

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Which Measures Better Discriminate Language Minority Bilingual Children With and Without Developmental Language Disorder? A Study Testing a Combined Protocol of First and Second Language Assessment

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Bonifacci P., Atti E., Casamenti M., Piani B., Porrelli M., Mari R. (2020). Which Measures Better Discriminate Language Minority Bilingual Children With and Without Developmental Language Disorder? A Study Testing a Combined Protocol of First and Second Language Assessment. JOURNAL OF SPEECH, LANGUAGE, AND HEARING RESEARCH, 63(6), 1898-1915 [10.1044/2020_JSLHR-19-00100].

Availability:

This version is available at: <https://hdl.handle.net/11585/764871> since: 2024-05-20

Published:

DOI: http://doi.org/10.1044/2020_JSLHR-19-00100

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

Running head: *Discriminating DLD in LMBC*

Title

Which measures better discriminate language minority bilingual children with and without Developmental Language Disorder? A study testing a combined protocol of L1 and L2 assessment.

Authors

Paola Bonifacci¹, Elena Atti², Martina Casamenti³, Barbara Piani⁴, Marina Porrelli⁵, Rita Mari⁶

¹Department of Psychology, University of Bologna, Bologna, Italy

²Villa Esperia – Rehabilitation Institute, Pavia, Italy

³Private Speech-Language Pathologist, Imola, Italy

⁴Nursing and technical direction (DIT), AUSL ROMAGNA, Faenza, Italy

⁵ Department of Primary Care, AUSL Bologna, Bologna, Italy

⁶ Studio Di Psicologia Clinica, Formazione, Linguaggio e Apprendimento Anna Valentini, Carpi, Italy

Corresponding author: Paola Bonifacci, Department of Psychology, University of Bologna, Viale Berti Pichat 5, 40126, Bologna. Phone: 0039 0512091824; E-mail: paola.bonifacci@unibo.it

Conflict of interest statement: there are no conflicts of interest

Funding statement: the authors did not receive funding for the study

Abstract

24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Purpose: The present study aimed to assess a protocol for the evaluation of Developmental Language Disorder (DLD) in language minority bilingual children (LMBC). The specific aims were: 1) to test group differences; 2) to evaluate the discriminant validity of single measures included in the protocol; 3) to define which model of combined variables had the best results in terms of efficacy and efficiency.

Method: Two groups of LMBC were involved, one with typical development (TD) (n=35), selected from mainstream schools and one with DLD (n=20). The study protocol included the collection of demographic information and linguistic history, a battery of standardized tests in L2 (Italian), including nonword repetition, morphosyntactic comprehension and production, and vocabulary and narrative skills, and included direct (children’s evaluation) and indirect (parents’ questionnaire) assessment of linguistic skills in L1.

Results: Results showed that the two groups differed in almost all linguistic measures. None of the single measures reached good specificity/sensitivity scores. A combined model, that included direct and indirect assessment of L1 skills, morphosyntactic comprehension and production, and nonword repetition, reached good discriminant validity, with 94.5% of cases correctly classified.

Discussion: The study defines a complex picture of the linguistic profile in bilingual children with DLD, compared to TD bilingual peers. The results reinforce the idea that no single measure can be considered optimal in distinguishing children with DLD from typical peers. The study offers a concrete example of an effective and efficient protocol with which to discriminate LMBC with and without DLD.

Keywords: Developmental Language Disorder, Language Minority Bilingual Children, Morphosyntactic skills, L1 assessment, Nonword repetition.

1 Which measures better discriminate language minority bilingual children with and without Developmental
2 Language Disorder? A study testing a combined protocol of L1 and L2 assessment.

3
4 Within Grosjean's (1989) pragmatic definition of bilinguals as "those people who use two or more
5 languages in their everyday life" (p. 4), the present study focused on a subgroup of bilingual children who
6 are exposed to a variety of minority languages in their home environments and to the societal language
7 within the school context, leading them to be considered as sequential bilinguals. However, in many cases,
8 they were born in the country of schooling from immigrant families (In Italy, 80% of non-Italian citizen
9 children attending Infant school were born in Italy, MIUR, 2018); thus, their exposure to the societal
10 language might be heterogeneous.

11 Despite the fact that bilingualism *per se* is not a risk factor for Developmental Language Disorders
12 (DLD), language minority bilingual children (LMBC) might encounter an increased chance of
13 over/underdiagnosis, or of misdiagnosis (Bedore & Peña, 2008; Grimm & Schulz, 2014; Lehti, Gyllenberg,
14 Suominen, & Sourander, 2018; Salameh, Nettelbladt, Håkansson, & Gullberg, 2002). The L2 language skills
15 of these children may vary immensely depending on several factors, such as the amount and quality of
16 bilingual exposure (Thordardottir, Rothenberg, Rivard, & Naves, 2006; Scheele, Leseman, & Mayo, 2010;
17 Sorenson-Duncan & Paradis, 2018), and LMBC might underperform in their L2 linguistic skills compared to
18 monolingual Typically Developing (TD) peers (Bedore & Peña, 2008). The L2 gap may persist for periods
19 of time that vary from child to child, and, at least for some children, might be quite long (Paradis, 2016),
20 with possible consequences on scholastic achievements (Graham, Minhas, Paxton, 2016, Bonifacci,
21 Lombardo, Pedrinazzi, Terracina, Palladino, 2019; Bellocchi, Bonifacci, Burani, 2014). The basic question is
22 whether and how it can be determined that language difficulties in a bilingual child are due to a *disorder* and
23 are not the reflex of a particular stage of typical L2 development. Different terminology has been used in the
24 literature to refer to disorders in language development: Primary Language Disorder/Impairment, Specific
25 Language Impairment (SLI), Language impairment (LI), and Developmental Language Disorder (DLD). The
26 latter term is employed here in accordance with more recent suggestions (Bishop, Snowling, Thompson,
27 Greenhalgh, & CATALISE-2 Consortium, 2017).

28 In previous literature, an increasing number of studies, that are briefly summarized in the following
29 sections, have tried to investigate which linguistic measures better allow for a correct identification of a DLD
30 in bilingual children; a subset of these studies specifically focused on LMBC. However, most studies tested
31 single measures or subsets of measures within the same linguistic area (e.g., morphological skills, nonword
32 repetition), whereas very few studies specifically addressed the issue of testing clinical protocols. These
33 studies mainly included children exposed to English as L2, with limited ascertained transferability of these
34 protocols to other languages of assessment. Furthermore, studies using the same methodology in different
35 bilingual contexts with children speaking a variety of minority languages are still lacking.

36 As suggested by Paradis, Schneider & Sorenson Duncan (2013), one strength of the bilingual
37 assessment approach is the emphasis on comparing bilingual children with each other, instead of with
38 monolinguals, for the purposes of identifying children with DLD. In the study by Paradis et al. (2013), the
39 authors compared, through a combined protocol of English measures and a parent questionnaire, 152
40 typically-developing bilingual children with 26 bilingual children with DLD. Children came from different
41 linguistic backgrounds and the protocol included English standardized tests of nonword repetition, tense
42 morphology, narrative story grammar, and receptive vocabulary. The ALDEQ questionnaire was
43 administered to parents to obtain measures for children's first-language development. Children with DLD
44 underperformed compared to the TD group in all measures, except vocabulary. Then, through Linear
45 Discriminant Function Analyses they tested two models. In the first, with all measures included, they found
46 91% of sensitivity and specificity indexes. Then, in Model 2, vocabulary was excluded and the model, which
47 resulted significant, revealed 92% of specificity and 91% of sensitivity. The strongest discriminator was the
48 ALDEQ, followed by nonword repetition and tense morphology; story grammar had a minor discriminant
49 power. Other models that were tested had minor specificity/sensitivity indexes and the authors report that
50 when the ALDEQ was removed discriminant scores fell below 80%.

51 The present study replicates and extends the structure of Paradis, et al.'s study (2013). In particular,
52 this study retains the measures tested in the model developed by Paradis et al. (2013) and included additional
53 measures: L1 linguistic skills, morphosyntactic comprehension, and microstructural aspects of narrative
54 skills. The aim was to identify which linguistic measures better discriminate bilingual language minority
55 children with and without DLD.

56 ***Methodological issues in the assessment of DLD in bilingual children***

57 When testing bilingual children, the use of L2 standardized measures alone might not provide
58 sufficient or reliable evidence in the absence of an accurate evaluation of L1 skills together with information
59 about linguistic history (e.g., age of exposure, languages spoken at home, etc.). Professional organizations
60 such as the American Speech and Hearing Association (ASHA), the Royal College of Speech and Language
61 Therapists (RCSLT), and many research groups (see reviews in De Lamo White & Jin, 2011; Ebert, &
62 Kohnert, 2016) suggest that the best solution for assessing bilingual children would be to assess each of their
63 languages. Testing the L1, however, encounters methodological shortcomings such as the poor availability of
64 minority L1 standardized measures and the difficulty faced by speech therapists when assessing a child in
65 another language (see also Boerma et al. 2017, Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013).
66 Assessing competencies in L1 should, however, be accompanied by an equivalent assessment of L2, which
67 may also account for possible language attrition processes in L1. In response to this issue, Contento,
68 Bellocchi & Bonifacci (2013) developed the Babil Test, which includes an assessment of linguistic
69 comprehension skills in Italian (L2 for sequential bilinguals) and a set of other languages (L1 for bilingual
70 children). This task allows for the definition of a bilingual profile, and it does not require L1 knowledge on
71 the part of the clinician.

72 The investigation of the development of linguistic competence may alternatively, or
73 complementarily, be achieved through parents' reports. Asking parents for specific information about the
74 child's linguistic development milestones is a widely-used procedure in clinical settings, both for
75 monolingual and bilingual children. Data from parental reports are reliable indexes of language impairment
76 and have been shown to correlate with objective measures (standardized tests) of language proficiency
77 (Bedore, Pena, Joyner, & Macken, 2011; Gutierrez-Clellen & Kreiter, 2003; Rescorla, 1993).

78 Paradis, Emmerzael and Sorenson Duncan (2010) developed the ALDEQ questionnaire, that was
79 specifically designed for the evaluation of clinical markers of DLD in sequential bilingual children, mainly
80 by collecting information from parents on their child's linguistic development in L1. A key characteristic of
81 the questionnaire is that it is non-language/culture-specific since it does not ask about the knowledge/mastery
82 of specific words or linguistic structures. Results from ALDeQ total scores showed robust between-group
83 differences between bilingual children with and without DLD. Paradis et al. (2010) found that the scores

84 from the questionnaire discriminated well overall, but with superior specificity to sensitivity, suggesting that
85 it could be a useful clinical instrument if it were used in conjunction with other measures. The
86 questionnaire's reliability in discriminating LMBC with and without DLD has been replicated in a sample of
87 bilingual children exposed to Italian as L2 (Bonifacci et al., 2016).

88 In summary, the challenge is to understand how clinical markers might be combined within clinical
89 protocols for the identification of DLD in bilingual populations. Dollaghan & Horner's (2011) meta-analysis
90 found that no single measure was optimal for discriminating DLD children from typical peers, and the
91 authors state that "*the results of any single measure must be viewed as no more than somewhat suggestive of*
92 *diagnostic status marker in LLI, given the heterogeneity of children with LLI as well as developing*
93 *bilinguals*" (p. 1086). Moreover, clinical evaluation in bilingual profiles should not rely exclusively on L2
94 assessment, and previous literature has highlighted the need for additional methodological issues that should
95 be implemented in LMBC linguistic assessment, including L1 assessment.

96

97 ***Markers of Developmental Language Disorder in bilingual populations.***

98 In this section, we will briefly review the areas of linguistic competence that previous studies found
99 to be potential markers of DLD in bilingual and monolingual assessment.

