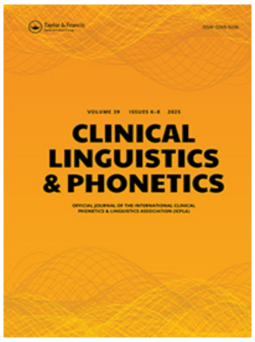

Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: www.tandfonline.com/journals/iclp20

A comparison of two dynamic assessment situations for detecting development language disorder in monolingual and bilingual children

Salomé Schwob, Yves Tillé & Katrin Skoruppa

To cite this article: Salomé Schwob, Yves Tillé & Katrin Skoruppa (2025) A comparison of two dynamic assessment situations for detecting development language disorder in monolingual and bilingual children, Clinical Linguistics & Phonetics, 39:6-8, 540-558, DOI: 10.1080/02699206.2024.2435010

To link to this article: https://doi.org/10.1080/02699206.2024.2435010

CLINICAL LINGUISTICS & PHONETICS
2025, VOL. 39, NOS. 6–8, 540–558
https://doi.org/10.1080/02699206.2024.2435010

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

A comparison of two dynamic assessment situations for detecting development language disorder in monolingual and bilingual children

Salomé Schwob[a], Yves Tillé[b], and Katrin Skoruppa[a]

[a]Institut des sciences logopédiques, Université de Neuchâtel, Neuchâtel, Suisse; [b]Institut de statistique, Université de Neuchâtel, Neuchâtel, Suisse

ABSTRACT
Bilingual children's language skills are strongly influenced by exposure to each of their languages, among other linguistic, environmental, and cognitive factors. In the speech and language therapy clinic, it is difficult to disentangle developmental language disorders from insufficient exposure. Dynamic assessment, which directly tests the learning potential of children, offers a promising solution for this dilemma. This study compares the clinical potential of two dynamic assessment situations, varying amount of adult mediation (autonomous computer game vs. interactive story reading with graduated cues), as well as item types (nouns, verbs, and inflections in sentences) and linguistic modalities and tasks (comprehension – word picture matching and acceptability judgement, production – free recall and picture naming). Forty-nine French monolingual and French-Portuguese bilingual children with and without developmental language disorder, aged 5;0 to 7;11 years, were include in the final analyses. Using Lasso regressions, we were able to determine which variables best explain the presence of disorder. A combination of all item types and predominantly receptive tasks, mostly from the interactive situation, was retained for very high classification accuracy (up to 100% sensitivity and 96% specificity). Language status showed no influence, which encourages the use of dynamic assessment in the context of speech and language assessment with bilingual children. This study adds to evidence that dynamic assessment is a promising task for identifying bilingual and monolingual children with developmental language disorder, particularly when the situation involves interaction with graduated cues.

ARTICLE HISTORY
Received 20 September 2023
Revised 18 October 2024
Accepted 20 November 2024

KEYWORDS
Developmental language disorder; bilingualism; dynamic assessment

Introduction

Language development and assessment in bilingual children

At present, the majority of children in the world grow up with more than one language. In general, these bilingual children reach their first developmental milestones (first words and word combinations) at the same age as monolingual children (Nicoladis & Genesee, 1997). However, their further language development is even more heterogeneous than that of their monolingual peers, due to a variety of additional influences: First, the quantity and quality

CONTACT Salomé Schwob salome.schwob@unine.ch Institut des Sciences logopédiques, Université de Neuchâtel, Pierre-à-Mazel 7, Neuchâtel 2000, Suisse
Supplemental data for this article can be accessed online at https://doi.org/10.1080/02699206.2024.2435010

of exposure in each language clearly impacts bilingual development (Paradis, 2023), although there are variations by language domain (e.g. lexicon vs. grammar, Thordardottir, 2011, 2015) and language modality (receptive vs. expressive, e.g. Keller et al., 2015). Furthermore, other factors such as the typology, proximity, and the prestige of the child's languages, migration background, learning style, and socioemotional wellbeing (Paradis, 2019, 2023), may also affect bilingual language development.

In principle, bilingual children are not more likely than their monolingual peers to be affected by developmental language disorder (DLD), a persistent oral language impairment with functional impacts (see full definition in Bishop et al., 2017), thus yielding an expected prevalence of ~ 8% in preschool children (Norbury et al., 2016). However, in the case of bilingualism, it is difficult for speech and language therapists (SLTs) to differentiate between a true DLD and temporary language difficulties due to insufficient exposure (Camilleri & Law, 2007). Qualitatively, the nature of the language difficulties and errors is also often similar during the initial stages of learning a new language and in pathological contexts (Armon-Lotem, 2012). Bilingual children are therefore sometimes over-identified – when any failure on assessment tasks is interpreted systematically as a DLD – or under-identified – when all difficulties are considered to be the inevitable result of lack of language exposure (Camilleri & Law, 2007). It is often only the persistence of the difficulties over time that indicates the presence of a DLD and demonstrates clear problems with language learning (Gray, 2003).

In order to reduce these identification errors, several assessment paths have been proposed in the scientific literature. Dual language assessment is advocated by professional associations (e.g. ASHA, 2024; IALP, 2020) but it is not always feasible due to multiples reasons: lacking assessment tools in some languages; SLT do not necessarily speak the children's first languages (Volpin et al., 2020); results of tests could be biased by exposure to the language of the tests (for a synthesis see McLeod et al., 2017) and, as such, a combination of monolingual tests in both languages often remains inconclusive (e.g. Schwob & Skoruppa, 2021). Alternatively, and because it appears to be least biased by language exposure, non-word repetition (NWR) has proven to be of discriminative value in both monolingual (Estes et al., 2007) and bilingual contexts (Schwob et al., 2021). However, such a single task is not sufficient to obtain a full profile of the child's difficulties (Peña et al., 2016) and children with DLD with little or no phonological processing difficulties may show good performance.

Dynamic assessment (DA), which directly probes the learning potential of children (Camilleri & Law, 2007) rather than knowledge of one particular language per se, has also been proposed to limit biases and floor effect when evaluating children speaking different languages and coming from different cultures (Budoff, 1987; Peña et al., 2014). For a variety of reasons (e.g. lack of time, lack of protocol, lack of tools and literature currently emerging, see Hasson & Joffe, 2007, and the meta-analysis of Hunt et al., 2022), clinicians still make little use of DA. Moreover, a recent study by Scharff Rethfeldt et al. (2024) shows that bilingual clinicians are currently more inclined to use it informally. More studies on the subject with a clinical link are therefore needed to enable it to be used in the SLT clinic.

