Received: 13 February 2022 | Revised: 7 September 2023 | Accepted: 12 September 2023

DOI: 10.1002/dys.1753

RESEARCH ARTICLE

WILEY

Disentangling dyslexia from typical L2-learning in emergent literacy

Turid Helland[1] | Frøydis Morken[1] | Wence A. Helland[1,2]

[1]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

[2]Department of Research and Innovation, Helse Fonna Health Authority, Norway Helse Fonna HF, Haugesund, Norway

Correspondence
Turid Helland, Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway.
Email: turid.helland@uib.no

The present paper assessed how dyslexia can be identified in school children with another language than their first language. Participants were school children with Norwegian as their second language (L2), and two groups of children with Norwegian as their first language (L1): a control group (L1-Con), and a dyslexia group (L1-Dys). All were 2nd and 3rd graders who had attended Norwegian schools from 1st grade on. None of the individuals in L1-Con or the L2 group were identified with any learning disability. However, slow literacy progress was seen in some L2-children. The children were tested individually within the symptomatic and cognitive levels. Results were analysed in two steps: (1) group comparisons; (2) L2 individual profiles and tentative L2 subgrouping. An unexpected L2 profile showed language scores below norm, coupled with some scores within and some scores above norm within the cognitive domain. Case assessment of the L2 group resulted in three subgroups: one comparable to L1-Con, one comparable to L1-Dys, and one with a result in between these two groups. Low linguistic scores cannot be considered valid markers of dyslexia in L2. Within the cognitive domain, a variety of low scores can indicate dyslexia, while high scores can be compensatory.

KEYWORDS
assessment, cognition, dyslexia, literacy, multilingualism

Practitioner Points

- Low literacy skills are not valid markers of dyslexia in L2.

- L2 children with unexpected low literacy skills should be assessed using the four-level model (symptomatic, cognitive, biological, environmental levels).
- In the present study assessment at the symptomatic and cognitive levels supplied by information at the biological and environmental levels emerged three L2 profiles.
- One profile was comparable to L1 controls, one comparable to L1 dyslexia, and one profile was in a middle position between these two groups.
- Unexpected high L2 scores on perceptual auditive and visual tests was interpreted as compensatory of low L2 linguistic skills.

1 | INTRODUCTION

It is estimated that more than half of the world's population is multilingual to some extent (Bialystok, 2017). Several studies have compared the performance of monolingual and multilingual school children, and according to some of these studies, there are cognitive benefits from multilingualism (Fox et al., 2019; Ramírez & Kuhl, 2016), while other studies point to disadvantages (Karlsen et al., 2017). In her review of studies of bilingual adaptation, Bialystok (2017) concludes in this way: 'It is perhaps not surprising that bilingualism changes the way language processing is carried out; it is certainly less expected that it also changes the way that nonverbal cognitive processing is conducted' (p. 41). The aim of this study was to assess how typical literacy L2 problems can be differentiated from L2 literacy problems associated with dyslexia.

It is often unclear how multilingual problems can be understood, and even more so when dyslexia is in question. Dyslexia is a disorder independent of languages and orthographies (Pugh & Verhoeven, 2018) and estimates indicate a prevalence of 5%–10% depending on definition (Miles, 2004). Definitions of dyslexia vary, but point to a constitutional disorder, that affects the development of literacy and language-related skills (BDA, 2007; Helland, 2022b; Lyon et al., 2003). Typical cognitive benchmarks are impaired processing speed, phonological processing, working memory and rapid naming (Baddeley, 2003; Pennington, 2006). In sum, dyslexia is a multifactorial impairment with idiosyncratic outcomes. Usually, but not always, it needs appropriate, specific intervention (Cooke & Adams, 2007).

Separating observed language problems from dyslexia in multilinguals is a challenge for teachers and clinicians. Studies focusing on foreign language learning (mostly English) in individuals with dyslexia have shown differences in both linguistic and brain measures (Helland & Kaasa, 2005; Helland & Morken, 2016; Ylinen et al., 2019). According to a recent meta-analysis of foreign language learning in multilinguals with poor literacy skills more research is needed (von Hagen et al., 2020). A way of piloting this problem is to assess children with and without dyslexia whose first language (L1) is also their school language, and a matched group of multilingual children whose L1 is different from their school language, here labelled second language (L2) (for terminology, see The Council of Europe, 2007). However, an investigation of dyslexia and multilingualism must be grounded in an understanding of literacy development and different levels of explanation. To structure the content of the present study a model of literacy development and a causal model were applied (see, i.e., Helland, 2022a).