100 *Morphosyntactic and grammar skills.* Morphological difficulties are considered a core component of
101 DLD, although with somewhat different clinical markers across languages. For example, English children
102 with DLD have been proven to have severe difficulties in producing tense morphology and in judging
103 accuracy in morphology in English (Rice, & Wexler, 1996), whereas French and Italian children's
104 performance is significantly poorer than TD peers in producing object clitics (Paradis, 2010; Bortolini,
105 Caselli & Leonard 1997), and German children with DLD show striking difficulties in verbal agreement
106 (Hamann 2012). Children exposed to a second language usually make more morphological errors than
107 monolingual peers, and it can take time for them to achieve monolingual-like performances (Blom, Paradis,
108 & Sorenson Duncan, 2012; Chondrogianni & Marinis, 2011; Jia & Fuse, 2007; Paradis, Schneider, &
109 Sorenson Duncan, 2013; Paradis, Tulpar, & Arppe, 2016; see for Italian Bellocchi, Tobia & Bonifacci, 2017;
110 Bonifacci, Barbieri, Tomassini & Roch, 2018; Bonifacci, Tobia, Bernabini, Marzocchi, 2016). Nevertheless,
111 second language learners with DLD have remarkable and more severe difficulties with morphology

112 compared to their TD L2 peers. This suggests that morphology might be considered a sensitive clinical
113 marker in discriminating TD from DLD among L2 children, as is the case for monolinguals (Blom &
114 Paradis, 2013; Gutierrez-Clellen, Simon-Cerejido, & Wagner, 2008; Jacobson & Livert, 2010; Paradis et al.,
115 2013). At the sentence level, bilingual children with DLD have been found to produce more grammatical
116 errors, shorter utterances, and reduced sentence comprehension. Some authors suggest that language
117 minority bilingual children might show cumulative effects (Cumulative Effects Hypothesis, CEH; Orgassa &
118 Weerman, 2008), in that bilingualism might impose extra language learning difficulties on children with
119 DLD, also leading bilingual children with DLD to underperform in comparison to monolingual peers with
120 DLD. There are, however, contradictory results to this regard. For example, Rothweiler, Chilla, and Clahsen
121 (2012), (Turkish–German) found that monolingual and bilingual children with DLD did not differ in tense
122 marking and produced syntactically complex sentences such as embedded clauses and wh-questions, but
123 were limited in producing correct agreement-marked verb forms. Gutierrez-Clellen, et al. (2008) (Spanish-
124 English) did not find differences between bilinguals and monolinguals with DLD in subject or verb use.
125 Paradis, Crago, Genesee, and Rice (2003) (French-English) and Paradis, Jia, & Arppe, (2017) (Language
126 Minority exposed to English L2) found that the two groups had similar accuracy in the production of tense
127 morphology. In summary, morphological disorders may be considered a core feature of DLD both in
128 monolingual and bilingual populations and results from previous literature suggest including morphological
129 tasks in assessment protocols. However, there might be possible behavioral similarities between the language
130 profiles of bilingual children and children with DLD, and this should be considered within a broader
131 assessment perspective, in order to avoid missed and mistaken identities (Gutiérrez-Clellen, 1996).

132 *Vocabulary.* Children with Developmental Language Disorder have limited expressive vocabulary
133 (Gray, Plante, Vance, & Henrichsen, 1999; Leonard, 2014) and have difficulty learning new words (Alt &
134 Spaulding, 2011). Weakness in vocabulary size, when tested in only one language, is, however, a core
135 characteristic of bilingual language minority children, as documented by many studies (see Bialystok, Luk,
136 Peets, & Yang, 2010; Verhoeven, Steenge, van Weerdenburg, & van Balkom, 2011). Vocabulary knowledge
137 is distributed across two languages (Patterson, 2004; Sheng, Peña, Bedore, & Fiestas, 2012) and bilingual
138 learning is context dependent (Oller & Pearson, 2002). The “vocabulary gap” in bilinguals should disappear
139 when both languages or “conceptual knowledge” are considered (Hoff et al., 2012). Bilinguals’

140 underperformance in vocabulary tasks, therefore, should not be considered as a marker of a *disorder* in the
141 first instance. Furthermore, the assessment through standardized tests, which are strongly “knowledge-
142 dependent” tasks (Campbell, Dollaghan, Needleman, & Janosky, 1997), might negatively affect the
143 evaluation of bilinguals’ competence, due to less experience with tasks and possible cultural biases (De
144 Lamo White & Jin, 2011). However, there is evidence that bilingual children with DLD underperform
145 compared to typically-developing bilingual peers (Sheng et al. 2012) and, if tested in both languages (Bedore
146 & Pena, 2008), children with DLD are also expected to underperform in terms of L1 vocabulary compared to
147 their TD bilingual peers (Gibson, Pena & Bedore, 2014).

148 *Narrative skills.* Narrative tasks reflect linguistic and communicative competence (Botting, 2002)
149 and are considered valuable tools for the assessment of linguistic abilities of children with language
150 disorders. Narratives are usually coded considering either microstructural (lexicon, morpho-syntactic skills)
151 or macrostructural (story grammar, causal relationships) features (LITMUS- Multilingual Assessment
152 Instrument for Narratives, Gagarina, et al., 2012). Bilingual children with typical development, when tested
153 in L2, usually underperform, compared to monolinguals, at the microstructural level, whereas their
154 performance is average in terms of the macro-structural level (Hipfner-Boucher et al., 2014; Fey, Catts,
155 Proctor-Williams, Tomblin, & Zhang, 2004; Bonifacci et al., 2018; Tsimpli, Peristeri, & Andreou, 2016)
156 possibly because macrostructure, more than microstructure, is thought to benefit from cross-linguistic
157 transfer from L1. On the contrary, bilinguals with DLD often underperform compared to TD bilinguals in
158 macrostructural and microstructural aspects of narratives (Squires et al., 2014; Fichman et al. 2017), because
159 of a language disorder that affects both L1 and L2. In summary, the assessment of narrative skills,
160 particularly at the macrostructural level, can be considered a valuable tool for the identification of DLD
161 (Botting, 2002) also in language minority bilingual children (Boerma, Leseman, Timmermeister, Wijnen, &
162 Blom, 2016; Paradis et al., 2013).

163 *Nonword repetition (NWR).* NWR involves temporary storage and retrieval of novel strings and is
164 considered a measure of phonological memory (see Ebert et al., 2014 for discussion). NWR is thought to
165 mimic word learning (Gathercole, 2006), and significant relationships between NWR task performance and
166 vocabulary acquisition have been documented in the literature (e.g., Gathercole & Baddeley, 1989), also
167 suggesting an involvement of long-term representations in NWR (Gathercole, 1995, McDonald & Oetting,

168 2019). NWR has been found to be severely impaired in monolingual children with DLD (e.g., De Bree,
169 Rispens, & Gerrits, 2007; Dispaldro, Leonard, & Deevy, 2013; Gathercole & Baddeley, 1990) and many
170 studies expanded the strength of this task also to bilingual populations, suggesting NWR as one of the best
171 candidates for discriminating bilingual children with and without DLD (Boerma et al., 2015). One of the
172 main advantages of NWR is that it is considered one of the purest behavioral tasks, being less dependent on
173 language knowledge and tapping into more basic cognitive underpinnings of language (Thordardottir &
174 Brandeker 2013; Paradis, et al., 2013; Ebert, 2014; Gathercole, 2006). Results on clinical populations
175 robustly indicate that bilingual children with DLD underperform compared to typically developing bilingual
176 peers in NWR tasks in both the L1 and the L2, as shown in English-French bilinguals in Canada
177 (Thordardottir & Brandeker 2013), Turkish-Dutch bilinguals in the Netherlands (Verhoeven et al. 2012), and
178 Spanish-English bilinguals in the United States (Gutierrez-Clellen & Simon-Cereijido, 2010). Recently,
179 further evidence has been collected through NWR tasks specifically developed with the aim of reducing
180 lexical and language-dependent influences (LITMUS project, see de Almeida et al. 2017, Armon-Lotem &
181 Meir, 2016; Boerma et al. 2015, Chiat and Polišenská, 2016); results from these studies have shown fair to
182 excellent diagnostic accuracy. Nonetheless, some authors have suggested that this task seems to have better
183 specificity than sensitivity (Gutiérrez-Clellen and Simon-Cereijido, 2010). Strong performance may
184 effectively rule out DLD for typical bilingual children (Gutiérrez-Clellen & Simon-Cereijido 2010; Windsor
185 et al. 2010). On the other hand, poor performance in a NWR task might not be sufficient, as a single
186 measure, to identify a DLD, and further assessments would be needed to provide an unequivocal positive
187 identification of DLD in bilingual children (Kohnert, Windsor, and Yim, 2006; Engel de Abreu, 2011).

188

189 ***Research Questions and Hypotheses***

190 Here, we present a study in which a clinical protocol for the valuation of DLD was administered to
191 LMBC, with and without DLD, exposed to Italian as L2. The aim of the study was to test a protocol of
192 combined L1/L2 measures for discriminating DLD in bilingual language minority children from different
193 linguistic backgrounds. This study replicates and extends the structure of the previous research by Paradis, et
194 al. (2013). First, the study intends to replicate findings from Paradis et al. in a different cultural and linguistic
195 context, that is, bilingual children exposed to Italian as L2. This should allow us to increase the

196 generalizability of the results of clinical protocols in different linguistic contexts. In addition, this study
197 extends previous findings by adding measures that were not included in Paradis et al.'s study but that the
198 literature suggests as potential markers for DLD in bilingual children: L2 morphosyntactic comprehension,
199 L1 vocabulary and morphosyntactic comprehension skills, and the fact of being conducted on a sample of
200 children exposed to Italian as L2. Third, the study adds the analysis of different discrimination models in
201 order to test for models with possible increased sensitivity/specificity indexes compared to Paradis et al.
202 (2013) and to define the combination of measures with highest indexes of efficacy (sensitivity/sensibility
203 scores) and efficiency (length of protocol).

204 The protocol included L2 standardized measures (morphosyntactic/grammar skills, vocabulary),
205 nonword repetition, and narrative skills. Furthermore, there were measures of L1 linguistic skills (receptive
206 vocabulary and language comprehension) and parents' reports on the linguistic history and clinical markers
207 of DLD in L1.

208 More specifically, the aims of this study were:

- 209 1) To assess differences between bilingual children with and without DLD in the measures included in the
210 clinical protocol. We expected children with DLD to underperform compared to their TD monolingual
211 peers in all measures, except for intellectual functioning.
- 212 2) To evaluate the discriminant power of single measures (and their subscales, when available) in
213 differentiating children with and without DLD. Based on previous literature, we expected parents' reports,
214 L2 nonword repetition and L2 grammar/morphosyntactic skills to better differentiate compared to other
215 standardized L2 measures such as vocabulary.
- 216 3) To evaluate which combination of linguistic and parents' measures best discriminate between children
217 with and without DLD. To accomplish this aim, we selected the best single measures and combined them
218 in discriminant analysis models. We expect language processing measures, L1 competence, and parents'
219 reports to constitute the most robust model in terms of efficacy (sensibility and sensitivity scores) and
220 efficiency (protocol length).

221

222 **METHOD**

223 *Study design and participants*

224 The participants were a total of 55 bilingual children (mean age: 83.34 Months, SD: 6.46; 28
225 Females) exposed to Italian as L2 and to a minority language (for the Italian context) in their home
226 environment. The languages spoken at home were: Moroccan-Arabic (50.9%), Albanian (16.4%), Romanian
227 (9.1%), Urdu (9.1%), Tunisian-Arabic (7.3%), Polish (3.1%), Bengali (1.8%), and Chinese (1.8%). Inclusion
228 criteria for all participants were: at least 2 years of intensive exposure to Italian within the school context
229 (mainly preschool); both parents speaking a language different from Italian at home; intellectual functioning,
230 as measured through Raven's Matrices, within the normal range (> 20^o centile); absence of neurological
231 impairment and sensory deficits.