Methods of DA

Based on the postulates of Vygotsky's socio-cognitive theory (1978; notably the concept of 'zone of proximal development'), assessments of learning potential were explored from the mid-20th century, in particular for IQ tests in psychology (see also Lidz, 1987 for a historic perspective). There is considerable debate about the definition and classification of DA types, which depend, among others, on the objectives pursued: diagnosis or intervention (Hessels-Schlatter & Hessels, 2009). For diagnosis, a learning task with standardised training is usually proposed, often making use of pre-defined cues (Hessels-Schlatter & Hessels, 2009). A frequent format consists of three phases: static pre-test, training, and static post-test (e.g. Hasson et al., 2013; Petersen et al., 2020), allowing for pre- and post-tests. However, when training the child on new, invented language materials, pre-tests are not necessary (e.g. Roseberry & Connell, 1991). Studies' analyses focus on post-test results (e.g. Hasson et al., 2013), or on the change score if the DA contains non-invented language (pre to post-test change, e.g. Petersen et al., 2017). In several studies (e.g. Hasson et al., 2013), post-test score analyses show significant differences between a group of children with DLD and a group of TD children, which does not always seem to be the case for change score analyses (e.g. some scores in Petersen et al., 2017; see; Hunt et al., 2022 for a summary). Diagnostic accuracy analyses have also been carried out in some studies, with results ranging from poor to very good (see Hunt et al., 2022). Studies focusing on existing languages (as opposed to newly invented ones) have shown higher sensitivity percentages (e.g. see Petersen et al., 2017; Roseberry & Connel, 1991).

The nature of the training phase can vary as well, ranging from pre-defined explicit mediation units (e.g. Peña et al., 2014) to cue hierarchies (Campione & Brown, 1987), given to the child interactively depending on his/her reactions (e.g. Camilleri & Law, 2007). Other procedures only involve minimal feedback during the training phase, reducing the use of verbal language (Hessels, 1997) to neutralise potential linguistic biases (e.g. a computerised procedure in Skoruppa, 2019). Finally, studies vary in their analysis of the training, some assessing the number and type of cues needed (e.g. Camilleri & Law, 2007), while others using more global modifiability scales (in which the clinician details and rates their observations as to their teaching effort and the learner's response e.g. Petersen et al., 2020). Several studies have shown significant differences between TD and children with DLD when comparing the number and types of cues provided (e.g. Camilleri & Law, 2007). Finally, modifiability scales have been shown to be relatively valid (see Orellana et al., 2019) but further investigation seems necessary as clinicians sometimes appear to have cultural biases and their responses may be modulated by their awareness of the clinical status of participants (Hunt et al., 2022). In conclusion, it seems promising to combine several measures, for example post-test results and process analyses (cues given and/or modifiability scales, Hunt et al., 2022).

Just as in children's daily life, DA training phases increasingly use digital media. For preschool children, Van Horn and Kan (2015) found similar novel word learning effects for a TV cartoon and live story reading (but see Lytle et al., 2018, for a discussion of the superiority of human interaction especially for very young children; and Sysoev et al. (2022) for limits of automatised scaffolding for children with weaknesses). To sum up, learning situations can take many forms, and it seems that it is the procedure used (e.g. type of scaffolding) and the objective pursued, in relation to the child's profile, that has an influence rather than the situation itself, as long as it is motivating for the child (Budoff, 1987).

We will now turn to the different linguistic domains and item types for which DA-type procedures have been used in the SLT domain and detail the existing literature on performance differences between bilingual (or monolingual) children with typical development (TD) and children with (or at risk of) DLD (for a synthesis see also Hunt et al., 2022; Orellana et al., 2019).

Word learning

In the lexical domain, word learning skills are a well-studied candidate for DA: Many studies involving 'fast mapping' or 'quick incidental learning' have shown that through only a few presentations of new word-meaning mappings, TD children already build up rudimentary phonological and semantic representations, whilst children with DLD are known to have more difficulties in learning new words (for a review see Kan & Windsor, 2010). At the same time, novel word learning skills are largely similar in mono- and bilingual children (Matrat et al., 2022; but see Skoruppa, 2019; Yoshida et al., 2011 for slight bilingual advantages as well as Alt et al., 2013, 2019 for slight bilingual disadvantages).

Thus, such tasks can provide high diagnostic accuracy in distinguishing children with TD and DLD regardless of their linguistic status (Kan & Windsor, 2010; Petersen et al., 2020). Children with DLD typically require more presentations and more cues for learning new words, which has been related to their typical deficits in short-term verbal memory (Gray, 2003) and difficulties with semantic analysis (Nash & Donaldson, 2005). The effect of clinical status seems to be greater in receptive tasks than in productive tasks (Kan & Windsor, 2010) and for verbs than for nouns (Alt et al., 2004), which may be due to the higher syntactic load of verbs. For verb learning, however, both groups seem to be sensitive to the presence of morphophonological markers (Alt et al., 2004; Gray, 2003).

Morphosyntactic learning

Fewer learning studies have focused on the morphosyntactic domain, although it is widely recognised as one of the biggest challenges for children with DLD (Bishop et al., 2017). Again, monolingual and bilingual children's morphosyntactic learning abilities seem comparable, as Kaushanskaya et al. (2017) showed for the ability to learn a new derivational morpheme marking a part-whole distinction. Kohnert and Danahy (2007), however, caution that the performance of bilingual children depends on the morphological systems of the languages they speak.

With respect to DLD, an older study by Roseberry and Connell (1991) showed that bilingual children with DLD struggled significantly (in comparison with bilingual children with TD) to learn to understand and produce new morphemes marking said part-whole distinction. Hasson et al.'s (2013) expressive syntax task with mediation and gradual scaffolding targeted different skills, that is, an increase in the number of syntactic elements used per sentence. As expected, TD children required less mediation than children with DLD, whilst also showing higher improvement at post-test. However, there were already group differences between TD and children with DLD in the static pre-test, making further comparisons challenging. Finally, Hadjadj et al. (2022)

proposed gradual cues for teaching French morphosyntactic structures to monolingual and bilingual children. Their results highlight the accuracy of their task in distinguishing between TD and children with DLD. The study found no differences between bilingual and monolingual children.

A few authors (see the meta-analysis of Hunt et al., 2022) have proposed to analyse several language domains in their DA. This is the case of Petersen et al. (2017) who carried out analyses on grammar and discourse elements in narration. DA in several language domains is recommended to cover the different areas that may be affected by DLD, which is known to be heterogeneous, and thus avoid children appearing as false positives or false negatives when testing only one domain (see Hunt et al., 2022).

To sum up, there is a growing interest for different types of DA in the SLT literature, although few studies have directly compared different procedures in terms of adult mediation and linguistic domains and modalities. In the lexical domain, word learning has been particularly researched, but more work is needed on the morphosyntactic domain, which is frequently affected by DLD. Very few studies have looked at a combination of several language domains (see Petersen et al., 2017 for an exception). Moreover, only a few studies (e.g. Camilleri & Law, 2007) compared monolingual and bilingual populations with and without DLD. Finally, not all DA studies have calculated the diagnostic accuracy (that is, the sensitivity and specificity) of their tasks (beyond group differences). For those who did, Hunt et al. (2022) usually note good classification accuracy, that is, over 80% diagnostic accuracy following Plante and Vance (1994), as in Petersen et al. (2017). However, so far, no conclusive standardised and diagnostically accurate clinical applications have been constructed.