1.1 | Literacy stages

In general, children go through three literacy stages (Ehri, 1987; Frith, 1986). The *pre-literacy stage* is before formal training starts in school. Some children 'play read' and 'play write', some 'break the code' while in kindergarten, but most children do not. *The emergent literacy stage* starts with formal literacy training, focusing on learning the principles of the literacy code through grapheme-phoneme combinations. *The literacy stage* is when reading and writing have become automatized and are expected to be tools for academic work. At-risk factors of dyslexia can be identified already in the pre-literacy stage (Helland et al., 2021; Snowling et al., 2015), while dyslexia becomes evident mainly at the emergent literacy stage. In general literacy development in L2 is a complex transition between L1 and L2 not only as to language typology and pragmatics but also as to policy of language education (Luk & Bialystok, 2013).

1.2 | A causal model

The many theories about what dyslexia is and how dyslexia can be understood have led to uncertainty and disagreements among researchers, clinicians, teachers and not least those concerned. One conclusion to this discussion is that several theories can explain dyslexia, but no one alone can give an unambiguous explanation (Peterson & Pennington, 2015).

The causal model put forth by Morton and Frith (1995) provided a structural approach to several types of deviations. Figure 1 illustrates the four basic interactive levels of explanation. The symptomatic level concerns observable behaviour, the cognitive level concerns the typical cognitive benchmarks of the actual impairment, which again can be understood in light of known or anticipated biological factors. Further, the environmental level implies the culture in which the person grows up.

Frith (1999) elaborated on this model by focusing on different aspects of dyslexia. These were dyslexia due to a phonological failure, a magnocellular or a cerebellar abnormality, or due to an attention deficit disorder. She also pointed out the need for more knowledge of information-processing mechanisms, encompassing difficulties with visual, auditory, and temporal processing. Accordingly, Frith argued that there should be no conflicts between the theories at the different levels. But that valid and reliable assessments are required at all levels, as well as good clinical judgement. The flexibility of the model, as demonstrated by Frith, and its structural approach, has made it sustainable and classical not at least in assessments of learning disabilities (McGrath et al., 2020).

At each literacy stage, skills and impairments should be assessed from different levels of explanation: symptomatic, cognitive, biological, and environmental (Frith, 1986; Morton & Frith, 1995).

1.2.1 | Symptomatic level

The behavioural expression of dyslexia and multilingualism can often be similar. In the emergent literacy stage reading and writing scores are typically low in both dyslexia and L2 (van Setten et al., 2017). Moreover, many children with dyslexia have comorbid Developmental Language Disorder (DLD) or other oral language issues, meaning that the problems they face in the transition between oral and written language are comparable to those that meet L2 children (Ramírez & Kuhl, 2016). In school, similar intervention principles are often applied to the two groups. However, no conclusions about diagnosis or the aetiology of literacy problems should be based on symptoms alone. The behavioural expression we observe may have both internal (cognitive or biological) and external (environmental) causes. Therefore, qualified assessments of underlying cognitive processes, biological factors and the child's environment are of essence.

1.2.2 | Cognitive level

Single deficit models as the phonological deficit model in dyslexia have proved to be inadequate (McGrath et al., 2020). Rather, in dyslexia, different cognitive profiles, especially within the framework of the multi-component working memory model, are shown (Baddeley, 2003; Gray et al., 2019) Children's literacy learning is affected by their

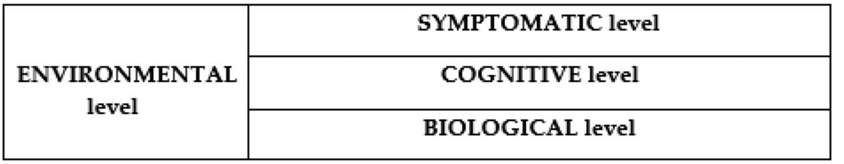

FIGURE 1 Basic causal model.