232 The sample included two different groups of LMBC. The first group comprised 35 bilingual children
233 with typical development (TD) (mean age: 82.48 Months, SD: 6.34; 21 Females). They were selected from
234 primary schools in a region of northern Italy. These children had not been diagnosed as having any
235 Neurodevelopmental Disorder, nor did they have any neurological or sensorial loss. The second group was
236 made up of children with Developmental Language Disorders (DLD) (mean age: 84.85 Months, SD: 6.55; 7
237 Females). These children had received, within the Italian National Health System, a clinical diagnosis of
238 Specific developmental disorders of speech and language (F 80), according to ICD-10 (WHO, 1992) criteria,
239 within the past 12 months. The diagnostic protocol included an in-depth interview with parents on the
240 children's medical and linguistic history, including questions regarding language of exposure, age of
241 exposure, and language delays in L1. In both clinical centres all children underwent an ENT assessment for
242 the exclusion of auditory deficits, and speech-language pathologists excluded primary speech-sound
243 disorders. Other exclusionary criteria were: sensory/neurological deficits, emotional disturbances, attention
244 deficit hyperactivity disorder (ADHD), and other neurodevelopmental disorders. All children underwent a
245 cognitive and speech-language assessment, conducted by a multidisciplinary team of speech therapists,
246 psychologists, and neuropsychiatrists. Inclusionary criteria for the diagnosis were defined according to the
247 ICD-10 classification manual for Specific developmental disorders of speech and language (F 80) and in
248 both centres the diagnosis followed a clinical evaluation, rather than being based on cut-off scores of
249 standardized tasks. Common inclusionary criteria were IQ in the normal range, significant impairments in
250 receptive/expressive language measures (vocabulary, morphosyntactic skills), at least 2 years of intensive
251 exposure to Italian within the school context (mainly preschool). The tests used in the assessment protocols

252 partially differed according to the team's experience and instruments available (e.g., different tests for the
253 assessment of Intellectual Quotient - IQ). The clinical evaluation was independent from the protocol used in
254 the present study, in which the tests and questionnaire used were selected based on a theoretical framework
255 (Paradis et al., 2013) and administered by two of the authors (speech-language pathologists), which were not
256 part of the clinical team that conducted the diagnosis. Further information concerning the languages spoken
257 at home is reported in Table 1.

258 Insert Table 1 here

259

260 Information on group characteristics (SES, age of exposure) and statistics on group differences in
261 background variables are detailed in the results section and in Table 2.

262 The TD group was selected to match the DLD group for chronological age at the moment of
263 assessment for the present study. It has to be underlined that the children's age at the moment of the
264 diagnosis was slightly lower, since they had been administered the study protocol after their diagnostic
265 assessment was concluded. The age of the diagnosis was therefore between five and six, which is actually an
266 adequate time frame in the case of bilingual children, as at least two years of scholastic exposure are needed
267 for a diagnosis of DLD.

268 ***Measurements***

269 All children were administered the following tasks. Parents completed the questionnaire on linguistic
270 history and the ALDEQ-IT interview (see below for a description).

271 **Children's cognitive assessment**

272 *Intellectual functioning.* Raven Coloured Progressive Matrices (Belacchi, Scalisi, Cannoni, & Cornoldi,
273 2008).

274 **Children's assessment in Italian (L2)**

275 *Morphosyntactic and grammar comprehension.* Children were administered the TROG-2 test
276 (Bishop, 1989, Italian adaptation Suraniti, Ferri, & Neri, 2009), a standardized measure of receptive
277 grammar that examines 20 specific syntactic constructions. Each construction is tested with a block of four
278 items. The participant's task is to select the one drawing out of four choices that corresponds to a sentence
279 read aloud by the examiner. Foil drawings differ from the target drawing by either a lexical element (noun,

280 verb, adjective) or a grammatical element (word order, function word, inflection). Testing was discontinued
281 after five consecutive failed syntactic constructions (i.e., blocks). Failure is defined as one or more incorrect
282 responses in a block. Performance on the TROG–2 is quantified in terms of the number of blocks passed and
283 raw scores were converted into standard scores according to Italian norms. The Italian version of the TROG
284 retains the same morphosyntactic structures as those used in the English version. For most morphosyntactic
285 structures tested by TROG, Italian and English sentence structure is very similar (e.g., *la* (article) *pecora*
286 (noun) *sta correndo* (verb, gerund); the sheep is running). In a negative sentence, in Italian, the verb follows
287 the negation (L'uomo non (neg.) è (verb) *seduto*; The man is not sitting). Some differences are in
288 prepositions: in Italian there are compound or simple prepositions when for English two words or
289 prepositions are needed (e.g., *La tazza è nella scatola*, The cup is *in the* box, or, *L'anatra è più grande della*
290 *palla*, The duck is bigger *than the* ball; *La mucca è inseguita dalla ragazza*, The cow is chased *by the* girl).
291 The task includes the assessment of specific grammatical structures that are known for being potential
292 markers of DLD in Italian (Bortolini, Caselli & Leonard 1997), for example clitics (e.g., *L'uomo vede che il*
293 *ragazzo lo* sta indicando; The man sees that the boy is pointing *at him*), singular/plurals (e.g., *Il ragazzo*
294 *raccoglie i fiori*; the boy collects flowers), pronouns (e.g., *Loro* lo stanno portando, *They* are bringing it),
295 propositions relating to the subject, which include verb agreement (e.g., *L'uomo, che sta mangiando, guarda*
296 *il gatto*; The man who is eating *looks* at the cat).

297 *Morphosyntactic and grammar production.* The morphosyntactic production subtest of the Test
298 Neuropsicologico (TNP) [Neuropsychological test] (Cossu & Paris, 2007) was administered. Within a
299 pragmatic context described by the examiner, the child must describe the action taken by the examiner and
300 the task consists in the elicitation of 6 propositions, two for each syntactic structure: relatives, datives, and
301 negatives. For relative, taking as an example the first clause, the examiner places two bowls on the table in
302 front of two dogs and the character of a child in front of the dogs. Then, the examiner explains that there are
303 two dogs, one dog is eating (the examiner indicates a dog) and the other is not eating (the second dog is
304 indicated); the examiner continues by saying that a child arrives and touches a dog (the examiner takes the
305 child and makes him touch the dog that is eating). Afterwards, the child is asked which dog is touched by the
306 child; the expected answer is the relative "Il bambino tocca il cane che mangia" (the child touches the dog
307 that eats), which assesses the capacity to produce a relative clause with explicit reference to the object

308 complement. For datives, the examiner shows a child who is sick in bed and the mother performing the
309 action of bringing a bowl of soup to the child. The child is asked the direct question: "what is the mother
310 doing?", The expected answer will be the simple dative phrase "La madre porta la zuppa al bambino" (The
311 mother brings the soup to the child). The negatives are not elicited with a direct question as for the two
312 previous syntactic classes but through a completion of a direct statement. For example, in one of the
313 sentences, the mother puts cherries on the table as a snack for John. Then, the child arrives at the table but
314 goes off to play, leaving the cherries on the table. In the following scene the mother comes in again and
315 wonders whether John has eaten the cherries. The answer is prompted by saying to the child: "you tell me:
316 Giovanni ..." the expected answer will be the simple negative phrase "Giovanni non mangia le ciliege"
317 (Giovanni has not eaten the cherries). A score of 1 is given for each correct answer, with scores ranging from
318 0 to 6. Raw scores are converted into z-scores. The test manual reports test-retest reliability with $R = .76$.

319 *Vocabulary.* The Italian version of the Peabody Picture Vocabulary Test (PPVT) (Dunn, & Dunn,
320 1981; Stella, Pizzoli, & Tressoldi, 2000) was administered. In this test, the examiner says a word, and the
321 examinee must choose the picture that best corresponds to the word from a selection of four presented
322 pictures. There is a total of 175 stimuli; standard scores are reported. The reliability of the PPVT-R reported
323 in the test manual is Chronbach's Alpha = 0.88.

324 *Nonword repetition.* Participants performed a nonword repetition task included in the Batteria per la
325 valutazione neuropsicologica 5–11 (test for neuropsychological assessment for 5- to 11-year-old children;
326 Bisiacchi et al., 2005). In this task, participants are instructed to listen to the 15 meaningless words spoken
327 by the examiner and to repeat it exactly as they hear it, without modifying it in any way. There were 5 bi-
328 syllabic nonwords, all with a CVCCV structure (e.g., cosco), and 10 tri-syllabic nonwords. Of the three-
329 syllabic nonwords, five had a CVCCVCV structure (e.g., torgame), two a CCVCVCV structure (e.g.,
330 glotoba), two had a CVCCVCCV structure (e.g., fuscorgo) and one a CCVCVCCV structure (e.g., frinosto).
331 Most of the Italian phonetic repertoire is tested, including occlusives (bilabial / p /, / b /; alveolar / t /, / d /;
332 velars / k /, / g /), fricatives (labiodentali / f /, / v /, alveolar / s /, / z /), nasal (bilabial / m /; alveolar / n /; palatal
333 /ɲ /); vibrants (/ r /); alveolar laterals (lateral: / l /) and vowels: (/ a /, / ε /, / e /, / i /, / ɔ /, / o /, / u /). Compared to the
334 LITMUS-NWR (Chiat, 2015; Dos Santos, & Ferré, 2018), the maximum length of nonwords is equal (three),
335 but minimum length is different (the BVN starts from bi-syllabic nonwords, whereas the LITMUS also

336 includes monosyllabic nonwords). In summary, the BVN task has a similar syllable structure and
337 syntagmatic axis compared to LITMUS-NWR but might differ in having higher segmental complexity; we do
338 not have data regarding between language-dependent and language independent sounds. Compared to other
339 nonword tasks used in English-speaking children, such as the CTOPP (Wagner, Torgesen, Rashotte, &
340 Pearson, 1999), used in Paradis et al.'s study, the BVN task is similar in length (BVN: 15 items, CTOPP: 18
341 items; in both cases of increasing difficulty) but different in administration procedure: in the CTOPP
342 nonsense words are reproduced on a CD and children's responses are recorded and scored later while in the
343 BVN the examiner says the nonwords and the child is asked to repeat it. The nonwords are repeated one by
344 one. The task is composed of 15 items. The examiner records the number of correct responses (accuracy) for
345 each child. Scores range from 0 to 15 and z-scores calculated based on norms reported in the test manual are
346 reported.

347 *Narrative skills.* An adapted version of the Nest Story (Paradis, 1987), included in the Batteria
348 Valutazione Linguaggio (BVL, Marini, Marotta, Bulgheroni, & Fabbro, 2015) was administered. A set of 6
349 pictures is presented to the child, and the examiner asks the child to tell the story. For the present study, four
350 indexes were considered: words per minute, mean length utterance (MLE), type (number of different words
351 produced in the narrative), and macrostructure. For the latter parameter, a set of 15 main actions were
352 identified (Marini, personal communication) and a score of 1 was given for each element correctly reported.
353 Raw scores were converted into z-scores based on the test's norms and TD sample mean scores for the
354 macrostructure index.

355 **Assessment of children's L1 linguistic competence**

356 *Linguistic competence in L1.* In order to test linguistic competences in L1 each child was
357 administered the Prove BaBIL (BaBIL Test; Contento, Bellocchi, Bonifacci, 2013). The BaBIL provides
358 information on bilingual profiles via four receptive tasks given in both L1 and L2. For the present study, only
359 the version in L1 was administered. The test is presented on a PC through a PowerPoint presentation, and
360 through a pair of earphones the child listens to instructions and stimuli in his/her L1, recorded in audio files
361 implemented in the power point presentation of the tasks. The examiner is sitting near the child and manages
362 the administration of the tasks on the computer. A scoresheet is available for the examiner with the correct
363 answers for each numbered item presented, and the examiner just has to mark whether the answer, given by

364 the child pointing at the screen, corresponds to the correct option. For example, in slide 1 of the vocabulary
365 task, the child listens to the word “feather” (in the L1) and sees four images (a feather, a puma, a duvet, a
366 bird) on the computer screen. Then, the child points to the correct answer (e.g., image number one) and the
367 examiner, on the score sheet, notes down the answer given by the child and checks if it is correct (for item 1,
368 response 1 is correct). Therefore, the examiner does not need to understand the different languages because
369 he/she knows which is the correct expected answer for each item, and the instructions and stimuli (in the L1
370 version) are recorded. The task was developed in Italian and in many different minority (for the Italian
371 context) languages (e.g., Arabic, Albanian, Twi, Tagalog, Romanian, Bengali, Chinese, and others). The
372 adaptation into different languages was conducted through the involvement of native speakers who not only
373 translated the stimuli but gave their contribution regarding the cultural and linguistic adjustments needed.
374 For example, in the Arabic (Moroccan) version, the item with the word “basket” (سلة) was replaced with
375 “bucket” (دلو), because it has been suggested that basket in Arabic was longer and less frequent compared to
376 the second. There was, however, a limited number of linguistic adjustments (0 to 3 changes for each
377 language on the entire task).