Aims of the study

The present study aims to fill, at least partially, these gaps identified in the literature. Our work can be understood as an early-efficacy study, in which we will assess the clinical potential of different learning situations, tasks and items cross-sectionally in predefined groups of already diagnosed children. We will compare two forms of DA in a population of French- and Portuguese-speaking children aged between 5;0 and 7;11 years, with and without previously diagnosed DLD. We chose this language combination because it is one of the most frequent ones in French-speaking Switzerland (Federal Statistical Office, 2020), where our study was run, especially in an SLT context (Volpin et al., 2020), and we selected this age range because children with DLD are frequently identified at these ages (Bishop et al., 2017).

We are interested in the learning of news words, both nouns and verbs, but also in the area of morphosyntax acquisition, which is less studied in DA research. In addition to productive tasks, we want to test the receptive modality because it seems to be less subject to floor effects and participants' language status (Alt et al., 2013). Finally, we also want to compare two ecological learning situations that are frequent in children's lives (interactive story reading and autonomous computer game), and that vary in the procedure used and adult assistance provided. Our objective being diagnosis, we want to determine which of our tasks best predict the clinical status of participants (TD or DLD). Specifically, we will attempt to answer the following research question aiming the diagnosis of DLD in monolingual and bilingual children:

Which learning situation (interactive story reading vs. autonomous computer game), language modality (productive or receptive), task (word-picture matching, free recall, naming, judgement acceptability) and item type (nouns, verbs, inflections in sentences) contribute most to predicting the clinical status of our participants? In addition, (a) do any participant characteristics (language status, parental education, age) contribute to predicting their clinical status? And (b) what are the sensitivity and specificity values for the most promising task combinations?

Method

Participants

Seventy-two children aged 5;0 to 7;11 living in the French-speaking part of Switzerland participated in the study. They had no history or suspected cognitive, socio-affective or hearing impairment or neurodevelopmental disorders. The parents of the participants completed a questionnaire (based on Tuller, 2015, see Supplementary material 1). In order to calculate the exposure rate and determine the child's dominant language, the parents indicated the frequency of use of French and Portuguese at home and at school (French always being the language of instruction at school). Parents also indicated the age of the children's first contact with the languages and their highest level of education (see also additional information about the participants in the Supplementary Material 8). According to the parental questionnaire 36 children were exposed to French at least 85% at the time of testing and thus classified as monolingual. The remaining 36 bilingual children had a minimum of 25% exposure to Portuguese at the time of testing (and no more than 15% exposure to a third language, following Byers-Heinlein & Lew-Williams, 2013).

In order to minimise the chicken-and-egg problem of initial diagnosis common in the field of bilingual assessment, we decided to use a strict two-step identification procedure in order to classify children as having a confirmed DLD. To meet the criteria for inclusion in our study, children with DLD firstly had to have been clinically diagnosed by a SLT and have received therapy for at least 3 months (to rule out spurious initial over-identification). Secondly, they also had to score clearly below age-equivalent norms on static standardised language tests (see details below) carried out alongside our experimental tasks, equal to or below −1.28 *SD* (following Thordardottir et al., 2011) or percentile 10 on at least one subtest per language, to exclude pure lack of exposure to one language in bilinguals (Peña et al., 2016). To be equally sure about the status of children classified as having a TD, they should never have been identified as having DLD, never have received SLT nor fail any standardised pre-test. In addition, all children passed an audiometric screening (audibility thresholds ≤20 dB at 0.5, 1, 2 and 4 kHz) and a non-verbal intelligence screening, using the matrices subtest of the Wide-Range Intelligence Test by Glutting et al. (2000).

Based on these criteria, 12 participants were excluded from the analyses because they did not score as expected on the language pre-tests (3 monolinguals and 9 bilinguals, 4 presumed TD and 8 presumed DLD), highlighting once more the difficulty of DLD diagnosis in bilinguals. A further 11 participants were excluded because they failed the audiometry ($n = 3$) or the non-verbal intelligence ($n = 2$) screening, did not fit the language exposure criteria ($n = 2$) or due to a technical problem at the time of testing ($n = 4$).

Table 1. Participant demographic information.

Clinical status	Linguistic background	Number	Gender G	Gender B	Age in months	Parental education	[a]Non-verbal IQ	Bilingualism [b]simultaneous	Bilingualism successive	Current dominance FR	Current dominance PT
TD	Monolingual	15	11	4	75 (11.83)	3.90 (1.08)	53.20 (20.06)				
DLD	Monolingual	13	8	5	76 (9.27)	3.10 (0.68)	47.30 (18.64)				
TD	Bilingual	11	6	5	85 (9.59)	3.00 (1.31)	64.90 (24.42)	7	4	9	2
DLD	Bilingual	10	7	3	81 (10.89)	2.60 (0.57)	48.00 (28.90)	7	3	7	3

Abbreviations: G = girls; B = boys; FR = French; PT = Portuguese.
Numbers in parentheses are standard deviations.
[a]Raw score.
[b]Simultaneous bilinguals = exposed to both languages from birth or before age of three; successive bilinguals = exposure to a second language after 3 years (Paradis, 2007).

Table 1 summarises the demographic information of our final sample, including details on parental education (on a scale from 1 to 5: 1 = elementary school, 2 = secondary school, 3 = apprenticeship, 4 = high school graduation, 5 = university), the type of bilingualism (simultaneous vs. sequential) and children's dominant language. There was no difference in age ($H = 5.80$, $p = 0.121$), non-verbal IQ ($H = 4.20$, $p = 0.240$) nor gender (X^2 (3, n = 49) = 1.14, $p = 0.767$), between the four participant groups in Kruskal-Wallis and Chi-square tests, respectively. However, we found a significant difference in parental education between the bilingual DLD (mean: 2.60) and the monolingual TD children (mean: 3.90; $H = 14.27$, $p = 0.003$), the other groups did not differ significantly in this respect. Note that we controlled for all of these parameters in our analysis (see details below).

Recruitment procedure

Participants were recruited through their schools and SLTs. Families interested in participation received an information sheet, a parental report and an informed consent form, validated by the local ethics commission (Swiss ethics Vaud; project no.: 2017-01881, authorisation granted on 7 February 2018).

Static language tests used to confirm clinical status

We chose three representative language tasks often used in the clinic, evaluating different domains and language modalities that are known to be difficult for children with DLD (Bishop et al., 2017): (1) picture naming for the productive lexicon, (2) oral comprehension of sentences for receptive morphosyntax, and (3) NWR for phonological processing. The lexical and syntactic domains are frequently assessed to identify a DLD (Denman et al., 2017) and NWR seems to be a discriminant processing task in the field of phonology (Estes et al., 2007; Schwob et al., 2021). The tasks were taken from two standardized French tests: Evaluation of Oral Language ('ELO', Khomsi, 2001) for naming and comprehension and Exalang for non-word repetition (Helloin & Thibault, 2006; Thibault & Helloin, 2010) and from two standardised Portuguese tests: The Evaluation of Oral Language test ('ALO', Sim-Sim, 2006) for naming and comprehension and from Language and Aphasia

Assessment Tests in Portuguese ('PALPA-P', Castro et al., 2007) for NWR. Children were tested in three sessions, mostly in the family home or at school. They participated in an additional cross-linguistic NWR task (for details, see Schwob & Skoruppa, 2022).