cognitive, linguistic, metalinguistic, and metacognitive skills. These are often divided into *domain-specific* skills (e.g., phonological and morphological awareness, vocabulary, grammatical knowledge, and metacognition) and *domain-general* factors, (e.g., working memory, executive functions, and processing speed) (O'Brien et al., 2014). Studies have shown weak domain-specific skills both in children at risk of developmental dyslexia (Brinchmann et al., 2015) and in emergent L2 learners (Karlsen et al., 2017). However, it has been argued that phonological awareness in L2 is language-specific with varied outcomes due to L1-L2 phonological distances (Saiegh-Haddad, 2019). Hence, impaired phonological awareness should not be seen as a reliable indicator of dyslexia in L2. It is likely that this same argument holds for other domain-specific skills. Impairments within domain-general factors, on the other hand, are associated with dyslexia (Hachmann et al., 2020; Norton & Wolf, 2012; O'Brien et al., 2014).

1.2.3 | Biological level

At the biological level studies of language learning focus mainly on heredity, gender, laterality, and brain structure and function. Estimations point to 40%–60% heredity in dyslexia (Peterson & Pennington, 2015). The gender distribution is about equal (Shaywitz et al., 1990), but with earlier at-risk signs in boys compared to girls (Snowling et al., 2015; Zambrana et al., 2015). Also, boys seem to be consistently more impaired than girls (Berninger et al., 2008b). Multiple theories have been proposed in linking hand dominance, brain lateralization, and genetics to language functions, but most questions remain unsolved (Bishop & Bates, 2020; McManus, 2019). For language lateralisation in L2 two hypotheses have been proposed. *The age of language acquisition* hypothesis proposes that early, simultaneous language learning will lead to a more left-hemispheric lateralisation, while later acquisition will lead to a stronger involvement of the right hemisphere. *The stage of language acquisition* hypothesis proposes that high degree of L2 proficiency will lead to a more left-hemispheric dominance compared to a lower degree of proficiency (Ocklenburg & Gunturkun, 2017).

1.2.4 | Environmental level

Studies of language impairment and dyslexia have shown that environmental factors are important for outcome (Hofslundsengen et al., 2019). Both the home environment of L2 children and the age at which they were introduced to L2 influence their learning (Karlsen et al., 2017). Children with another home language than Norwegian are usually either children of asylum seekers (refugees) or children of immigrant workers.

Figure 2 adds to the original model by Morton and Frith (1995) summing the current multifactorial viewpoints of what dyslexia is.

School systems around the world vary in accordance with political systems, economy, and philosophy of education. Naturally, these variations affect the educational opportunities of all children, and not least those who come from other countries and cultures. Also, this makes comparisons across societies and languages difficult. The Norwegian school system is mainly public, with only a few private schools. It is unitary, for all and inclusive with rights of special education. According to The Norwegian Directorate for Education and Training (2022) children with another first language have rights to

<table>
<tr><td rowspan="3">Environmental level

Home
Education
Intervention</td><td>Symptom level
Reading, writing, literacy</td></tr>
<tr><td>Cognitive level
Domain specific: language (comprehension, processing), phonological awareness,
Domain general: Attention, STM/WM/LTM (auditory, visual), executive functions</td></tr>
<tr><td>Biological level
Brain, genetics/heredity, gender, laterality</td></tr>
</table>

Notes: levels (in bold) refer to the original model form Morton & Frith (1995). Descriptions of each level from Helland (2022a)

FIGURE 2 Basic dyslexia model based on the causal model of Morton and Frith (1995). Levels (in bold) refer to the original model form Morton and Frith (1995). Descriptions of each level from Helland (2022a).

'Introductory support in primary, lower secondary and upper secondary schools. Children and youth who have recently arrived in Norway, may receive all or some education in separate groups, classes or schools for a period of up to two years. The aim is to learn the Norwegian language quickly, before entering regular schools or classes' (p. 6).