378 In all tasks, a score of 1 is given for each correct answer (see below maximum scores for each task),
379 and the total number of correct responses is transformed into z-scores, based on standardized mean and
380 standard deviation (SD) values contained in the test manual. The standardization sample was of bilingual
381 children (exposed to Italian as L2), tested in their L1. The test was not developed for a diagnostic purpose;
382 psycholinguistic variables (word frequency and length, morphosyntactic complexity) in each language could
383 not be precisely controlled for, and there is no previous evidence concerning the discriminant validity for the
384 identification of DLD in bilingual children. The test is intended to collect direct information on L1
385 knowledge (receptive vocabulary, morphosyntactic comprehension, receptive grammar) and to define a
386 bilingual profile of language competence. The test usually requires the administration of both Italian and L1
387 versions (within a 15-day time interval), and the examiner can draw a profile of linguistic competence in the
388 two languages (e.g., if the child has a good knowledge of L1 but low scores in Italian, this is interpreted as
389 an insufficient exposure to L2; if the child has poor linguistic comprehension in both languages, this suggests
390 the need of further investigation for DLD, the case of a dominance in Italian suggests attrition in L1
391 competence). This is useful for orienting assessment and developing educational programs, etc. Cronbach’s

392 alpha for the whole test is .86. For the purpose of the present study, in line with evidence that suggests the
393 importance of assessing L1 (De Lamo White & Jin, 2011; Ebert, & Kohnert, 2016), we decided to include
394 the L1 version of the Babil in order to have a direct assessment of L1 receptive skills. We did not administer
395 the L2 (Italian) version, because for Italian we chose a more comprehensive linguistic assessment through
396 the most frequently and widely-used instruments.

397 The four subtests included in the battery are:

- 398 1) Vocabulary: The task includes 20 words. The choice of words for Test 1 was based on an age of
399 acquisition for Italian of below 4.5 years (Burani, Barca and Arduino, 2001). The score range is 0 - 20.
- 400 2) Morphosyntactic comprehension: The task includes 20 sentences containing locatives (e.g. the dog is
401 under the table), quantifiers (e.g. There are fewer flowers in the vase), negatives (There are no apples in the
402 basket), plurals (there are two bees). The task includes the assessment of diverse grammar structures:
403 articles, pronouns, prepositions, adverbs, adjectives, and verb agreement. The score range is 0 - 20.
- 404 3) Knowledge of body parts and colors (Basic Vocabulary). This task addresses basic linguistic
405 knowledge (body parts, colours), which is in everyday use in the scholastic context. There are 15 sentences
406 and the child is required to draw what the instruction says (e.g., "Colora i capelli di giallo" [Color the hair
407 yellow]). The score range is 0 - 15.
- 408 4) Inferences. The test consists of 15 items with increasing difficulty, which evaluates the
409 understanding of simple sentences (items from 1 to 5; for example: "The child is drawing"), complex
410 sentences (item from 6 to 10; "On Sunday morning Mrs. Maria goes to the park by bicycle. What does Mrs
411 Maris do on Sunday morning(s)?") and pragmatic judgments, (items 11 to 15; "The child has just woken up
412 and does not see his mom. What does he do?"). The score range is 0 - 15.

413 *Parents' questionnaires:* Parents were interviewed in order to assess the linguistic background of the
414 children and to collect demographic variables (QuBIL questionnaire, Contento et al., 2013) and socio-
415 economic status (SES) (Hollingshead Four Factor Indexes, 2011). For this study, Chronological Age, Age of
416 Exposure (AoE), and Months of Exposure (MoE) are included in the analyses. AoE refers to the age at which
417 the child began exposure to Italian as a second language within the scholastic setting. MoE corresponds to
418 the number of months the child has been consistently exposed to Italian in his/her everyday life within a
419 scholastic context. Other background variables (languages spoken at home, place of birth, etc.) were used as

420 criteria for inclusion/exclusion but were not included in the analyses. For SES, indexes of educational level
421 (EL) and occupation (O) were adopted. For the level of education, a score from 1 to 9 was indicated and for
422 employment, a score from 1 to 9. SES scores for fathers and mothers were determined with the formula
423 $EL*3 + O*5$, and an aggregate SES score for children resulted from the mean of the two values.

424 They were then administered the Italian version of the ALDEQ Questionnaire (for a full description
425 see Paradis et al., 2010; Bonifacci et al., 2016) for the evaluation of markers of DLD in their linguistic
426 development in L1. The ALDeQ is a questionnaire for parents structured in four sections: A) early language
427 milestones, B) current first language abilities, C) activity preferences and behavior, D) family history.
428 Answers are scored on rating scales such that lower scores index an increased risk for DLD, and higher
429 scores are more consistent with typical development. The rating scale scores yield a total proportion score
430 (denominator derived from the number of questions answered) with a range of 0 – 1.0. Raw scores were
431 converted into z-scores based on the Italian validation study.

432 ***Administration setting and procedure***

433 Written informed consent was obtained from all parents. The study was conducted in accordance
434 with ethical principles of the Declaration of Helsinki and approved by the Ethical Review Board of the
435 CEIIAV (Comitato Etico Irst Irccs AVR, Regional Health Service Emilia-Romagna, prot. 4239/2017
436 I.5/129).

437 ***Data analysis***

438 First, a set of t-tests was conducted in order to test group differences in background variables and in
439 cognitive and linguistic tasks. Then, a set of discriminant analyses was performed, first on single measures
440 and later on a set of models combining different indexes. Sensitivity and sensibility indexes are reported
441 together with Lambda Wilks value and statistical significance from univariate analysis for each model. As in
442 Paradis et al. (2013), this study adopted Plante and Vance's (1994) criteria for assessing classification
443 results, namely that specificity/sensitivity of 80% – 89% can be considered fair, and specificity/sensitivity
444 of > 90% can be considered good.

445

446 **RESULTS**

447 ***Group comparisons***

448 Table 2 reports mean values and group differences for samples' background variables.

449 Insert Table 2 here

450

451 The two groups did not differ for gender ($\chi^2(1) = 3.18, p = .07$), chronological age ($t(53) = -1.31, p =$
452 $.19, d = -0.37$), Age of Exposure ($t(53) = 1.01, p = .32, d = 0.31$), Months of Exposure ($t(53) = -1.63, p =$
453 $.12, d = -0.5$), SES ($t(53) = -0.74, p = .46, d = -0.21$), or languages spoken at home ($\chi^2(6) = 11.39, p = .07$).

454 Although non-significantly, the DLD group tended to have a higher number of males and a greater length of
455 exposure compared to the TD group. The majority of children (96.4%) were from low-SES families, in the
456 absence of difference of SES ranges between the two groups ($\chi^2(2) = 2.52, p = .28$).

457 Insert Table 3 here

458

459 In Table 3, group differences are reported for all variables included in the study protocol.

460 As expected, the two groups did not differ in intellectual functioning ($t(53) = .20, p = .84, d = 0.06$).

461 Bilingual children with DLD underperformed compared to their bilingual TD peers in all L2 measures:

462 vocabulary ($t(53) = 3.26, p < .01, d = 0.95$), morphosyntactic comprehension ($t(53) = 6.16, p < .01, d =$

463 2.01) and production ($t(53) = 4.4, < .01, d = 1.20$), nonword repetition ($t(53) = 4.10, p < .01, d = 1.10$), and

464 microstructural level of narratives (WpM: $t(53) = 3.04, p < .01, d = 0.86$; MLU $t(53) = 3.79, p < .01, d =$

465 1.11 ; Type: $t(53) = 2.89, p < .01, d = 0.79$). There was a tendency to significance in the macrostructural

466 aspects of narratives ($p = .056, d = 0.56$). There were significant differences also in L1 measures; all of the

467 four measures of the Babil tasks showed better L1 comprehension skills in the TD group compared to the

468 DLD group (Vocabulary: $t(53) = 5.15, p < .01, d = 1.57$; Morphosyntactic skills: $t(53) = 3.32, p < .01, d =$

469 0.89 ; Basic Vocabulary: $t(53) = 3.73, p < .01, d = 1.07$; Oral comprehension: $t(53) = 3.30, p < .01, d =$

470 0.89). Finally, the DLD group resulted as having higher indices of language difficulties as reported by

471 parents in the ALDEQ questionnaire: Section A ($t(53) = 5.16, p < .01, d = 1.39$); Section B: ($t(53) = 7.13, p$

472 $< .01, d = 2.11$); Section C: ($t(53) = 4.30, p < .01, d = 1.33$). The only subscale that did not yield a

473 significant difference was Section D, which referred to family risk for DLD ($t(53) = 1.2, p = .24, d = 0.33$).

474 ***Discriminant power of single measures***

475 Table 4 reports the output of discriminant analyses for all measures included in the study and, when
476 available, their subscales. The sample size is sufficient, based on the recommendation by Hair et al. (2006)
477 that each group should have at least 20 observations, and that at the very least, the smallest group size must
478 exceed the number of predictor variables.

479 *Insert Table 4 here*

480

481 The analyses showed that background variables (Raven matrices, SES, AoE and MoE) were not
482 statistically significant discriminators (all $p > .1$), except for a tendency for MoE ($p = .07$).

483 Subsequently, considering standardized measures in L2, morphosyntactic comprehension emerged as
484 the only measure that, alone, reached acceptable indexes of correctly classified cases (80%), although with
485 low specificity (77.10%). On the contrary, L2 nonword repetition had good specificity (91.4%) but low
486 sensitivity (50%), with a total percentage of 76.4% cases being correctly classified. A similar trend was
487 found for L2 macrostructural aspects of narrative skills (specificity: 94.3%) but with very low sensitivity
488 scores (20%) and poor predictive value (67.3%). Considering L2 narrative skills, the mean total score was
489 the index that furnished the highest percentage of correctly classified cases (72.7%). L2 Vocabulary showed
490 poor specificity (77.10%) and sensitivity (50%), with a rate of below 70% of cases being correctly classified,
491 whereas L2 morphosyntactic/grammar production had fair specificity (85.7%) but low sensitivity (50%), and
492 72.7% of cases were correctly classified.

493 As far as measures in L1 are concerned, the vocabulary subscale reached the highest number of
494 correctly classified cases (83.6%) whereas the other subscales and the total score had fair to good specificity
495 indexes, but poor sensitivity (below 60%).

496 Finally, the ALDEQ-IT parents' questionnaire reached fair to good discriminant validity when
497 considering the total score (specificity 88.6%; sensitivity 90%; percentage of correctly classified cases:
498 89.1%). Considering the different sections, section B (current first-language abilities) offered a fair score of
499 correctly classified cases (81.8%) with similar indexes of specificity (82.9%) and sensitivity (80%). The
500 other sections (A, C, D) had fair to good specificity but low sensitivity (below 60%) when considered
501 independently.

502 ***Discriminant analyses on combined model***

531 time taken to interview parents or find appropriate L1 objective measures. We therefore developed Model 6
532 to test whether a combination of L2 standardized measures could offer a valuable protocol in the absence of
533 L1 measures. Results showed that the discriminant values were weaker compared to other models, with a
534 total of 78.2% cases being correctly classified, and specificity and sensitivity indexes of 77.1% and 80%,
535 respectively.

536 The present study was developed from that of Paradis et al. (2013), with the same (or a language
537 equivalent) set of measures as in the original study. However, in the present study, we also included
538 additional measures (L2 morphosyntactic comprehension, L1 linguistic assessment, the microstructure of L2
539 narrative skills) and we developed our models based on a data-driven approach using the best measures that
540 emerged from our study. In order to increase the replicability of results we developed an additional model
541 (Model 7) in which we tried to reproduce, although with different standardized measures, the optimal model
542 that emerged from the study of Paradis and colleagues. Thus, in this model, we included the ALDEQ-IT
543 Total score, L2 nonword repetition, L2 story grammar, and L2 morphosyntactic/grammar production. The
544 model resulted as fair and, in comparison with what was found by Paradis et al. (2013), it showed equal
545 values of specificity (91%) but with slightly lower values of sensitivity (85% vs. 91%). This model was,
546 however, less strong compared to our optimal model (Model 3A).