DA materials

We developed two DA situations that will be detailed in the following. Each situation was run during a separate session and consisted of having the children learn new invented nouns, verbs, and inflections in different phases, and then evaluating their learning via different tasks. All participants completed both DA situations. The order of stimuli lists (non-word list A or B) and learning situations (autonomous computer game or interactive story reading) was counterbalanced across children.

Stimuli

For word learning, auditory stimuli were six non-words per list that do not exist in French or Portuguese, in order to avoid exposure bias (Kapantzoglou et al., 2012). We took care to ensure that they were well distinct from each other and varied in length (one and two syllables) and in complexity (with or without cluster). We compose our non-words with vowels common to both languages (/a,i,o,u/) as well as early-acquired consonants common to both languages (/l, t, f, v, b, p, s, k, n/; following Chiat, 2015). Our items respect the phonotactic patterns of both French and Portuguese. The stimuli and their illustrations are presented in Supplementary Material 2. Specifically, there were three nouns to be learned per situation (list A or list B). Two were designed as masculine and ended with the vowel /i/ (/fli/, /lati/) and one was feminine and ended with the vowel /u/ (/plifu/). Visually, two nouns were represented by aliens, and one was inanimate object. Three verbs were learned per situation (list A or list B). These were created to represent actions that cannot be named by a single verb in both French and Portuguese (e.g. turning and descending at the same time). The verbs ended with the consonant /t/ (/flinat/, /vat/, /aplut/). We hypothesised that these morphophonological cues could help children acquire these new words (Alt et al., 2004; Gray, 2003). One verb was intransitive (involving one alien), and two verbs were transitive (involving one alien and one object). Finally, the children also learned sentences by associating previously learned nouns and verbs. We generated a total of six sentences per situation. In order to test morphosyntactic learning skills, we added an inflectional rule not present in either Portuguese or French: the verb ended with the consonant /t/ for masculine subjects (e.g./lati flinat/), and with the consonant /m/ for feminine subjects (e.g./plifu flinam/). Note that in both languages, the gender of the subject is indicated by their determiner and by adjective and noun endings but is not marked on the verb.

DA procedure

In each learning situation, the children learned the new items in three consecutive blocks (one for nouns, one for verbs and one for the morphosyntactic rule in sentences) containing different phases (adapted from Skoruppa, 2019, see Figure 1), presented on a laptop in front of the child via Microsoft Power-point (2018). During the first ***receptive exposure phase***, children were presented with images appearing and moving on the screen and heard the corresponding labels. The items were presented six times each in this first phase (following the

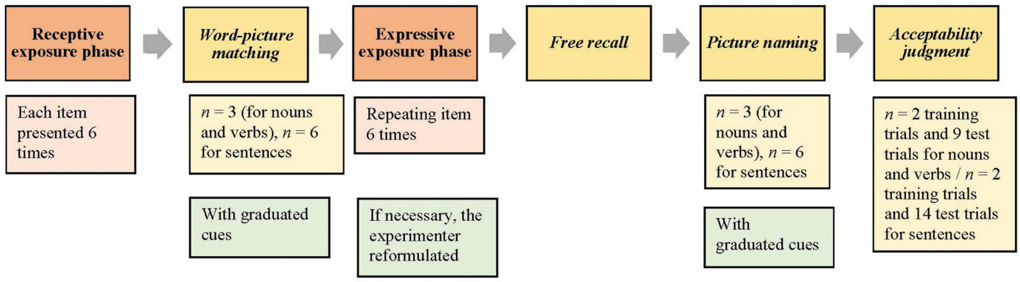

Figure 1. Exposure phases and assessment procedures. Note. Yellow was used to represent the assessment tasks and orange to identify the exposure phase. The specificities according to the interactive reading situation are mentioned in green.

recommended exposure in the meta-analysis of Kan & Windsor, 2010). Immediately after these first exposures, children were asked to complete a ***word-picture matching task***, to evaluate if the new items were acquired receptively. This task included three pictures for nouns and verbs (two distractors and the correct item) and four pictures for sentences (three distractors and the correct item). The children then participated in a second ***expressive exposure phase,*** which consisted of repeating the new items six times each, while the referents appeared on the screen. After this, the children completed a ***free recall task*** and a ***picture naming task,*** to evaluate if they could produce the names of the items learned. Finally, an ***acceptability judgment task*** was proposed to evaluate the accuracy of the children's semantic and phonological representations. An audio item was presented to the child, and they had to judge whether it corresponded to the item presented visually. Semantic (e.g./lati/ as a distractor for /plifu/), phonological (e.g./plifi/ for /plifu/) and syntactic (for sentences; order distractor, e.g./vam otu/ for /otu vam/; inflection distractor, e.g./otu vat/ for /out vam/) distractors were inserted (see complete list in Supplementary Material 3). Before starting this task, two practice trials with unrelated distractors and corrective feedback were proposed in order to familiarise the child with the procedure. Each situation (with these different phases and item types) required approximately 45 min, including a short break.

Situation 1: Autonomous computer game

The first situation was designed to involve only minimal French (very simple instructions), thus avoiding linguistic biases as much as possible. The child was immersed in an alien world where they would learn a new 'language'. The new items were recorded by a female Swedish speaker, so that none of the children would have an advantage in terms of the intonation and accent used to produce the items. The speaker produced a tonic accent on the last syllable, as is possible in French and Portuguese. During the adventure, the children were accompanied by an alien teacher, Mrs. Flamap, who introduces them to the different items on the screen and rewarded them with stars at the end of each learning phase. Instructions were pre-recorded in French by a male speaker. The experimenter only recorded the child's receptive responses via mouse-click and gave general encouragement, but there was no feedback (except for the practice trials in the acceptability judgement task).

Situation 2: Interactive story reading with mediation

The second situation included more child-experimenter interaction, even if it also involved screen-based presentation. For this situation, the experimenter was the same for all the children (the first author, who is a trained SLT). The experimenter had an active role (Hasson & Joffe, 2007), performing the live reading of three stories in French (one story per block) forming a logical sequence, with aliens who visit other planets and learn magic gymnastics figures (see an example in Supplementary Material 4). The new items were embedded into these stories, which offered semantic cues to facilitate learning (three cues per non-word, see Supplementary Material 5 for examples). Each time a new non-word was presented, the experimenter accentuated it verbally and pointed to the corresponding picture or movement on the screen (Weismer & Hesketh, 1993). The child and the experimenter interacted during the receptive training (story reading), and when repetition errors were detected during the productive training, the experimenter reformulated them. Finally, the experimenter provided graduated cues (Campione & Brown, 1987) during the word-picture matching and naming tasks in order to observe their learning potential with assistance. The cues were first very general and then became more and more specific and detailed, ranging from semantic cues over first-phoneme cues to full repetition (see Supplementary Material 5 for details on the cues and an analysis of their effect on children's performance).

Inter-rater reliability

Inter-judge agreement indices on the number of correctly produced/or designated novel items on 10% of our sample were calculated, using the statistical software SPSS (Version 27.0, IBM Corp., 2020) and Cohen's Kappa test. We obtain an agreement of .98 indicating a near perfect agreement (Landis & Koch, 1977).

