55: 1289–300. [CrossRef]

Gray, Shelley, Hope Lancaster, Mary Alt, Tiffany P. Hogan, Samuel Green, Roy Levy, and Nelson Cowan. 2020. The structure of word learning in young school-age children. *Journal of Speech, Language, and Hearing Research* 63: 1446–66. [CrossRef]

Grimm, Angela, and Petra Schulz. 2014. Specific language impairment and early second language acquisition: The risk of over- and underdiagnosis. *Child Indicators Research* 7: 821–41. [CrossRef]

Gutiérrez-Clellen, Vera F., and Elizabeth Peña. 2001. Dynamic assessment of diverse children. *Language, Speech, and Hearing Services in Schools* 32: 212–24. [CrossRef]

Hammer, Carol Scheffner, Frank R. Lawrence, and Adele W. Miccio. 2008. Exposure to English before and after entry into head start: Bilingual children's receptive language growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism* 11: 30–56. [CrossRef] [PubMed]

Hasson, Natalie, and Victoria Joffe. 2007. The case for Dynamic Assessment in speech and language therapy. *Child Language Teaching and Therapy* 23: 9–25. [CrossRef]

Hasson, Natalie, Bernard Camilleri, Caroline Jones, Jodie Smith, and Barbara Dodd. 2013. Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy* 29: 57–75. [CrossRef]

He, Angela Xiaoxue, and Sudha Arunachalam. 2017. Word learning mechanisms. *Wiley Interdisciplinary Reviews: Cognitive Science* 8: e1435. [CrossRef]

Hoff, Erika, Rosario Rumiche, Andrea Burridge, Krystal M. Ribot, and Stephanie N. Welsh. 2014. Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly* 29: 433–44. [CrossRef]

Kan, Pui Fong, and Jennifer Windsor. 2010. Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research* 53: 739–56. [CrossRef]

Kapantzoglou, Maria, Adelaida M. Restrepo, and Marilyn S. Thompson. 2012. Dynamic assessment of word learning skills identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools* 43: 81–96. [CrossRef]

Khomsi, Abdelhamid, Jabril Khomsi, Frédéric Pasquet, and Aude Parbeau-Gueno. 2007. *Bilan Informatisé de Langage Oral pour le cycle 3 et le Collège (BILO 3C)*. Paris: ECPA, Les Editions du Centre de psychologie appliquée.

Kohnert, Kathryn. 2010. Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders* 43: 456–73. [CrossRef]

Leonard, Laurence B. 2014. *Children with Specific Language Impairment*. Cambridge: MIT Press.

Leonard, Laurence B., Sharon L. Christ, Patricia Deevy, Jeffrey D. Karpicke, Christine Weber, Eileen Haebig, Justin B. Kueser, Sofía Souto, and Windi Krok. 2021. A multi-study examination of the role of repeated spaced retrieval in the word learning of children with developmental language disorder. *Journal of Neurodevelopmental Disorders* 13: 20. [CrossRef]

Lidz, Carol S. 1987. *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential*. New York: Guilford Press.

Lidz, Carol S. 1991. *Practitioner's Guide to Dynamic Assessment*. New York: Guilford Press.

MacLeod, Andrea A. N., Ann Sutton, Natacha Trudeau, and Elin Thordardottir. 2011. The acquisition of consonants in Québécois French: A cross-sectional study of pre-school aged children. *International Journal of Speech-Language Pathology* 13: 93–109. [CrossRef]

Maddieson, Ian, Sébastien Flavier, Egidio Marsico, and François Pellegrino. 2014. LAPSyD: Lyon-Albuquerque Phonological Systems Databases. Version 1.0. Available online: http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd/ (accessed on 5 October 2015).

Marini, Andrea, Paola Sperindè, Isabella Ruta, Christian Savegnago, and Francesco Avanzini. 2019. Linguistic skills in bilingual children with developmental language disorders: A pilot study. *Frontiers in Psychology* 10: 493. [CrossRef] [PubMed]

Morgan, Paul L., George Farkas, Marianne M. Hillemeier, Carol Scheffner Hammer, and Steve Maczuga. 2015. 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development* 86: 1351–70. [CrossRef] [PubMed]

Nayeb, Laleh, Dagmar Lagerberg, Anna Sarkadi, Eva-Kristina Salameh, and Mårten Eriksson. 2021. Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta Paediatrica* 110: 265–72. [CrossRef] [PubMed]

New, Boris, Christophe Pallier, Marc Brysbaert, and Ludovic Ferrand. 2004. Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, and Computers* 36: 516–24. [CrossRef] [PubMed]

Office fédéral de la statistique. 2021. *Pratiques Linguistiques en Suisse—Premiers Résultats de l'Enquête sur la Langue, la Religion et la Culture 2019*. Neuchâtel: Office fédéral de la statistique. Available online: https://www.bfs.admin.ch/bfs/fr/home/statistiques/catalogues-banques-donnees/publications.assetdetail.15324910.html (accessed on 6 December 2021).

Oller, D. Kimbrough, Barbara Z. Pearson, and Alan B. Cobo-Lewis. 2007. Profile effects in early bilingual language and literacy. *Applied Psycholinguistics* 28: 191–230. [CrossRef]

Orellana, Carla I., Rebekah Wada, and Ronald B. Gillam. 2019. The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology* 28: 1298–317. [CrossRef]

Ouellette, Gene P. 2006. What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology* 98: 554–66. [CrossRef]

Paradis, Johanne, Fred Genesee, and Martha B. Crago. 2021. *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*, 3rd ed. Baltimore: Paul H. Brookes Publishing Co.

Peña, Elizabeth D. 1993. Dynamic Assessment: A Non-Biased Approach for Assessing the Language of Young Children. Ph.D. dissertation, Temple University, Philadelphia, PA, USA. unpublished.

Peña, Elizabeth D., Aquiles Iglesias, and Carol S. Lidz. 2001. Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology* 10: 138–54. [CrossRef]

Petersen, Douglas B., Helen Chanthongthip, Teresa A. Ukrainetz, Trina D. Spencer, and Roger W. Steeve. 2017. Dynamic assessment of narratives: Efficient, accurate identification of language impairment in bilingual students. *Journal of Speech, Language, and Hearing Research* 60: 983–98. [CrossRef]

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: https://www.R-project.org/ (accessed on 6 December 2021).

Raven, John J. 1998. *Raven Progressive Matrices*. Oxford: Oxford Pyschologists Press.

Schwob, Salomé, Laurane Eddé, Laure Jacquin, Mégane Leboulanger, Margot Picard, Patricia Ramos Oliveira, and Katrin Skoruppa. 2021. Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research* 64: 3578–93. [CrossRef]

Stanford, Emily, Maren Eikerling, Olivia Hadjadj, and Hélène Delage. Forthcoming. Supporting multilingual children with language impairment in a multilingual environment: Experience and perspectives from practitioners in Switzerland.

Thordardottir, Ellin. 2011. The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism* 15: 426–45. [CrossRef]

Tuller, Laurice, Layal Abboud, Sandrine Ferré, Alice Fleckstein, Philippe Prévost, Christophe dos Santos, Maureen Scheidnes, and Racha Zebib. 2013. Specific language impairment and bilingualism: Assembling the pieces. In *Language Acquisition and Development: Proceedings of GALA*. Cambridge: Cambridge Scholars Publishing, pp. 533–67.

Vygotsky, Lev. 1978. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Zoom. n.d. Zoom Video Communications, Inc. 2020–2021. Available online: https://zoom.us/ (accessed on 1 April 2020).