547

548 **DISCUSSION**

549 The present study was aimed at testing the discriminant validity of a protocol for the evaluation of
550 Developmental Language Disorder in bilingual language minority children who were exposed to Italian as
551 L2 and who spoke different minority languages in their home environment. The study was devised with the
552 intention of replicating and extending a previous study by Paradis et al. (2013), and, importantly,
553 generalizing previous results on combined models of assessment in BLMC with DLD in a different linguistic
554 and cultural context, with new combinations of measures being tested in order to possibly reach higher
555 sensitivity/specificity indexes. Finally, in the models that were tested we addressed the issue of efficacy and
556 efficiency, defining those models that, with the minor number of measures gave the best discriminatory
557 indexes. Paradis et al. (2013) developed a protocol that included a parents' questionnaire on L1
558 development, and tasks in L2 (English): tense morphology, vocabulary, story grammar, and nonword

559 repetition. In the present study we kept similar measures as in Paradis et al.'s study (morphosyntactic
560 production, vocabulary, story grammar, and nonword repetition) and added the measures of children's L1
561 comprehension, L2 morphosyntactic comprehension, and microstructural aspects of narratives (in L2). The
562 study protocol also included the collection of demographic information and linguistic history (SES, AoE,
563 MoE, Age). Two groups of LMBC were involved: one with typical development (TD), selected from
564 mainstream schools and one with DLD, diagnosed within the Italian National Health System by a
565 multidisciplinary team of experts.

566 The first specific aim of the study was to evaluate the differences between the two groups of LMBC
567 regarding the measures included in the study protocol. The two groups did not differ in background
568 demographic or linguistic history variables, but LMBC with DLD underperformed compared to bilingual TD
569 peers for all measures excluding intellectual functioning, with only a tendency toward statistical significance
570 for the macrostructural aspects of narratives. This pattern of results is in line with previous studies, reviewed
571 in the introduction section, that evidenced how measures of L2 morphological comprehension
572 (Chondrogianni, Marinis, Edwards, & Blom, 2015), L2 morphological production (Blom, & Paradis, 2013),
573 L2 nonword repetition (Boerma et al., 2015), and L2 narrative skills (Squires et al., 2014) are capable of
574 differentiating bilingual children with DLD compared to bilingual TD peers. At a descriptive level, bilingual
575 TD children had L2 vocabulary scores around -1 SD compared to standardized values for monolingual peers,
576 and were at -0.85 SD in L2 morphological production. In the other L2 measures (narratives, morphological
577 comprehension, nonword repetition), their mean scores were within the average range. In contrast, LMBC
578 with DLD obtained very low scores in L2 morphological production (-3.04 SD), L2 vocabulary (74.95
579 Standard Score), and L2 morphological comprehension (< 15° percentile), as well as in nonword repetition
580 (around -1 SD). This pattern demonstrated that, although LMBC with typical development might fall behind
581 their monolingual peers in some measures of L2 achievement, and particularly in L2 vocabulary (Bialystok
582 et al., 2009), the profile of LMBC with DLD is severely impaired and different from both their bilingual and
583 monolingual peers with TD. The analysis of group difference is therefore of help in understanding the
584 functional linguistic profile of LMBC; however, it does not offer specific information as to which measures
585 better allow for the identification of DLD in LMBC.

586 To accomplish the two further aims of the study, we performed a set of discriminant analyses on the
587 measures included in the study protocol. First, we tested the discriminant validity of single measures and
588 then we combined a set of measures in different discriminant models, in order to define the best model in
589 terms of efficacy (specificity/sensitivity score) and efficiency (protocol length).

590 Considering single measures, it emerged that vocabulary scores in L1 (BABIL task), parents'
591 questionnaire (ALDEQ-IT), and L2 morphosyntactic comprehension reach fair indexes, even when
592 considered independently from each other. On the contrary, all the other measures taken separately had fair
593 to good specificity scores, but very poor sensitivity scores and an overall percentage of correctly identified
594 cases of below 80%. Data on the ALDEQ questionnaire replicate previous evidence (Paradis et al., 2010;
595 Bonifacci et al., 2016), that supports the good discriminating power of the questionnaire in sequential
596 bilingual children with and without DLD. In the study by Paradis et al., the ALDEQ showed high specificity
597 (96%) but medium-low sensitivity (66%). In the Italian validation, it reached high specificity (93.3%) and
598 fair sensitivity (Italian: 83.3%). In the present study, the sensitivity score increased to 90%, with fair
599 specificity (88.6%). In the following analyses, we kept only section B of the questionnaire, because it had
600 fair specificity (82.9%) and sensitivity (80%) scores and also because it is the most original and specific
601 section of the questionnaire for the indirect assessment children's linguistic skills in L1, based on the
602 parents' perspective. Furthermore, the choice to keep only one section was motivated by the need to develop
603 clinical protocols that combine validity with ease and rapidity of administration.

604 Regarding the models of discriminant analyses that aimed to define the best protocol for the
605 identification of DLD in LMBC, it resulted that model 3A was the one that reached the highest
606 specificity/sensitivity scores (91.4% and 100%, respectively), with a total percentage of correctly classified
607 cases of 94.5%, with the minimum number of measures included. The measures included in the model were:
608 L1 vocabulary (BABIL task), Section B of ALDEQ questionnaire, L2 morphosyntactic comprehension, L2
609 morphosyntactic production, and nonword repetition. Models 4 and 5 also reached 94.5 classification
610 accuracy scores, but with minor efficiency compared to model 3A. Therefore, the inclusion of L2 vocabulary
611 or narrative skills, which require a significant amount of time to administer (up to thirty minutes for the
612 PPVT) and score (particularly in the case of narrative tasks), did not significantly improve the classification
613 scores. Thus, model 3A reached the highest efficiency and efficacy scores.

614 The role of morphosyntactic comprehension found in the present study is in line with previous
615 evidence (e.g., Verhoeven, Steenge, van Weerdenburg, van Balkom, 2011, Paradis et al., 2013) which found
616 that skills in the morpho-syntax area were those that suffered more from the conjoint condition of
617 bilingualism and DLD. In the study by Verhoeven et al. (2011) the authors also found a disadvantage in the
618 lexical area, but they suggested that bilingualism had more influence than DLD on the scores of the lexicon
619 tasks, whereas language impairment was more specifically associated with deficits in the morpho-syntax
620 area. In other words, we might explain the important role of morphosyntactic comprehension in light of the
621 fact that bilingual children with typical development might catch up faster with their monolingual peers in
622 oral comprehension (see also Bonifacci and Tobia, 2016) compared to the time they need to reach
623 monolingual-like performance in vocabulary tasks. The same rationale might hold true for nonword
624 repetition (in line with Paradis et al., 2013), which is considered a measure that develops in a relatively short
625 time in bilingual children with typical development. Thus, delays in morphosyntactic comprehension and
626 nonword repetition, since they are not expected to be particularly influenced by having a bilingual profile,
627 might result as meaningful markers of DLD. In the present study, morphosyntactic production also resulted
628 as increasing discriminatory power significantly. As previously discussed, bilingual children with typical
629 development were around -0.85 sd compared to monolingual reference norms, suggesting that it might take
630 time to develop monolingual-like linguistic production skills in a bilingual condition (see also Paradis,
631 2016). However, the bilingual group with DLD scored more than -3 sd compared to monolingual referenced
632 norms and more than - 2 sd compared to bilingual peers with typical development. Therefore, as suggested
633 by many previous studies, morphosyntactic production can be considered as a robust marker of DLD in
634 LMBC. Considering narrative skills, the present study found that there was a significant difference between
635 bilingual children with and without DLD in microstructural (lexicon, MLU) aspects but only a tendency in
636 macrostructural aspects; these variables did not add discriminatory power when included in combined
637 models.

638 Finally, these results strongly reinforce the importance of the evaluation of L1 skills, either through
639 parents' questionnaires or direct assessment of receptive vocabulary conducted with children. Model 6, in
640 which the assessment of L1 skills was removed, and in which only standardized measures in L2 were
641 maintained, ended up as being the weakest model, with poor specificity/sensitivity indexes. The importance

642 of L1 assessment has been well established by ASHA guidelines and previous studies (e.g., Bedore & Peña,
643 2008; Gillam, Peña, and Miller, 1999). However, this good practice is still far from being easily translated
644 into everyday clinical practice, with particular reference to countries with a more recent history of
645 immigration and an increase of LMBC children in mainstream schools, as is the case in Italy. The present
646 study suggests that a short section of a parental interview and a task of receptive vocabulary in L1 might
647 offer sensitive and efficient tools that can be easily adopted in a clinical setting, even in the absence of
648 linguistic knowledge of minority languages on the part of speech language pathologists or the psychologist
649 conducting the assessment. The task of L1 vocabulary proposed in the present study has been developed in
650 many different minority languages, with cultural and linguistic adaptations for each language, and it is
651 presented through a pc with recorded audio so that the child merely needs to indicate the correct picture on
652 the pc monitor. Other ways to assess L1 vocabulary have been described in the literature (Peña, Bedore, &
653 Kester, 2016; Anaya, Peña, & Bedore, 2018).

654 Taken as a whole, the present study offers replication and extension of the usefulness of developing
655 and testing combined protocols for the identification of DLD in LMBC. Comparing results from the present
656 study and those obtained by Paradis et al. (2013), it emerged that in both studies, the ALDEQ questionnaire,
657 nonword repetition, and morphosyntactic production had significant discriminating power, suggesting that
658 these measures can be considered strong markers of DLD in LMBC, when compared to bilingual peers.
659 Considering sensitivity/specificity scores, the model that replicated Paradis et al.'s study obtained a similar
660 specificity index (91.4% vs. 92%) but lower sensitivity (85% vs. 91%). This was possibly due to the lowest
661 discriminating power observed in our study for story grammar. In point of fact, in the present study, the
662 optimal model did not include narrative skills and included two more measures than those used by Paradis et
663 al. (2013), namely morphosyntactic comprehension and L1 vocabulary. Furthermore, to increase protocol
664 efficiency, only section B of the ALDEQ questionnaire was included. Finally, in both studies, vocabulary, as
665 measured by the PPVT, did not yield additional discriminant power.

666 These results allow us to generalize the findings by Paradis et al., and suggest that the parental
667 questionnaire on L1 development, nonword repetition, and grammar/morphosyntactic production are good
668 discriminant measures for the identification of DLD in bilingual children. Furthermore, the results suggest
669 that this combined measure protocol also has fair validity in a different linguistic and cultural context.

670 However, contrasting results emerged for the role of story grammar. As previously discussed, earlier
671 literature failed to find consistent results regarding the role of narrative skills as a marker for DLD; this point
672 requires further investigation. Finally, the study adds further clues regarding the additional role of
673 morphosyntactic production and direct assessment of L1 vocabulary. The study was conducted in Italy,
674 where little evidence has been collected on the identification of DLD in LMBC; therefore, it offers a new
675 perspective into the generalization of previous results applied to a different linguistic and cultural context,
676 adding insights into the replicability of results mainly obtained in children with English as a second
677 language. In addition, the entire protocol length (Model 3A) is thought to have an acceptable length in
678 clinical practice since it takes around 1 hour and a half to administer: 30 minutes for the administration of
679 Section B of ALDEQ and linguistic history, 5 minutes for nonword repetition, 20 minutes for the TROG,
680 around 10 minutes for the grammar production task, and another 15 minutes for the L1 assessment (BABIL
681 Task).

682 **Potential Limitations**

683 Considering the measures included in the study, although we made our best effort to select Italian
684 versions of international standardized tasks, some of the measures were unique for the Italian context (e.g.,
685 morphosyntactic production, nonword repetition). Even if their characteristics are described in detail,
686 heterogeneity of measures might limit replicability of results. In addition, due to language diversity in L1, we
687 do not have specific control over inter-linguistic distance and influence of L1 (see Blom et al. 2012).
688 Furthermore, the assessment in L1, which was administered in a partially different condition (audio
689 recording) compared to the L2 assessment, and in the absence of a direct equivalent measure in L2, requires
690 further investigation. The use of the BABIL task for the assessment of L1 linguistic comprehension skills
691 revealed to be a potentially positive tool, but it was not developed as a diagnostic tool and further evidence
692 should be collected about its clinical validity. Although language-specific characteristics of measures are of
693 importance, it has to be underlined that the diagnostic issue addressed in the present study is not language-
694 specific. In particular, it was not possible to derive scores on specific grammar structures from the TROG
695 and, considering the BABIL task, we did not have sufficient information regarding markers of DLD in the
696 minority languages. Therefore, we do not have detailed information about the grammar measures that best

697 allow for the differentiation of bilinguals with and without DLD exposed to Italian as L2; future research is
698 needed to this regard.

699 Another limitation is related to the fact that, as in previous studies adopting a similar approach, the
700 identification of DLD was performed by clinicians who may have adopted different procedures; therefore,
701 we cannot exclude biases in their assessment tools and procedures. A differential approach could be that of
702 using the parent questionnaire as the gold standard and then classifying the experimental measures against a
703 well-established tool whose diagnostic utility has been proved in previous studies.