Results

Descriptive statistics

Table 2 presents mean scores (number of items corrects) and standard deviations of our four groups of children (monolinguals and bilinguals, with and without DLD). We notice that in general, children with DLD have lower scores than children with TD, whereas monolinguals and bilinguals' performance seem generally comparable. We also note that the performance of the different groups on the autonomous computer game situation is lower than on the interactive reading situation. Finally, some group means are close to or equal to 0, e.g. for the sentence naming subtests.

Diagnostic accuracy for DLD

Since our main aim was to determine the most discriminant elements of our two learning situations, we will not present group comparisons for each situation, task and item type in detail. A full set of statistical analyses of these, via generalised linear models for each situation (autonomous computer game and interactive story reading), task (e.g. word picture matching, naming, . . .), and item type (nouns, verbs, sentences) can be found in

Table 2. Summary of results in autonomous computer game and in interactive story reading situations.

	Items	TD MON	TD BIL	DLD MON	DLD BIL
Autonomous DA					
[a]Word-picture matching (max 3)	Nouns	2.27(0.88)	2.82(0.60)	1.62(1.04)	2.40(0.96)
Word-picture matching (max 3)	Verbs	1.67(1.17)	2.55(0.68)	1.38(1.04)	1.90(1.28)
Word-picture matching (max 6)	Sentences	2.40(1.50)	3.27(1.95)	1.92(1.11)	2.10(0.73)
Recall (max 3)	Nouns	1.00(1.00)	1.18(0.75)	0.54(0.96)	0.90(1.28)
Recall (max 3)	Verbs	0.87(0.91)	0.64(0.80)	0.46(0.77)	0.40(0.69)
Recall (max 6)	Sentences	0.27(0.59)	0.09(0.30)	0.00(0.00)	0.10(0.82)
Naming (max 3)	Nouns	0.80(0.77)	1.36(1.02)	0.31(0.48)	0.70(0.82)
Naming (max 3)	Verbs	0.53(0.64)	0.36(0.50)	0.15(0.37)	0.30(0.46)
Naming (max 6)	Sentences	0.13(0.35)	0.00(0.00)	0.00(0.00)	0.00(0.00)
Acceptability judgement (max 9)	Nouns	6.33(1.67)	8.27(1.00)	5.54(2.84)	6.50(1.35)
Acceptability judgement (max 9)	Verbs	5.73(1.22)	6.18(1.07)	5.15(2.07)	5.90(1.66)
Acceptability judgement (max 14)	Sentences	9.40(1.68)	9.64(1.69)	7.92(1.70)	8.30(2.00)
Interactive DA					
Word-picture matching (max 3)	Nouns	2.53(0.64)	2.09(0.70)	1.77(0.59)	2.30(0.82)
Word-picture matching (max 3)	Verbs	2.53(0.51)	2.36(0.92)	2.15(0.89)	2.20(0.63)
Word-picture matching (max 6)	Sentences	4.47(1.30)	5.27(0.90)	3.08(1.38)	3.10(1.59)
Recall (max 3)	Nouns	1.20(1.14)	1.36(0.80)	0.46(0.66)	0.40(0.51)
Recall (max 3)	Verbs	1.13(0.64)	1.18(0.60)	0.38(0.50)	0.10(0.31)
Recall (max 6)	Sentences	0.73(1.10)	0.45(0.68)	0.00(0.00)	0.00(0.00)
Naming (max 3)	Nouns	1.20(1.20)	0.64(0.80)	0.46(0.66)	0.40(0.51)
Naming (max 3)	Verbs	1.13(0.83)	1.09(0.53)	0.38(0.50)	0.30(0.48)
Naming (max 6)	Sentences	0.53(0.91)	0.36(0.50)	0.00(0.00)	0.00(0.00)
Acceptability judgement (max 9)	Nouns	7.67(1.34)	8.36(0.50)	5.69(1.65)	7.60(0.96)
Acceptability judgement (max 9)	Verbs	6.80(1.01)	7.55(0.93)	5.38(1.66)	4.70(1.05)
Acceptability judgement (max 14)	Sentences	9.60(2.44)	10.82(1.94)	8.00(1.68)	8.20(1.81)

Abbreviations: DA = dynamic assessment; Autonomous DA = autonomous computer game situation; Interactive DA = interactive story reading situation; DLD = developmental language disorder ; TD = typical language development.
[a]The tasks are arranged according to the chronological order of the assessment. The word-picture matching and acceptability judgement are receptive tasks. The others are expressive tasks.
N = 49 (*n* = 23 children with a DLD and 26 children with a TD/*n* = 21 bilingual children and 28 monolingual children).
We present the mean number of items corrects resp. mean number of cues given by group, with standard deviations in parentheses.

Supplementary Materials 6. In summary, for the autonomous computer game situation, few significant effects emerged, both for clinical status (3/14 sub-tasks) and for linguistic status (3/14 sub-tasks). For the interactive story reading situation, most sub-tasks showed significant effects of clinical status (10/14 sub-tasks), and fewer showed significant effects of linguistic status (5/14 sub-tasks).

In order to determine the overall diagnostic potential of our procedure, we rather focussed on a logistic regression, with a binary indicator variable for disorder (TD vs. DLD children), like previous studies in the field including many different measures (e.g. Boerma et al., 2016). All 28 sub-tasks were used as potential explanatory variables, complemented by age (in months), parental education and linguistic status (monolingual/ bilingual). Using the R package glmnet (Friedman et al., 2010; Simon et al., 2011), we implemented a Lasso penalty as proposed by Tibshirani (1996), in order to reduce the number of variables involved and to retain only those that variables were most important for predicting whether a child has a DLD or not. Since we had missing data on parental education of some monolingual children (*n* = 10), we performed imputation of missing data by group average to generate these observations and control for them in our statistical model.

The Lasso logistic regression reduced the number of important variables to nine, with an initial penalisation parameter λ[1] = 0.041. The estimated regression coefficients for these variables that contribute to the disorder classification are presented in Table 3.

In Table 3, we note that these variables mostly concern the interactive story reading situation, with only two tasks of the autonomous computer game situation retained. Concerning the variables related to the children's characteristics, we note both age and parental education seem to have an influence. Furthermore, the confusion matrix derived from the classification model using the variables identified by the Lasso regression shows that only one child is misclassified by this analysis, yielding 100% of sensitivity and 96% of specificity in terms of diagnostic accuracy (see Supplementary Material 7).

For small sample sizes, the identification of the penalty parameter by means of cross-validation can be relatively unstable (Riley et al., 2021; Roberts & Nowak, 2014). For this reason, we have identified another model by increasing the Lasso penalty parameter (to λ = 0.16), in order to obtain a more parsimonious model, which removes the influence of participants' characteristics (e.g. age and education of parents) and reduces the variables to four. These variables only concern in the interactive story reading situation (see Table 4). However, this model misclassifies six children, yielding 86% of sensitivity and 88% of specificity. Of these children, five are monolingual (three with a DLD and two with a TD) and only one child is bilingual (with a TD).