Most children in Norway attend kindergarten before entering school at age six. Before and after-school programmes are offered to all with age-adjusted activities. Literacy training starts in first grade. The Norwegian language has a semitransparent orthography, with 29 letters and 36 graphemes representing around 40 phonemes (Helland & Kaasa, 2005). In comparison, English has 26 letters and 561 graphemes, representing 44 phonemes (Dewey, 1971). Norwegian teachers typically combine a low-level, or phonic, approach with a high-level, or whole-language, approach (Hagtvet et al., 2006).

All children are entitled to adapted education, meaning that teaching should meet their individual needs. Some L2 children receive special education, but for literacy issues, customized training in the regular classroom is more common. There seems to be an overrepresentation of L2 children with language issues in special education (Kim & Helphenstine, 2017), which could indicate that schools interpret their challenges differently from those of L1 children. Moreover, even though evidence-based screening tools completed by parents and teachers can add valuable information about the individual child's development (Glascoe, 1997; Helland et al., 2021), language problems and cultural issues may result in fewer parents of L2 children raising questions about their child's development.

1.3 | Purpose and hypotheses

Based on the causal model as shown in Figure 2, the aim of the study was twofold: (1) to compare an L2 profile to a typical L1 profile and a typical L1 dyslexia profile; (2) tentatively to disentangle the L2 profile into a typical L2 profile and an L2 dyslexia profile. As to (1) we expected overall low L2 scores at the symptomatic level (reading and spelling) and low scores within the cognitive domain-specific area. In the cognitive domain-general area we expected scores comparable to the L1 typical group. As to (2) we expected that low domain-general scores would indicate dyslexia in L2.

2 | METHOD

2.1 | Participants and approvals

Participants ($N = 144$) consisted of three groups of children attending Norwegian public schools. All children were in the emergent literacy stage, more specifically in the last half of second and first half of third grade of primary school.

All had attended their home schools from 1st grade, with Norwegian as their primary instructional language. None of the children had any intellectual disability and/or diagnosis of any neurological impairments as reported by schools and parents. Overall, the participants attended 20 different Norwegian public primary schools located in a widespread geographical area.

2.1.1 | L2 participants

Four regular public schools from an urban district accepted to participate. In the formal information letter and in meetings with the schools (administrators, counsellors, teachers, and special educators) the project leaders presented the project. Inclusion criteria were (a) L2 children attending regular Norwegian classes and (b) should be a mixture of L2 children with and without suspected literacy problems according to school records and evaluations. From these presented criteria the respective schools selected participants. By informed consent from the parents and the children themselves the established L2 group consisted of 20 children who did not have Norwegian or any other Scandinavian language as their home language The school records and evaluations were not known to the project leaders or testers.

2.1.2 | L1 participants

For comparison with L2 children, data from three formerly published project were used. Permission from the Norwegian Centre for Research Data (NSD) was given to draw data from these projects: (1) 'the Bergen Longitudinal Dyslexia Project' (Clark et al., 2014; Helland et al., 2011; Helland et al., 2021; Specht et al., 2009); (2) 'Narrative writing in primary grade children' (Torkildsen et al., 2016), and (3) 'Auditive training effects from a dichotic listening app in children with dyslexia' (Helland et al., 2018). All projects were approved by the Regional Ethics Committee for Medical Research (REK) and the NSD. L1 participants consisted of a control group (L1-Con, $n = 95$) and a dyslexia group (L1-Dys, $n = 29$). Dyslexia was identified by professionals and based on the BDA (2007) definition and the causal model by Morton and Frith (1995). All studies were conducted in accordance with the Declaration of Helsinki.

2.2 | Background information

All parents or caregivers were asked to fill out a questionnaire, based on the original screening instrument 'Risk Index for age 5' (RI-5) designed to identify preschool children at risk of developmental dyslexia (for descriptive information and measures of reliability and validity, see Helland, 2015; Helland et al., 2021; Helland et al., 2011). The questionnaire collects information from caregivers along the same lines as what is discussed in clinical dyslexia assessments. It addresses the child's general health, motor and language development, special needs education, and incidence of language problems, dyslexia, or mathematic difficulties in the biological family. In sum, the questionnaire collects information at all four levels. A low score indicates low risk with different cut-off scores for boys (17.0) and girls (9.7). For this group of children, age adjusted questions for motor skills were used, but was otherwise kept in its original form. Hence it was denoted RI-5/8 in parallel to its use in Helland et al. (2021). The questionnaire consists of statements with response alternatives ('yes', 'no', 'do not know'). It is kept in an easily readable format and language. The L2 parents were encouraged to ask for help to fill out the questionnaire if needed. 16 parents from L1-Con and one from L2 had not filled out the RI-5/8. Table 1 gives an overview of the RI-5/8 scores, age, gender, and handedness by group and overall.