704 Finally, the study did not include dynamic assessment, which is considered one of the best practices
705 for the assessment of linguistic trajectories in bilingual children (De Lamo White & Jin, 2011; Peña, Gillam,
706 & Bedore, 2014). Further inquiries are needed that combine assessment protocols and dynamic assessment,
707 also considering children at a younger age in order to better prevent or minimize future difficulties.

708 **Conclusions**

709 Despite these limitations, this study offers important implications for the assessment of DLD in
710 language minority bilingual children. It reinforces the idea that no single measure can be considered optimal
711 for distinguishing children with DLD from typical peers (see meta-analysis by Dollaghan & Horner, 2011).
712 This study suggests that standard measures in L2, in the absence of L1 direct or indirect assessment, are not
713 the gold standard for identifying DLD in LMBC, although they have good discriminant validity when
714 included in a composite protocol. Ebert & Kohnert (2016) recently proposed that “*Creating composite*
715 *clinical markers – i.e., groups of tasks, perhaps implemented in both languages, which jointly possess*
716 *adequate sensitivity and specificity – may be a more valid approach for identifying LLI in bilingual*
717 *children*” (p. 317). The present study offers a concrete example of an effective and efficient protocol for the
718 discrimination of LMBC with and without DLD that may provide a valuable tool in different cultural and
719 linguistic settings.

720

721 **References**

722 Alt, M., & Spaulding, T. (2011). The effect of time on word learning: An examination of decay of the
723 memory trace and vocal rehearsal in children with and without specific language impairment. *Journal of*
724 *Communication Disorders, 44*(6), 640-654. <https://doi.org/10.1016/j.jcomdis.2011.07.001>

- 725 Anaya, J., Peña, E., & Bedore, L. (2018). Conceptual Scoring and Classification Accuracy of Vocabulary
726 Testing in Bilingual Children. *Language, Speech, and Hearing Services in Schools, 49*(1), 85–97.
727 https://doi.org/10.1044/2017_LSHSS-16-0081
- 728 Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of
729 specific language impairment (SLI) in bilingual children: evidence from Russian and
730 Hebrew. *International journal of language & communication disorders, 51*(6), 715-731.
731 <https://doi.org/10.1111/1460-6984.12242>
- 732 Bedore, L. M., & Pena, E. D. (2008). Assessment of bilingual children for identification of language
733 impairment: Current findings and implications for practice. *International Journal of Bilingual Education*
734 *and Bilingualism, 11*(1), 1-29. <https://doi.org/10.2167/beb392.0>
- 735 Bedore, L. M., Peña, E. D., Joyner, D., & Macken, C. (2011). Parent and teacher rating of bilingual language
736 proficiency and language development concerns. *International Journal of Bilingual Education and*
737 *Bilingualism, 14*(5), 489–511. <https://doi.org/10.1080/13670050.2010.529102>
- 738 Belacchi, C., Scalisi, T. G., Cannoni, E., & Cornoldi, C. (2008). Manuale CPM. *Coloured progressive*
739 *matrices. Standardizzazione italiana*. Firenze: Giunti OS
- 740 Bellocchi, S., Bonifacci, P., & Burani, C. (2014). Lexicality, frequency and stress assignment effects in
741 bilingual children reading Italian as a second language. *Bilingualism: Language and Cognition, 19*(1),
742 89–105. <https://doi.org/10.1017/s1366728914000297>
- 743 Bellocchi, S., Tobia, V., & Bonifacci, P. (2017). Predictors of reading and comprehension abilities in
744 bilingual and monolingual children: a longitudinal study on a transparent language. *Reading and Writing,*
745 *30*(6), 1311–1334. <https://doi.org/10.1007/s11145-017-9725-5>
- 746 Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2009). Receptive vocabulary differences in monolingual and
747 bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525–531.
748 <https://doi.org/10.1017/s1366728909990423>
- 749 Bishop, D. (2009). TROG 2–Test for reception of grammar Version 2. *Edizioni Giunti OS, Firenze*.
- 750 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T.; CATALISE-2 Consortium. (2017).
751 CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language

- 752 development. Phase 2. Terminology. *Journal of Child Psychology & Psychiatry*, 58, 1068–1080.
753 <https://doi.org/10.1111/jc>
- 754 Bisiacchi, P. S., Cendron, M., Gugliotta, M., Tressoldi, P. E., & Vio, C. (2005). *BVN 5-11: batteria di*
755 *valutazione neuropsicologica per l'età evolutiva*. Trento: Centro studi Erickson.
- 756 Blom, E., & Paradis, J. (2013). Past Tense Production by English Second Language Learners With and
757 Without Language Impairment. *Journal of Speech, Language, and Hearing Research*, 56(1), 281–294.
758 <https://doi.org/10.1044/1092-4388>
- 759 Blom, E., Paradis, J., & Duncan, T. S. (2012). Effects of Input Properties, Vocabulary Size, and L1 on the
760 Development of Third Person Singular -s in Child L2 English. *Language Learning*, 62(3), 965–994.
761 <https://doi.org/10.1111/j.1467-9922.2012.00715.x>
- 762 Boerma, T., & Blom, E. (2017). Assessment of bilingual children: What if testing both languages is not
763 possible?. *Journal of communication disorders*, 66, 65-76. <https://doi.org/10.1016/j.jcomdis.2017.04.001>
- 764 Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A Quasi-Universal
765 Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second
766 Language. *Journal of Speech, Language, and Hearing Research*, 58(6), 1747–1760.
767 https://doi.org/10.1044/2015_jslhr-l-15-0058
- 768 Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of
769 monolingual and bilingual children with and without language impairment: implications for clinical
770 practice. *International Journal of Language & Communication Disorders*, 51(6), 626–638.
771 <https://doi.org/10.1111/1460-6984.12234>
- 772 Bonifacci, P., & Tobia, V. (2016). Crossing barriers: Profiles of reading and comprehension skills in early
773 and late bilinguals, poor comprehenders, reading impaired, and typically developing children. *Learning*
774 *and Individual Differences*, 47, 17-26. <https://doi.org/10.1016/j.lindif.2015.12.013>
- 775 Bonifacci, P., Barbieri, M., Tomassini, M., & Roch, M. (2017). In few words: linguistic gap but adequate
776 narrative structure in preschool bilingual children. *Journal of Child Language*, 45(1), 120–147.
777 <https://doi.org/10.1017/s0305000917000149>

- 778 Bonifacci, P., Mari, R., Gabbianelli, L., Ferraguti, E., Montanari, F., Burani, F., & Porrelli, M. (2016).
779 Sequential bilingualism and Specific Language Impairment: The Italian version of ALDeQ Parental
780 Questionnaire. *BPA-Applied Psychology Bulletin (Bollettino di Psicologia Applicata)*, 64(275).
- 781 Bonifacci, P., Lombardo, G., Pedrinazzi, J., Terracina, F. & Palladino, P. (2019). Literacy Skills in
782 Bilinguals and Monolinguals with Different SES. *Reading and Writing Quarterly*. [https://doi.org/](https://doi.org/10.1080/10573569.2019.1635057)
783 10.1080/10573569.2019.1635057
- 784 Bonifacci, P., Tobia, V., Bernabini, L., & Marzocchi, G. M. (2016). Early Literacy and Numeracy Skills in
785 Bilingual Minority Children: Toward a Relative Independence of Linguistic and Numerical Processing.
786 *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01020>
- 787 Bortolini, U., Caselli, M. C., & Leonard, L. B. (1997). Grammatical Deficits in Italian-Speaking Children
788 With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 40(4), 809–
789 820. <https://doi.org/10.1044/jslhr.4004.809>
- 790 Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child*
791 *language teaching and therapy*, 18(1), 1-21. <https://doi.org/10.1191/0265659002ct224oa>
- 792 Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing Bias in Language Assessment.
793 *Journal of Speech, Language, and Hearing Research*, 40(3), 519–525.
794 <https://doi.org/10.1044/jslhr.4003.519>
- 795 Chiat, S. (2015). Nonword repetition. In: Armon-Lotem, de Jong, Meir (ed) *Assessing multilingual children:*
796 *Disentangling bilingualism from language impairment*, 125-150.
- 797 Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of
798 bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and*
799 *Hearing Research*, 59(5), 1179-1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293
- 800 Chondrogianni, V., & Marinis, T. (2011). Production and processing asymmetries in the acquisition of tense
801 morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, 15(1), 5–21.
802 <https://doi.org/10.1017/s1366728911000368>
- 803 Chondrogianni, V., Marinis, T., Edwards, S., & Blom, E. (2015). Production and on-line comprehension of
804 definite articles and clitic pronouns by Greek sequential bilingual children and monolingual children with

- 805 specific language impairment. *Applied Psycholinguistics*, 36(5), 1155-1191.
806 <https://doi.org/10.1017/S0142716414000101>
- 807 Contento, S., Bellocchi, S., & Bonifacci, P. (2013). BaBIL. Prove per la valutazione delle competenze
808 verbali e non verbali in bambini bilingui.[BaBIL. Battery for the evaluation of Verbal and Non verbal
809 Skills skills in bilingual children]. Firenze: Giunti OS.
- 810 Cossu, G., & Paris, E. (2007). TNP: Test neuropsicologico prescolare. Firenze: Giunti OS.
- 811 de Almeida, L., Ferré, S., Morin, E., Prévost, P., Dos Santos, C., Tuller, L., ... & Barthez, M. A. (2017).
812 Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches*
813 *to Bilingualism*, 7(3), 331-358. <https://doi.org/10.1075/lab.15019.alm>
- 814 De Bree, E., Rispens, J., & Gerrits, E. (2007). Non-word repetition in Dutch children with (a risk of)
815 dyslexia and SLI. *Clinical Linguistics & Phonetics*, 21(11-12), 935–944.
816 <https://doi.org/10.1080/02699200701576892>
- 817 De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with
818 bilingual children. *International Journal of Language & Communication Disorders*, 46(6), 613–627.
819 <https://doi.org/10.1111/j.1460-6984.2011.00049.x>
- 820 Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-Word and Nonword Repetition in Italian-Speaking
821 Children With Specific Language Impairment: A Study of Diagnostic Accuracy. *Journal of Speech,*
822 *Language, and Hearing Research*, 56(1), 323–336. [https://doi.org/10.1044/1092-4388\(2012/11-0304\)](https://doi.org/10.1044/1092-4388(2012/11-0304))
- 823 Dollaghan, C. A., & Horner, E. A. (2011). Bilingual Language Assessment: A Meta-Analysis of Diagnostic
824 Accuracy. *Journal of Speech, Language, and Hearing Research*, 54(4), 1077–1088.
825 [https://doi.org/10.1044/1092-4388\(2010/10-0093\)](https://doi.org/10.1044/1092-4388(2010/10-0093))
- 826 Dos Santos, C., & Ferré, S. (2018). A nonword repetition task to assess bilingual children’s
827 phonology. *Language Acquisition*, 25(1), 58-71. <https://doi.org/10.1080/10489223.2016.1243692>
- 828 Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test-revised*. American guidance service,
829 Incorporated.
- 830 Ebert, K. (2014). Role of auditory non-verbal working memory in sentence repetition for bilingual children
831 with primary language impairment. *International Journal of Language & Communication*
832 *Disorders*,49(5), 631–636. <https://doi.org/10.1111/1460-6984.12090>