Table 3. Regression coefficients for the Lasso logistic regression with λ = 0.041.

Variable	*b*
Intercept	11.8535
Interactive story reading situation – recall verbs	−0.9805
Interactive story reading situation – acceptability judgement verbs	−0.7395
Parental education	−0.5726
Interactive story reading situation – word picture matching sentences	−0.4371
Interactive story reading situation – acceptability judgement nouns	−0.3273
Autonomous computer game – acceptability judgement sentences	−0.2733
Interactive story reading situation – word picture matching nouns	−0.0239
Autonomous computer game – word picture matching sentences	−0.0581
Age	0.0223

N = 49 (*n* = 23 children with a DLD and 26 children with a TD/*n* = 21 bilingual children and 28 monolingual children).

Table 4. Regression coefficient for the Lasso logistic regression with λ = 0.16.

Variable	*b*
Intercept	2.6407
Interactive story reading situation – recall verbs	−0.5944
Interactive story reading situation – acceptability judgement verbs	−0.2890
Interactive story reading situation – word picture matching sentences	−0.1330
Interactive story reading situation – acceptability judgement nouns	−0.0048

N = 49 (*n* = 23 children with a DLD and 26 children with a TD/*n* = 21 bilingual children and 28 monolingual children).

[1]The nonnegative tuning parameter λ controls the amount of shrinkage (Riley et al., 2021). In practice, the lasso is tested for different values of λ. The value is chosen automatically using a cross-validation method to maximise the predictive capacity.

Discussion

The aim of our research was to compare two DA situations, as well as different tasks and item types, and determine which one(s) are best suited for the diagnosis of DLD in a population of French- and Portuguese-speaking children aged between 5;0 and 7;11 years. Our results show that, even with a small number of DA tasks, we can detect the presence of a DLD in our linguistically diverse population relatively well. Although children also learned some new linguistic elements during our computerised procedure, our interactive story reading situation seemed most relevant to predict DLD, as both our Lasso regression models retained more elements from this part of the test. It is possible that the autonomous computer game was too complex for all groups of children, as it was done with less interaction and without graduated cues. These elements make it possible to highlight the fact that it is not the situation itself that plays a role, as long as it is motivating for the child (Budoff, 1987), but rather the procedure used (with graduated help, in interaction).

Both models classified our sample into TD vs. DLD children with acceptable diagnostic accuracy (according Plante & Vance, 1994), like some previous studies in the field (e.g. Petersen et al., 2017, 2020). Furthermore, our diagnostic accuracy scores are relatively good for DA with new, invented language materials, which have been little explored in previous studies (see Hunt et al., 2022). Unsurprisingly, the first model containing more variables showed an advantage and reached good accuracy (>90% of sensitivity and specificity, following Plante & Vance, 1994), whereas the second, more parsimonious model showed fair accuracy (90% of sensitivity and specificity, following Plante & Vance, 1994). Despite its disadvantage, the second model seems more promising for the SLT clinic as it would allow for a single learning situation and does not rely on further participant characteristics (age and parental education), potentially facilitating the establishment of clinically useful norms. However, this version is hypothetical and has not been tested in this form, and both situations should clearly be replicated with larger, representative samples. We can also note that in both regression models, linguistic status never appears as a primary predictor of clinical status (even in the interactive story reading task which is based on French), despite the slight effects we found in the group comparisons (see Supplementary Materials 6) for some sub-tasks. Moreover, interactions between linguistic status and clinical status are only visible in two sub-tasks that are not included in our second regression model (word-picture matching for nouns and naming of nouns), and that should be excluded from any further, shortened version. In summary, our analyses seem to confirm that DA is less prone to language bias than static tests (Maragkaki & Hessels, 2017), making it an ideal tool for the bilingual SLT clinic.

Furthermore, our study provided new results on DA in French, a language that is still much less studied than English at present. Other recent research on French (e.g. Hadjadj et al., 2022; Matrat et al., 2022) has found similar results suggesting that the children's difficulties in DA situations are inherent to the presence of DLD and not to other characteristics.

As for linguistic domains and item types, it is a combination of different tasks (especially acceptability judgement and word-picture matching, but also recall) and items (especially verbs, but also nouns and sentences) that seems to contribute to diagnostic accuracy. Most current studies have focused on different language domains within DA, showing their effectiveness (e.g. morphosyntax- Roseberry & Connell, 1991; lexical- Matrat et al., 2022).

A few studies (see meta-analysis by Hunt et al., 2022) have tried to combine several language domains in their DA (e.g. Petersen et al., 2017 with an analysis of grammar and discourse in narrative), this process seems the most promising according to our study. The most predictive tasks seem to be receptive (except for verbs recall), confirming previous review literature (Kan & Windsor, 2010). Is it possible, again, that production is more difficult for all children, and some of these tasks yielded floor effects. These floor effects may have been found because the children had to learn new (invented) language structures, which is not the case in some earlier studies that did not find such effects (e.g. Hadjadj et al., 2022). Our results on new structures to learn encourage the use of receptive tasks in DA which are also very easy to score automatically. Children retained only a small number of items, especially when it came to verbs and sentences, which is consistent with the existing literature that nouns seem to be more easily learned (e.g. Alt et al., 2004). The free recall task, however, appears to be more accessible to children than classic picture naming, and our results show that it is useful for diagnosis.

Our supplementary analysis of children's use of cues shows that, as in several other studies (e.g. Hasson et al., 2013; Maragkaki & Hessels, 2017), children with DLD need more support. Following Hunt et al. (2022), combining our product scores (number of correct responses) and the cues provided to the children could lead to further improvements in diagnostic accuracy. Furthermore, the amount of improvement and the number of cues needed may also give important information with regards to a child's intervention planning. Finally, it should be noted that the help provided to the children undoubtedly contributed to their motivation during the situation. However, our analyses show that, for DLD diagnosis it may be sufficient to rely on the number of correct items as per the child's first response, which would be much easier to keep track of in a standardised DA battery than if all individual cues would need to be counted and analysed.

In several studies, the authors note that DA is not well known in the clinical setting and appears to be time-consuming for practitioners (Petersen et al., 2017). However, the time required to perform our assessment was not excessive (about one 45-min session per situation) since there was no need for pre-tests, encouraging its use in the clinic (Hasson & Joffe, 2007). Our second model suggests that this time could be further reduced to a single session, with minor loss in diagnostic accuracy. Note that other authors have already tested a short version of their DA showing good diagnostic potential in the narrative domain (Petersen et al., 2017), and encourage the reduction of time to perform a DA in SLT clinic. We think that our (shortened) DA assessment in the lexical and morphosyntactic domains could be of interest for clinical test development, if it can be validated in further studies.