The RI-5/8 questionnaires were filled out individually by parents. It is unclear to what degree the L2 parents received any help with translation or interpretation of the text, and therefore the RI-5/8 scores for the L2 group

TABLE 1 Age, gender hand preference, risk index by groups.

Groups	L2, $N = 20$	L1-con, $N = 95$	L1-Dys, $N = 29$	ALL, $N = 144$
RI-5/8 (SD)				
All	7.2 (6.6)	10.3 (9.4)	19.2 (13.7)	11.9 (11.0)
M	6.7 (6.4)	11.1 (10.4)	24.4 (11.6)	13.1 (11.7)
F	7.9 (7.2)	9.0 (7.5)	13.7 (14.08)	10.1 (9.7)
Age, years (SD)				
	8.5 (0.8)	8.4 (0.3)	8.7 (0.3)	8.5 (0.3)
Gender				
M	12 (60%)	57 (60%)	15 (52%)	84 (58%
F	8 (40%)	38 (40%)	14 (48%)	60 (42%)
Hand				
R	18 (90%)	78(82%)	27 (93%)	123 (85%)
L	2 (10%)	17 (18%)	2 (7%)	21 (15%)

Note: RI-5/8: Risk Index screening for risk of dyslexia ages 5 and 8. Missing data 15 RI-5/8 from CON and 1 from L2. LSD Test showed significant higher age and RI-8 score in DYS versus CON and L2 ($p < 0.01$ and $p < 0.001$, respectively).

should be interpreted with caution. For the L2 group additional information was collected regarding home language, L2 (Norwegian) support in class, and the educational level of the parents. The additional questions for the L2 parents showed the following:

1. Home language: Arabic (5), English (1), Hungarian (1), Kurdish (1), Moroccan (1) Polish (1), Serbian (1), Slovakian (1), Somali (5), Spanish (1), Tamil (1), unknown (1)
2. Years of schooling: Fathers (16 responses): mean 13.5; range 6–19; SD 3.78. Mothers (17 responses): mean 12.71; range 2–21; SD 4.93. T-test for independent samples assessing gender difference in educational level was non-significant. Also, there was no correlation between parents' educational level and the RI-8 scores, or the children's reading and spelling scores.
3. L2 (Norwegian) support in class was reported in 7 L2 children, not received by 9 and not reported in 4 children. T-test showed no differences in RI-8 score between received support (mean 9.18, SD 7.67) and not received support (6.73, SD 6.45).
4. Heredity (familial incidence of language, dyslexia, or mathematic problems) was reported in one of the L2 participants.

2.3 | Procedures

For all projects testing was done individually in separate rooms by professionals (psychologists, speech-language therapists and master students in logopaedics trained and supervised by the project leaders).

2.4 | Tests

The test battery is based on the causal model by Morton and Frith (1995), on definitions of dyslexia (BDA, 2007; Peterson & Pennington, 2015). It is applied in several Norwegian dyslexia studies assessing children 10–15 years old with Norwegian as L1 (Helland et al., 2018; Helland & Asbjørnsen, 2000, 2001, 2003, 2004). and English as L2 (Helland & Morken, 2016). Measures of validity and reliability are reported in the listed references for all tests.

2.4.1 | Symptomatic level

Single word reading and spelling was assessed by the 'Standardisert test i avkoding og staving' (STAS) [Standardized test in decoding and spelling] (Klinkenberg & Skaar, 2001). The *reading* test (Read) consists of four lists of 85 words each to be read aloud. The lists are a mix of high and low frequency real words with phonetic and non-phonetic spellings (please note the earlier paragraph on Norwegian orthography). For each of the lists, there was a time limit of 40 seconds. One point was given for each correctly read word. A compound score comprising all four lists was used for analyses. The norm score for 2nd grade is 81 (SD 48), and 3rd grade is 99 (SD 55). The *spelling* test (Spell) consists of 79 real words, all to be written to dictation. The list is again a mix of high and low frequency words with phonetic and non-phonetic spelling. One point was given for each correctly spelled word. The norm score for 2nd grade is 19 (SD 9) is and 3rd grade is 22 (SD 9).