- 833 Ebert, K., & Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Language*
834 *Teaching*, 49(3), 301–338. <https://doi.org/10.1017/S0261444816000070>
- 835 Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and Written Story
836 Composition Skills of Children With Language Impairment. *Journal of Speech, Language, and Hearing*
837 *Research*, 47(6), 1301–1318. [https://doi.org/10.1044/1092-4388\(2004/098\)](https://doi.org/10.1044/1092-4388(2004/098))
- 838 Fichman, S., Altman, C., Voloskovich, A., Armon-Lotem, S., & Walters, J. (2017). Story grammar elements
839 and causal relations in the narratives of Russian-Hebrew bilingual children with SLI and typical language
840 development. *Journal of Communication Disorders*, 69, 72–93.
841 <https://doi.org/10.1016/j.jcomdis.2017.08.001>
- 842 Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., ... & Walters, J. (2015).
843 Assessment of narrative abilities in bilingual children. In: Armon-Lotem, de Jong, Meir (ed) *Assessing*
844 *multilingual children: Disentangling bilingualism from language impairment*, 243-276.
- 845 Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied*
846 *psycholinguistics*, 27(4), 513-543. <https://doi.org/10.1017/s0142716406060383>
- 847 Gathercole, S. E., & Baddeley, A. D. (1989). The Role of Phonological Memory in Normal and Disordered
848 Language Development. *Brain and Reading*, 245–255. [https://doi.org/10.1111/j.2044-](https://doi.org/10.1111/j.2044-8295.1990.tb02371.x)
849 [8295.1990.tb02371.x](https://doi.org/10.1111/j.2044-8295.1990.tb02371.x)
- 850 Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all
851 depends on the nonwords. *Memory & Cognition*, 23(1), 83-94. <https://doi.org/10.3758/BF03210559>
- 852 Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014). The Receptive–Expressive Gap in Bilingual Children
853 With and Without Primary Language Impairment. *American Journal of Speech-Language Pathology*,
854 23(4), 655–667. https://doi.org/10.1044/2014_ajslp-12-0119
- 855 Gillam, R. B., Peña, E. D., & Miller, L. (1999). Dynamic assessment of narrative and expository
856 discourse. *Topics in Language Disorders*, 20(1), 33-47. [https://doi.org/10.1097/00011363-199911000-](https://doi.org/10.1097/00011363-199911000-00005)
857 [00005](https://doi.org/10.1097/00011363-199911000-00005)
- 858 Gillam, R., Peña, E., Bedore, L., Bohman, T., & Mendez-Perez, A. (2013). Identification of Specific
859 Language Impairment in Bilingual Children: I. Assessment in English. *Journal of Speech, Language, and*
860 *Hearing Research*, 56(6), 1813–1823. [https://doi.org/10.1044/1092-4388\(2013/12-0056\)](https://doi.org/10.1044/1092-4388(2013/12-0056))

- 861 Graham, H. R., Minhas, R. S., & Paxton, G. (2016). Learning Problems in Children of Refugee Background:
862 A Systematic Review. *Pediatrics*, 137(6), e20153994. <https://doi.org/10.1542/peds.2015-3994>
- 863 Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The Diagnostic Accuracy of Four Vocabulary
864 Tests Administered to Preschool-Age Children. *Language, Speech, and Hearing Services in Schools*,
865 30(2), 196–206. <https://doi.org/10.1044/0161-1461.3002.196>
- 866 Grimm, A., & Schulz, P. (2014). Specific language impairment and early second language acquisition: the
867 risk of over-and underdiagnosis. *Child Indicators Research*, 7(4), 821-841. [https://doi.org/](https://doi.org/10.1007/s12187-013-9230-6)
868 10.1007/s12187-013-9230-6.
- 869 Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and*
870 *language*, 36(1), 3-15. [https://doi.org/ 10.1016/0093-934x\(89\)90048-5](https://doi.org/10.1016/0093-934x(89)90048-5)
- 871 Gutiérrez-Clellen, V. F. (1996). Language diversity: Implications for assessment. *Assessment of*
872 *communication and language*, 6, 29-56. <https://doi.org/10.1177/152574019701900107>
- 873 Gutiérrez-Clellen, V. F., Simon-Cerejido, G., & Wagner, C. (2008). Bilingual children with language
874 impairment: A comparison with monolinguals and second language learners. *Applied*
875 *psycholinguistics*, 29(1), 3-19. [https://doi.org/ 10.1017/s0142716408080016](https://doi.org/10.1017/s0142716408080016)
- 876 Gutiérrez-Clellen, V. F., & Simon-Cerejido, G. (2010). Using nonword repetition tasks for the identification
877 of language impairment in Spanish-English-speaking children: Does the language of assessment
878 matter?. *Learning Disabilities Research & Practice*, 25(1), 48-58. [https://doi.org/ 10.1111/j.1540-](https://doi.org/10.1111/j.1540-5826.2009.00300.x)
879 5826.2009.00300.x
- 880 Gutiérrez–Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and
881 teacher reports. *Applied Psycholinguistics*, 24(2), 267–288. <https://doi.org/10.1017/s0142716403000158>
- 882 Hair, J., Black, W., Babin, B., Anderson, R., Tatham, R. (2006). *Multivariate Data Analysis*. 6th ed. Upper
883 Saddle River, NJ: Pearson Prentice Hall.
- 884 Hamann, C., Penner, Z. & Lindner, K. (1998). German Impaired Grammar: The Clause Structure Revisited.
885 Special issue 'Specific Language Impairment in Children.' *Language Acquisition* 7, 193-246.
886 https://doi.org/10.1207/s15327817la0702-4_5

- 887 Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015).
888 Narrative abilities in subgroups of English language learners and monolingual peers. *International*
889 *Journal of Bilingualism*, 19(6), 677-692. <https://doi.org/10.1177/1367006914534330>
- 890 Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early
891 bilingual development. *Journal of child language*, 39(1), 1-27.
892 <https://doi.org/10.1017/s0305000910000759>
- 893 Hollingshead, A. D. B. (1975). Four factor index of social status [1975].
- 894 Jacobson, P., & Livert, D. (2010). English past tense use as a clinical marker in older bilingual children with
895 language impairment. *Clinical Linguistics & Phonetics*, 24(2), 101-121. [https://doi.org/](https://doi.org/10.3109/02699200903437906)
896 [10.3109/02699200903437906](https://doi.org/10.3109/02699200903437906)
- 897 Jia, G., & Fuse, A. (2007). Acquisition of English Grammatical Morphology by Native Mandarin-Speaking
898 Children and Adolescents: Age-Related Differences. *Journal of Speech, Language, and Hearing*
899 *Research*, 50(5), 1280–1299. [https://doi.org/10.1044/1092-4388\(2007/090\)](https://doi.org/10.1044/1092-4388(2007/090))
- 900 Kohnert, K., Windsor, J., & Yim, D. (2006). Do Language-Based Processing Tasks Separate Children with
901 Language Impairment from Typical Bilinguals? *Learning Disabilities Research and Practice*, 21(1), 19–
902 29. <https://doi.org/10.1111/j.1540-5826.2006.00204.x>
- 903 Lehti, V., Gyllenberg, D., Suominen, A., & Sourander, A. (2018). Finnish-born children of immigrants are
904 more likely to be diagnosed with developmental disorders related to speech and language, academic skills
905 and coordination. *Acta Paediatrica*, 107(8), 1409–1417. <https://doi.org/10.1111/apa.14308>.
- 906 Leonard, L. B. (2014). Specific language impairment across languages. In *Speech and language impairments*
907 *in children* (pp. 129-144). Psychology Press.
- 908 Leonard, L. B., Caselli, M. C., Bortolini, U., McGregor, K. K., & Sabbadini, L. (1992). Morphological
909 deficits in children with specific language impairment: The status of features in the underlying
910 grammar. *Language Acquisition*, 2(2), 151-179.
- 911 McDonald, J. L., & Oetting, J. B. (2019). Nonword repetition across two dialects of English: Effects of
912 specific language impairment and nonmainstream form density. *Journal of Speech, Language, and*
913 *Hearing Research*, 62(5), 1381-1391.

- 914 Marini, A., Marotta, L., Bulgheroni, S., & Fabbro, F. (2015). *BVL 4-12 Batteria per la Valutazione del*
915 *Linguaggio in Bambini dai 4 ai 12 anni*. Firenze: Giunti Psychometrics
- 916 MIUR (2018). *Gli Alunni Con Cittadinanza Non Italiana A.s. 2016/2017* [Not-Italian citizenship students
917 2016/17]. Milano, Graphidea.
- 918 Oller, D. K., & Pearson, B. Z. (2002). Chapter 1: Assessing the Effects of Bilingualism: A Background.
919 *Language and Literacy in Bilingual Children*, 3–21. <https://doi.org/10.21832/9781853595721-002>
- 920 Orgassa, A., & Weerman, F. (2008). Dutch gender in specific language impairment and second language
921 acquisition. *Second Language Research*, 24(3), 333–364. <https://doi.org/10.1177/0267658308090184>
- 922 Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Applied*
923 *Psycholinguistics*, 31, 3–28. <https://doi.org/10.1017/s0142716409990373>
- 924 Paradis, J. (2016). The development of English as a second language with and without specific language
925 impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research*, 59(1), 171-182.
926 https://doi.org/10.1044/2015_jslhr-l-15-0008
- 927 Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French-English Bilingual Children With SLI. *Journal*
928 *of Speech, Language, and Hearing Research*, 46(1), 113–127. [https://doi.org/10.1044/1092-](https://doi.org/10.1044/1092-4388(2003/009))
929 [4388\(2003/009\)](https://doi.org/10.1044/1092-4388(2003/009))
- 930 Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent
931 report on first language development. *Journal of Communication Disorders*, 43(6), 474–497.
932 <https://doi.org/10.1016/j.jcomdis.2010.01.002>
- 933 Paradis, J., Jia, R., & Arppe, A. (2017). The acquisition of tense morphology over time by English second
934 language children with specific language impairment: Testing the cumulative effects hypothesis. *Applied*
935 *Psycholinguistics*, 38(4), 881–908. <https://doi.org/10.1017/s0142716416000485>
- 936 Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating Children With Language Impairment
937 Among English-Language Learners From Diverse First-Language Backgrounds. *Journal of Speech,*
938 *Language, and Hearing Research*, 56(3), 971–981. [https://doi.org/10.1044/1092-4388\(2012/12-0050\)](https://doi.org/10.1044/1092-4388(2012/12-0050))
- 939 Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time:
940 Individual variation in long-term outcomes. *Journal of Child Language*, 43(3), 553-580. [https://doi.org/](https://doi.org/10.1017/s0305000915000562)
941 [10.1017/s0305000915000562](https://doi.org/10.1017/s0305000915000562)

- 942 Paradis, M. (1987). *Bilingual aphasia test*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 943 Patterson, J. L. (2004). Comparing Bilingual and Monolingual Toddlers' Expressive Vocabulary
944 Size. *Journal of Speech, Language, and Hearing Research*, 47(5) 1213-1215. [https://doi.org/
945 10.1044/1092-4388\(2004/089\)](https://doi.org/10.1044/1092-4388(2004/089))
- 946 Peña, E., Bedore, L., & Kester, E. (2016). Assessment of language impairment in bilingual children using
947 semantic tasks: two languages classify better than one. *International Journal of Language &
948 Communication Disorders*, 51(2), 192–202. <https://doi.org/10.1111/1460-6984.12199>
- 949 Peña, E., Gillam, R., & Bedore, L. (2014). Dynamic Assessment of Narrative Ability in English Accurately
950 Identifies Language Impairment in English Language Learners. *Journal of Speech, Language, and
951 Hearing Research*, 57(6), 2208–2220. https://doi.org/10.1044/2014_JSLHR-L-13-0151
- 952 Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language,
953 Speech, and Hearing Services in Schools*, 25, 15–24. [https://doi.org/ 10.1044/0161-1461.2501.15](https://doi.org/10.1044/0161-1461.2501.15)
- 954 Rescorla, L. (1993, November). Use of parental report in the identification of communicatively delayed
955 toddlers. In *Seminars in Speech and Language* (Vol. 14, No. 04, pp. 264-277). Thieme Medical
956 Publishers, Inc..
- 957 Rice, M. L., & Wexler, K. (1996). Toward Tense as a Clinical Marker of Specific Language Impairment in
958 English-Speaking Children. *Journal of Speech, Language, and Hearing Research*, 39(6), 1239–1257.
959 <https://doi.org/10.1044/jshr.3906.1239>
- 960 Rothweiler, M., Chilla, S., & Clahsen, H. (2011). Subject–verb agreement in Specific Language Impairment:
961 A study of monolingual and bilingual German-speaking children. *Bilingualism: Language and Cognition*,
962 15(1), 39–57. <https://doi.org/10.1017/s136672891100037x>
- 963 Salameh, E. K., Nettelbladt, U., Håkansson, G., & Gullberg, B. (2002). Language impairment in Swedish
964 bilingual children: a comparison between bilingual and monolingual children in Malmö. *Acta
965 Paediatrica*, 91(2), 229-234. <https://doi.org/10.1080/080352502317285261>
- 966 Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2009). The home language environment of monolingual
967 and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1), 117–140.
968 <https://doi.org/10.1017/s0142716409990191>