Limitations and further research

Nevertheless, our current results should be interpreted with caution before being replicated with a larger sample size. Another limitation of our work is that we used pre-existing (potentially biased) SLT classifications and standardised pre-tests with monolingual norms in both languages of the child to form our initial diagnostic groups. We have chosen to categorise our monolingual and bilingual children based on a quantitative current exposure score, using a shortened version of the PabiQ parental questionnaire (Tuller, 2015). While it is true that we have little information on longitudinal cumulative exposure and on the richness of the exposure received (Paradis, 2023), in most of our tasks we do not even find differences between monolinguals and bilinguals, suggesting that they are indeed independent of exposure. Of course, further research will be necessary to confirm our results, ideally in a larger prospective study, where our tasks would be run before assessing the clinical status of the child, thus making them more objective and enabling full blinding of the professionals involved. For clinical use, it would also be necessary to replicate this study with different language profiles and/or sharing other language combinations to see if the results obtained here are generalisable. In fact, the interactive story reading situation has shown potential for monolingual and bilingual children exposed to French aged between 5–7;11 years old. We therefore do not know whether our task would be as relevant with younger children with less exposure to French (e.g. younger preschool children, new arrivals), or if a simplified version of autonomous computer situation with less French would be more discriminant. This is another area that should be investigated in future studies. Moreover, it would also be interesting to add a more subjective 'modifiability scale' to be completed the experimenter to assess if it can further improve diagnostic accuracy, as was the case in Petersen et al. (2017).

Finally, our method could be combined with other tasks that seem promising for the identification of a DLD. For example, NWR tasks show good diagnostic accuracy in monolingual (Estes et al., 2007) and bilingual (Schwob et al., 2021) populations. Boerma and Blom (2017) found good percentages of diagnostic accuracy of a combination of a NWR, a parental questionnaire and a narrative task. It could be very interesting to add our interactive DA situation to this promising mix.

Conclusion

The results of this study show that DA allows us to identify with good diagnostic accuracy which children successfully master one or two languages, and which children struggle with a DLD. We show that a relatively short, interactive learning situation with adult scaffolding, using a variety of new linguistic materials (nouns, verbs and inflections in sentences) in receptive and productive tasks, is diagnostically very accurate for young French and French-Portuguese children, and thus holds a promising potential for a valid diagnostic procedure to detect DLD in young monolingual and bilingual learners.

Acknowledgments

We would like to thank all participants and their families for taking part in this research, as well as colleagues, students and members of institutions who allowed us to contact the families. Parts of a prior version of this text were translated from French by the free version of the software Deep-L.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Alt, M., Arizmendi, G. D., Gray, S., Hogan, T., Green, S., & Cowan, N. (2019). Novel word learning in children who are bilingual: Comparison to monolingual peers. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2332–2360. https://doi.org/10.1044/2019_JSLHR-L-18-0009

Alt, M., Meyers, C., & Figueroa, C. (2013). Factors that influence fast mapping in children exposed to Spanish and English. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1237–1248. https://doi.org/10.1044/1092-4388(2012/11-0092)

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research*, *47*(2), 407–420. https://doi.org/10.1044/1092-4388(2004/033)

American Speech-Language-Hearing Association. (2024, May 5). *Multilingual service delivery in audiology and speech-language pathology*. Multilingu.

Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI – The nature of the problem*. *Bilingualism: Language and Cognition*, *15*(1), 1–4. https://doi.org/10.1017/S1366728911000599

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (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. https://doi.org/10.1111/jcpp.12721

Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not possible? *Journal of Communication Disorders*, *66*, 65–76. https://doi.org/10.1016/j.jcomdis.2017.04.001

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language & Communication Disorders*, *51*(6), 626–638. https://doi.org/10.1111/1460-6984.12234

Budoff, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic assessment. An interactional approach to evaluating learning potential* (pp. 173–195). Guilford Press.

Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the early years: What the science says. *LEARNing landscapes*, *7*(1), 95–112. https://doi.org/10.36510/learnland.v7i1.632

Camilleri, B., & Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, *9*(4), 312–322. https://doi.org/10.1080/14417040701624474

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–115). Guilford Press.

Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R., & Coltheart, M. (2007). *Provas de avaliação da linguagem e da afasia em português*. CEGOC.

Chiat, S. (2015). Non-word repetition. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 95–122). Multilingual Matters.

Denman, D., Speyer, R., Munro, N., Pearce, W., Chen, Y. W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4–12 years: A systematic review. *Frontiers in Psychology*, *1515*(8), 1–28. https://doi.org/10.3389/fpsyg.2017.01515

Estes, K. G., Evans, J., & Else-Quest, N. (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*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Federal Statistical Office. (2020, February 27). *Most frequent main languages*. https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.assetdetail.11947533.html

Friedman, J., Hastie, T., & Tibshirani, R. J. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Glutting, J., Adams, W., & Sheslow, D. (2000). *Wide range intelligence test*. Wide Range.

Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research*, *46*(1), 56–57. https://doi.org/10.1044/1092-4388(2003/005)

Hadjadj, O., Kehoe, M., & Delage, H. (2022). Dynamic assessment identifies morphosyntactic deficits in mono-and bilingual children with developmental language disorder. *Languages*, *7*(295), 1–26. https://doi.org/10.3390/languages7040295

Hasson, N., Camilleri, B., Jones, C., Smith, J., & Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy*, *29*(1), 57–75. https://doi.org/10.1177/0265659012459526

Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy*, *23*(1), 9–25. https://doi.org/10.1177/0265659007072142

Helloin, M. C., & Thibault, M. P. (2006). *Exalang 3-6*. Happyneuron.

Hessels, M. G. P. (1997). Low IQ but high learning potential: Why zeyneb and Moussa do not belong in special education. *Educational & Child Psychology*, *14*, 121–136.

Hessels-Schlatter, C., & Hessels, M. G. P. (2009). Clarifying some issues in dynamic assessment: Comments on Karpov and Tzuriel. *Journal of Cognitive Education and Psychology*, *8*(3), 246–251. https://doi.org/10.1891/1945-8959.8.3.246

Hunt, E., Nang, C., Meldrum, S., & Amstrong, E. (2022). Can dynamic assessment identify language disorder in multilingual children? Clinical applications from a systematic review. *Journal of Speech, Language, and Hearing Research*, *53*(2), 598–625. https://doi.org/10.1044/2021_LSHSS-21-00094

IBM Corp. (2020). *IBM SPSS Statistics for Macintosh (Version 27.0)* [Computer software].

Kan, P., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *53*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kapantzoglou, M., Restrepo, A., & Thompson, M. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in School*, *43*(1), 81–96.)

Kaushanskaya, M., Gross, M., Sheena, E., & Roman, R. (2017). Novel morpheme learning in monolingual and bilingual children. *American Journal of Speech and Language Pathology*, *26*(2), 301–315. https://doi.org/10.1044/2016_AJSLP-16-0011

Keller, K., Troesch, L., & Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology*, *6*(1284), 1–9. https://doi.org/10.3389/fpsyg.2015.01284

Khomsi, A. (2001). *Évaluation du langage oral*. ECPA

Kohnert, K., & Danahy, K. (2007). Young L2 learners' performance on a novel morpheme task. *Clinical Linguistics & Phonetics*, *21*(7), 557–569. https://doi.org/10.1080/02699200701374231

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lidz, C. S. (1987). Historical perspectives. In C. S. Lidz (Ed.), *Dynamic assessment. An interactional approach to evaluating learning potential* (pp. 3–35). Guilford Press.