2.4.2 | Cognitive level

Domain-specific tests

Vocabulary (Voc) was assessed by the British Picture Vocabulary Scale (BPVS; Dunn et al., 2003), which is a receptive vocabulary test. The Norwegian version contains 12 sets of 12 tasks each. For each item, the test administrator says a word, and the subject responds by selecting the picture (from an array of four) that best illustrates its meaning. Raw scores were used, with 144 as a maximum score. One point was given for each correct response. The norm for age 8:0 to 8:11 is 95.52 (SD 11.77).

Sentence comprehension (Comp) was assessed by the Norwegian version of The Test for Reception of Grammar—Version-2 (TROG-2; Bishop, 2009). This is a receptive language test containing 20 blocks, each representing a grammatical construction. Each block has four items – altogether 80 items. The blocks are arranged in order of increasing difficulty. For each item, the administrator reads a sentence, and the subject points to the picture (in an array of four) that correctly depicts the sentence. Raw scores for each correct response were used, with a mean score for age 8:0 to 8:11 $=$ 99.9 (SD 14.16).

Domain-general tests

Short-term memory (STM) and *working memory* (WM) were assessed by the Digit Span task from the Wechsler Intelligence Scale for Children—Third Edition (Wechsler, 2002). The tasks were administered and scored according to test instructions. Raw scores from each test (forward recall for STM and backward recall for WM) were used. 1 point was given for each correct response. For comparison, raw score data for this age group was retrieved from Gardner (1981): STM age 8: 6–8:11: boys 6.17 (SD 2.32); girls 6.61 (SD 1.80). WM: boys 3.85 (SD 1.22); girls 4.02 (SD 1.39).

Rapid naming (RAN) was assessed using the colour/word naming test from the Stroop test battery (Hugdahl, undated; Golden & Freshwater, 1978; Stroop, 1935). In this test, the participants were shown a sheet with $6 \times 8 = 48$ dots in four different colours (red, blue, green, yellow). The task was to name the colours as quickly and accurately as possible, and a stopwatch was used to report the amount of time (in seconds) to name the colours of all dots on the plate. To make sure that the child knows the colour names and the reading direction from left to right, the test is started with an example trial. There are no Norwegian age norms available for this task, but data from our own laboratory on 119 typical children ages 7–8 show a mean (SD) score of 52.6 (18.0) seconds.

Dichotic listening (DL; Hugdahl, 2011) is a non-invasive method for the study of hemispheric dominance for speech perception. The DL test applied here is a standardized consonant/vowel listening task where the subjects use earphones to listen to 36 stimuli combinations of CV-syllables: /ba/, /da/, /ga/, /pa/, /ta/, /ka/ presented simultaneously to both ears. One syllable is played to the right ear and the other played to the left ear (Bless et al., 2014). The typical finding across genders and languages is the so-called right ear advantage (REA) meaning that more correct responses are given from stimuli to the right ear compared to stimuli to the left ear (Bless et al., 2015; Hirnstein et al., 2013). Studies have shown a general increase in the DL scores by age (Hirnstein et al., 2013) indicating an interaction with other developmental factors.

The test includes three conditions. The non-forced (NF) condition is foremost an assessment of speech perception, while the forced left (FL) and forced right (FR) conditions tap cognitive processes as attention (Hugdahl, 2011). In this study only the NF condition score was used. DL Re is the number of correctly reported stimuli given to the right ear, and DL Le is the number of correctly reported stimuli given to the left ear. For the NF condition the instruction was to immediately say or tick the syllable that was 'most audible'. A response was counted as correct when it matched the syllable presented to either the right or left ear on each trial. For the raw scores, the maximum score was 30 points. Data from 'The Bergen dichotic listening database' of 53 typical 2. and 3. graders show a mean NF Re score of 12.62 (SD 3.10) and a NF Le score of 10.25 (SD 2.48). See also Hirnstein et al. (2013).