- 969 Sheng, L., Bedore, L. M., Peña, E. D., & Fiestas, C. (2012). Semantic Development in Spanish-English
970 Bilingual Children: Effects of Age and Language Experience. *Child Development, 84*(3), 1034–1045.
971 <https://doi.org/10.1111/cdev.12015>
- 972 Sorenson Duncan, T., & Paradis, J. (2018). How does maternal education influence the linguistic
973 environment supporting bilingual language development in child second language learners of English?
974 *International Journal of Bilingualism*. <https://doi.org/10.1177/1367006918768366>
- 975 Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story
976 retelling by bilingual children with language impairments and typically developing controls. *International*
977 *Journal of Language & Communication Disorders, 49*(1), 60-74. [https://doi.org/10.1111/1460-](https://doi.org/10.1111/1460-6984.12044)
978 [6984.12044](https://doi.org/10.1111/1460-6984.12044)
- 979 Stella, G., Pizzoli, C., & Tressoldi, P. E. (2000). *Peabody test di vocabolario recettivo*. Torino: Omega
980 Edizioni.
- 981 Suraniti, S., Ferri, R., & Neri, V. (2009). *TROG-2 adattamento italiano*. Firenze: Giunti OS.
- 982 Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on
983 nonword repetition and sentence imitation scores. *Journal of Communication Disorders, 46*(1), 1–16.
984 <https://doi.org/10.1016/j.jcomdis.2012.08.002>
- 985 Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall
986 proficiency be estimated from separate measurement of two languages? *Journal of Multilingual*
987 *Communication Disorders, 4*(1), 1-21. <https://doi.org/10.1080/14769670500215647>
- 988 Tsimpli, I. M., Peristeri, E., & Andreou, M. (2015). Narrative production in monolingual and bilingual
989 children with specific language impairment. *Applied Psycholinguistics, 37*(1), 195–216.
990 <https://doi.org/10.1017/s0142716415000478>
- 991 Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., Santos, C., et al. (2018). Identifying
992 language impairment in bilingual children in France and in Germany. *International Journal of Language*
993 *& Communication Disorders, 53*(4), 888–904. <https://doi.org/10.1111/1460-6984.12397>
- 994 Verhoeven, L., Steenge, J., & van Balkom, H. (2012). Linguistic transfer in bilingual children with specific
995 language impairment. *International Journal of Language & Communication Disorders, 47*(2), 176-183.
996 <https://doi.org/10.1111/j.1460-6984.2011.00092.x>

- 997 Verhoeven, L., Steenge, J., van Weerdenburg, M., & van Balkom, H. (2011). Assessment of second
998 language proficiency in bilingual children with specific language impairment: A clinical
999 perspective. *Research in Developmental Disabilities, 32*(5), 1798-1807. [https://doi.org/
1000 10.1016/j.ridd.2011.03.010](https://doi.org/10.1016/j.ridd.2011.03.010)
- 1001 Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (1999). *Comprehensive test of*
1002 *phonological processing: CTOPP*. Austin, TX: Pro-ed.
- 1003 Williams, C., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices
1004 with multilingual children. *International Journal of Speech-Language Pathology, 14*(3), 292–305.
1005 [https://doi.org/ 10.3109/17549507.2011.636071](https://doi.org/10.3109/17549507.2011.636071)
- 1006 Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language nonword repetition by
1007 bilingual and monolingual children. *American Journal of Speech-Language Pathology, 19*(4), 298–310.
1008 [https://doi.org/10.1044/1058-0360\(2010/09-0064\)](https://doi.org/10.1044/1058-0360(2010/09-0064))
- 1009 World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical*
1010 *descriptions and diagnostic guidelines*. Geneva: World Health Organization.
- 1011
- 1012
- 1013
- 1014
- 1015
- 1016
- 1017
- 1018
- 1019
- 1020
- 1021
- 1022
- 1023
- 1024

1025 Table 1. Languages spoken in the two groups (TD, DLD)

	Albanian	Arabic	Bengali	Chinese	Polish	Romanian	Urdu
TD	14.30%	71.40%	0.00%	0.00%	0.00%	8.60%	5.70%
DLD	20.00%	35.00%	5.00%	5.00%	10.00%	10.00%	15.00%

1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069

1070 Table 2. Descriptive statistics on Age, AoE, MoE, and SES for the two bilingual samples (TD,
 1071 DLD).

	Typical Development		DLD		<i>t</i>	<i>p</i>	Cohen's <i>d</i>
	Mean	SD	Mean	SD			
Age (months)	82.49	6.34	84.85	6.56	-1.31	.19	-0.37
AoE (Age of Exposure)	38.57	9.26	34.70	15.57	1.01	.32	0.31
Months of Exposition ITL2 (MoE)	44.09	9.45	50.55	16.24	-1.63	.12	-0.50
SES	18.09	6.39	19.40	6.25	-0.74	.46	-0.21

1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107
 1108
 1109
 1110

1111 Table 3. Mean scores for all variables included in the study protocol for the two groups (DLD, TD), together
 1112 with statistics and effect size (Cohen's *d*). In the last column mean scores based on standardized scores for
 1113 monolingual children (-1/+1 sd) are reported.

	Typical Development		DLD		<i>t</i>	<i>p</i>	Cohen's D	Mean scores based on standardized scores for monolingual children (-1/+1 sd)
	Mean	SD	Mean	SD				
Intellectual functioning (Raven) §	75.91	24.14	74.50	26.22	0.20	.84	0.06	25-75
Vocabulary (PPVT) °	85.71	12.69	74.95	9.97	3.26	**	0.95	85-115
Nonword repetition (BVN)^	0.46	0.93	-0.96	1.65	4.10	**	1.10	-1 / +1
Morphosyntactic comprehension (TROG) §	61.57	32.31	14.45	14.55	6.16	**	2.01	25-75
Morphosyntactic production (TNP)^	-0.85	1.58	-3.04	2.09	4.40	**	1.20	-1 / +1
Narratives (words per minute) (BVL) ^	0.48	0.79	-0.18	0.75	3.04	**	0.86	-1 / +1
Narratives (Mean Length Utterance) (BVL)^	0.41	1.66	-1.22	1.28	3.79	**	1.11	-1 / +1
Narratives Type (BVL)^	0.38	1.21	-0.57	1.19	2.89	**	0.79	-1 / +1
Narrative (Macrostructure) (BVL)^	0.19	1.03	-0.34	0.86	1.95	.056	0.56	NA
Narrative Total (BVL)^	0.37	0.94	-0.58	0.68	3.94	**	1.17	NA
L1 Vocabulary (BABIL)^	0.80	1.04	-0.53	0.66	5.15	**	1.57	NA
L1 Morphosyntactic skills (BABIL)^	0.40	1.00	-0.77	1.64	3.32	**	0.89	NA
L1 Basic vocabulary (BABIL)^	0.19	0.95	-0.75	0.80	3.73	**	1.07	NA
L1 Inferences (BABIL)^	0.33	0.77	-0.72	1.60	3.30	**	0.89	NA
L1 Total Score (BABIL)^	0.64	0.81	-0.86	1.12	5.74	**	1.56	NA
ALDeQ-IT section A^	-0.35	1.35	-3.37	2.99	5.16	**	1.39	NA
ALDeQ-IT section B^	-0.04	0.83	-1.54	0.60	7.13	**	2.11	NA
ALDeQ-IT section C^	0.13	1.09	-1.02	0.64	4.30	**	1.33	NA
ALDeQ-IT section D^	0.04	0.83	-0.25	0.85	1.20	.24	0.33	NA
ALDeQ-IT Total ^	-0.16	1.05	-2.69	1.06	8.54	**	2.39	NA

1114 ** *p* < .01; ^z scores, °standard scores, § percentile

1115
 1116
 1117
 1118
 1119
 1120
 1121
 1122

1123 Table 4. Output of discriminant analyses for all measures included in the study and, when available, their
 1124 subscales.

	Wilks Lambda	Chi Square	Sign.	Specificity	Sensitivity	% cases correctly classified
Intellectual functioning (Raven) §	.99	0.41	.84	100%	0%	63.6%
SES	.99	0.54	.46	97.1%	5%	63.6%
AoE	.98	1.31	.25	94.3%	25%	69.1%
MoE	.94	3.36	.07	94.3%	25%	69.10%
Morphosyntactic comprehension (TROG) §	.58	28.32	**	77.10%	85%	80%
Nonword repetition (BVN)^	.76	14.49	**	91.40%	50%	76.4%
Vocabulary (PPVT) °	.83	9.58	**	77.10%	50%	67.3%
Morphosyntactic production (TNP)^	.73	16.36	**	85.70%	50%	72.7%
Narrative Total (BVL)^	.77	13.49	**	85.7%	50%	72.7%
Narrative (Macrostructure) (BVL)^	.93	3.65	**	94.3%	20%	67.30%
Narratives Type (BVL)^	.87	7.38	**	82.90%	40%	67.3%
Narratives (words per minute) (BVL) ^	.85	9.21	**	82.90%	35%	65.5%
Narratives (Mean Length Utterance) (BVL)^	.79	12.61	**	88.60%	45%	72.7%
L1 Total Score (BABIL)^	.62	25.40	**	91.40%	55%	78.2%
L1 Vocabulary (BABIL)^	.67	21.28	**	91.40%	70%	83.6%
L1 Morphosynatctic skills (BABIL)^	.83	9.90	**	85.70%	35%	67.3%
L1 Basic vocabulary (BABIL)^	.79	12.24	**	82.90%	50%	70.9%
L1 Inferences (BABIL)^	.83	9.79	**	100.00%	30%	74.5%
ALDeQ-IT Total ^	.420	45.5	**	88.60%	90%	89.1%
ALDeQ-IT section A^	.68	20.28	**	94.30%	50%	78.2%
ALDeQ-IT section B^	.51	35.23	**	82.90%	80%	81.8%
ALDeQ-IT section C^	.72	17.61	**	80.00%	60%	72.2%
ALDeQ-IT section D^	.97	1.41	.23	91.40%	5%	60%

1125 ** $p < .01$ ** $p < .01$; ^z scores, °standard scores, § percentile

1126

1127

1128

1129 Table 5. Output of discriminant analyses for combined models, ordered for classification accuracy
 1130 (efficacy) and protocol length (efficiency).

	Wilks Lambda	Chi Square	Sign.	Standardized coefficients	Specificity	Sensitivity	% cases correctly classified
MODEL 3A: Model 1 + Nonword repetition + Morphosyntactic production	.38	48.87	**	ALDEQ-IT B: .577 TROG: .28; BABIL 1: .32; BVN: .14; TNP: .27	91.4%	100%	94.5
MODEL 4: Model 1 + Nonword repetition+ Narrative skills + Morphosyntactic production	.38	48.78	**	ALDEQ-IT B: .56 TROG: .27; BABIL 1: .34; BVN: .12; BVL: .13; TNP: .23	91.4%	100%	94.5
MODEL 5: Model 1 + Nonword repetition + Morphosyntactic production + L2 Vocabulary	.38	48.38	**	ALDEQ-IT B: .58 TROG: .28; BABIL 1: .33; BVN: .14; TNP: .28; PPVT: .004	91.4%	100%	94.5
MODEL 3B: Model 1 + Nonword repetition + Narrative skills	.385	48.14	**	ALDEQ-IT B: .53 TROG: .38; BABIL 1: .35; BVN: .12; BVL: .198	88.6%	100%	92.7
MODEL 2: Model 1 + Nonword repetition	.394	47.55	**	ALDEQ-IT B: .576 TROG: .43; BABIL 1: .34; Nonword rep: .16	85.7%	100%	90.9%
Model 1. L1 vocabulary + ALDEQ-IT B + Morphosyntactic comprehension	.399	47.35	**	ALDEQ-IT B: .576 TROG: .49; BABIL 1: .35	82.9%	100%	89.1%
Model 7: replication of Paradis et al. (2013) ALDEQ total score, nonword repetition, morphosyntactic production, story grammar	.40	46.71	**	ALDEQ-IT B: .89; BVN: .08; TNP: .30; BVL: - .21	91.4%	85%	89.1%
MODEL 6: L2 MEASURES (Vocabulary, morphosyntactic production and comprehension, nonword repetition)	.54	30.73	**	TROG: .68; BVN: .30; TNP: .25; PPVT: .04	77.1%	80%	78.2

** $p < .01$

1131