Lytle, S. R., Garcia-Sierra, A., & Kuhl, P. K. (2018). Two are better than one: Infant language learning from video improves in the presence of peers. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(40), 9859–9866. https://doi.org/10.1073/pnas.1611621115

Maragkaki, I., & Hessels, M. G. P. (2017). A pilot study of dynamic assessment of vocabulary in German for bilingual preschoolers in Switzerland. *Journal of Studies in Education*, *7*(1), 32–49. https://doi.org/10.5296/jse.v7i1.10392

Matrat, M., Delage, H., & Kehoe, M. (2022). Dynamic assessment of word learning to diagnose developmental language disorder in French-speaking monolingual and bilingual children. *Languages*, *7*(3), 181. https://doi.org/10.3390/languages7030181

McLeod, S., Verdon, S., & the International Expert Panel on Multilingual Children's Speech. (2017). Tutorial: Speech assessment for multilingual children who do not speak the same language(s) as the speech-language pathologist. *American Journal of Speech Language Pathology*, *26*(3), 691–708. https://doi.org/10.1044/2017_AJSLP-15-0161

Multilingual-Multicultural Affairs Committee of the IALP. (2020). *Common question by speech and language therapists/speech-language pathologists about bilingual/multilingual children and* *informed, evidence-based answers*. https://files.japanslht.or.jp/notifications/2020/05/27/Common-Questions-by-SLT-SLPabout-Bilingual-Multilingual-Children-and-Informed-Evidence-based-Answers_2020.pdf

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, *48*(2), 439–458. https://doi.org/10.1044/1092-4388(2005/030)

Nicoladis, E., & Genesee, F. (1997). Language development in preschool bilingual children. *Journal of Speech-Language Pathology and Audiology*, *21*(4), 258–270.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Orellana, C., Wada, R., & Gillam, R. (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*(3), 1298–1317. https://doi.org/10.1044/2019_AJSLP-18-0202

Paradis, J. (2007). Becoming bilingual. In P. A. et L. Wei (Ed.), *Handbook of multilingualism and multilingual communication* (pp. 15–44). Mouton de Gruyter.

Paradis, J. (2019). English second language acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In M. Brown (Ed.), *Proceedings of the 43rd Annual Boston University Conference on Language Development* (pp. 11–26). Cascadilla Press.

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793–817. https://doi.org/10.1017/S0305000922000708

Peña, E. D., Bedore, L., & Kester, E. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Communication Disorders*, *51*(2), 192–202. https://doi.org/10.1111/1460-6984.12199

Peña, E. D., Gillam, R., & Bedore, L. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2208–3330. https://doi.org/10.1044/2014_JSLHR-L-13-0151

Petersen, D. B., Chanthongthip, H., Ukrainetz, T., Spencer, T., & Roger, S. (2017). Dynamic assessment of narratives: Efficient, accurate identification of language impairment in bilingual students. *Journal of Speech, Language, and Hearing Research*, *60*(4), 983–998. https://doi.org/10.1044/2016_JSLHR-L-15-0426

Petersen, D. B., Tonn, P., Spencer, D., & Foster, M. E. (2020). The classification accuracy of a dynamic assessment of inferential word learning for bilingual English/Spanish-speaking school-age children. *Language, Speech, and Hearing Services in Schools*, *51*(1), 144–164. https://doi.org/10.1044/2019_LSHSS-18-0129

Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Power-Point. (2018). *Power-point for Macintosh* [Computer software]. Microsoft.

Riley, R. D., Snell, K. I., Martin, G. P., Whittle, R., Archer, L., Sperrin, M., & Collins, G. S. (2021). Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *Journal of Clinical Epidemiology*, *132*, 88–96. https://doi.org/10.1016/j.jclinepi.2020.12.005

Roberts, S., & Nowak, G. (2014). Stabilizing the lasso against cross-validation variability. *Computational Statistics & Data Analysis*, *70*, 198–211. https://doi.org/10.1016/j.csda.2013.09.008

Roseberry, C. A., & Connell, P. J. (1991). The use of an invented language rule in the differentiation of normal and language-impaired Spanish-speaking children. *Journal of Speech & Hearing Research*, *34*(3), 596–603. https://doi.org/10.1044/jshr.3403.596

Scharff Rethfeldt, W., McNeilly, L., Laasonen, M., Meir, N., Abutbul-Oz, H., Smolander, S., Niega Gargia Goulart, B., & Frances Hunt, E. (2024). Assessment of developmental language disorder in multilingual children: Results from an international survey. *Folia phoniatrica et logopaedica*, *76*(2), 127–150. https://doi.org/10.1159/000533139

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Ramos Oliveira, P., & Skoruppa, K. (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*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Schwob, S., & Skoruppa, K. (2021). Identifying developmental language disorders in French-Portuguese bilingual children: Resources available to speech-language pathologists. *Canadian Journal of Speech-Language Pathology and Audiology*, *45*(4), 251–266.

Schwob, S., & Skoruppa, K. (2022). Detecting developmental language disorder in monolingual and bilingual children: Comparison of language-specific and crosslinguistic non-word repetition tasks in French and Portuguese. *Journal of Speech, Language, and Hearing Research*, *65*(3), 1159–1165. https://doi.org/10.1044/2021_JSLHR-21-00017

Sim-Sim, I. (2006). *Avaliação da linguagem oral - um contributo para o conhecimento do desenvo lvimento linguístico das crianças portuguesas* (4e ed.). Fundação Calouste Gulbenkian.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. J. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, *39*(5), 1–13. https://doi.org/10.18637/jss.v039.i05

Skoruppa, K. (2019). Novel noun and verb learning in mono- and multilingual children. *Travaux neuchâtelois de linguistique*, *1*(71), 109–123. https://doi.org/10.26034/tranel.2019.2994

Sysoev, I., Gray, J., Fine, S., Sneha, P. M., & Roy, D. (2022). Child-driven, machine-guided: Automatic scaffolding of constructionist-inspired early literacy play. *Computers & Education*, *182*, 104434. https://doi.org/10.1016/j.compedu.2022.104434

Thibault, M. P., & Helloin, M. C. (2010). *Exalang 5-8*. Happyneuron.

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E. (2015). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, *17*(2), 97–114. https://doi.org/10.3109/17549507.2014.923509

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research*, *54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, K. de Jong, & N. Meir (Eds.), *Assessing multilingual children* (pp. 301–331). Multilingual Matters.

Van Horn, D., & Kan, P. F. (2015). Fast mapping by bilingual children: Storybooks and cartoons. *Child Language Teaching and Therapy*, *32*(1), 1–13. https://doi.org/10.1177/0265659015584975

Volpin, L., Rezzonico, S., & de Weck, G. (2020). Enfants bilingues et prise en charge logopédique : panorama des pratiques déclarées. *Glossa*, *129*(1), 16–31.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Weismer, S. E., & Hesketh, L. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research*, *36*(5), 1013–1025. https://doi.org/10.1044/jshr.3605.1013

Yoshida, H., Tran, D., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontier in Psychology*, *210*(2), 1–14. https://doi.org/10.3389/fpsyg.2011.00210