Visuo-spatial skills (VS) were assessed using the Rey–Osterrieth Complex Figure Test (RO; Meyers & Meyers, 1995). with the Copy (VS1) and Recall (VS2) subtests. The subjects were given a blank piece of paper and a pencil and were first asked to copy a complex line drawing figure put in front of them. After 20 min with other activities, they were asked to redraw the figure from memory. Not only visuo-spatial skills are involved in test performance, but also other cognitive functions, such as memory, attention, planning, WM, and executive functions (Franceschini et al., 2022; Watanabe et al., 2005). Each drawing is scored for accurate reproduction and placement of 18 specific design elements, with a maximum score of 36 points for each task. No Norwegian norm data are available. However, a Norwegian study of 29 typical 8-year-old L1 children showed a mean score of 16.91 (SD 6.14) on VS1 and 9.78 (SD 4.22) on VS2 (Helland & Morken, 2016).

3 | DATA ANALYSES AND RESULTS

Data analyses were done in two parts. The analyses in Part 1 compares the L2 scores with the L1-Con and L1-Dys scores and assesses the relationship between the literacy scores and the cognitive scores. Part 2 assesses variations in scores within the L2 group.

3.1 | Data analyses, part 1

To assess group differences one-way ANOVA was used with the design Group (3: L2, L1-Con, L1-Dys) by each of the test scores (12: RI-5/8, Read, Spell, Voc, Comp, STM, WM, RAN, DL Re, DL Le, VS1, VS2) followed by LSD planned comparison. Cohen's *d* was used as a measure of effect size. A right ear advantage (REA) is the typical response pattern in the DL NF condition (Bless et al., 2015). Preliminary analyses showed no effects in the DL Le condition, and therefore this condition was excluded in the further analyses.

To assess the relationship between the literacy scores (Read, Spell) and the eight cognitive scores two partial correlations (Product–Moment) were executed. First, the relationship combining L1-Con and L1-Dys was assessed, and second, the relationship when combining L1-Con and L2 was assessed.

To be able to compare group profiles all scores were transformed to *z-scores* using the Con scores as baseline (Cooksey, 2020, pp 121–134). Three GLM Repeated Measures analyses were performed. First, the literacy variables with the design Group (3: L2, L1-Con; L1-Dys) by Tests (2: Read, Spell); second, the domain-specific variables with the design Group (3: L2, L1-Con; L1-Dys) by Tests (2: Voc, Comp); third, the domain-general variables with the design Group (3: L2, L1-Con; L1-Dys) by Tests (6: STM, WM, RAN, DL Re, VS1, VS2). Fisher LSD test was used to follow up. Tentative analyses with and without mean substitution of missing data in Voc and Comp in Dys showed no difference, and therefore the substitution version was used.

In all Part 1 analyses the alpha level was set at $p < 0.05$, and Cohen's *d* effect sizes were interpreted as 0.20 Small; 0.50 Medium; 0.80 Large (Fan & Konold, 2010).

3.2 | Results, part 1

Table 2 shows the one-way ANOVA, followed up by LSD post hoc test and Cohen's *d* for effect size analyses. Except for WM all analyses showed significant variations, with F-values from 6.5 to 35.2 and *p*-values larger than 0.002. LSD follow-up test showed the following:

3.2.1 | Environmental level

RI-5/8

No significant difference was seen between L2 and L1-Con, but the lower score in L2 showed a small effect size. L1-Dys showed significantly higher risk scores on the RI-5/8 compared to the other two groups 2.

3.2.2 | Symptomatic level

Read and spell

Both L2 and L1-Dys scored significantly lower than L1-Con with large effect sizes. The lower scores in L1-Dys versus L2 did not reach significance but showed a medium effect size.

3.2.3 | Cognitive level

Domain specific scores

Voc: L2 showed significantly lower scores compared to L1-Con ($p < 0.05$) and L1-Dys ($p < 0.001$), with large effect size. No difference was seen in L1-Con versus L1-Dys, however with a tendency L1-Con > L1-Dys with a small effect size.

Comp: Con scored significantly higher than both L2 ($p < 0.05$) and L1-Dys ($p < 0.001$) with a medium to large effect size. No significant difference between L2 and L1-Dys was seen, however with a tendency towards L2 > L1-Dys with a large effect size.

Domain general scores

STM: Both L1-Dys and L2 scored significantly lower than L1-Con ($p = 0.05$ and 0.001, respectively) with medium to large effect sizes, and with a tendency towards L2 > L1-Dys with a small effect size.

WM: as mentioned above there were no significant differences. However, there was a tendency towards L2 > L1-Dys with a small effect size.

RAN: L2 and L1-Dys used significantly longer time ($p = 0.05$, and 0.01, respectively) compared to L1-Con, both with medium effect sizes. No difference was seen between L2 and L1-Dys.

DL Re: L2 showed significantly higher scores ($p < 0.01$) compared to L1-Con and L1-Dys with medium to high effect sizes, however with a small effect size in L1-Con > L1-Dys.

VS1: L2 showed significantly higher scores ($p < 0.01$) compared to L1-Con and L1-Dys with medium effect sizes.

TABLE 2 Scores, group comparisons.

	Mean (SD) by groups						One-way ANOVA		LSD follow up test			Cohen's *d*		
Level test	L2 mean	SD	L1-con mean	SD	L1-Dys mean	SD	Df	F	L2 vs. L1-con	L1-con vs. L1-Dys	L2 vs. L1-Dys	L2 vs. L1-con	L1-con vs. L1-Dys	L2 vs. L1-Dys
ENV														
RI-5/8	7.20	6.59	10.34	9.42	19.22	13.74	125	10.37		***	***	0.386	0.754	1.116
n	19		80		29									
SYMP														
Read	92.50	48.38	131.82	47.55	68.77	20.97	80	12.839	*	***		0.864	1.716	0.697
Spell	20.35	7.49	27.24	7.28	16.31	6.78	80	14.805	***	****		0.933	1.554	0.566
n	20		50		13									
COGN														
dom-spec														
Voc	64.55	15.85	92.31	13.30	87.92	9.19	121	35.213	****		****	1.897	0.374	1.804
Comp	78.70	13.27	88.59	20.06	67.92	6.56	121	8.693	*	***		0.582	1.385	1.030
n	20		91		13									
dom-gen														
STM	6.30	1.53	7.25	1.76	5.93	1.09	137	8.539	*	***		0.576	0.902	0.279
WM	3.45	1.28	3.69	1.21	3.39	0.99	137	0.825				0.193	0.271	0.052
RAN	59.53	17.12	50.50	16.51	61.86	17.30	138	6.286	*	**		0.537	0.672	0.136
DL Re	14.55	3.47	11.88	3.47	11.03	2.65	137	7.132	**		***	0.769	0.275	1.140
VS1	22.48	8.26	16.92	6.47	15.62	7.47	138	6.498	**		***	0.749	0.186	0.871
VS2	14.93	7.69	8.30	4.86	7.38	4.85	138	14.425	****		****	1.031	0.189	1.174
n	20		92		29									

Abbreviations: COGN, cognitive; dom-gen, domain-general; dom-spec, domain-specific; ENV, environmental; SYMP, symptomatic.
*$p \le 0.05$.
**$p \le 0.01$.
***$p \le 0.001$.
****$p \le 0.0001$.

TABLE 3 Correlations between symptom and cognitive scores.

Level	Symptomatic	Cognitive							
Groups	1) Literacy	2) Domain-specific		3) Domain-general					
	Tests	Voc	Comp	STM	WM	RAN	DL Re	VS1	VS2
L1-Con and L1-Dys	Read	0.123	0.049	0.315*	0.207	−0.297*	0.263*	0.249	0.178
	Spell	0.259*	0.100	0.433***	0.251	−0.246	0.084	0.449***	0.404**
L1-Con and L2	Read	0.278*	−0.090	0.175	0.299*	−0.301*	0.093	0.072	0.076
	Spell	0.478***	−0.015	0.301*	0.233	−0.269*	−0.109	0.197	0.149

Note: Abbreviations and *p*-values as in Table 2.

VS2: L2 showed significantly higher scores ($p < 0.001$) compared to L1-Con and L1-Dys with medium to high effect sizes.

Table 3 shows correlations between the symptom (Read, Spell) scores and the cognitive scores. As to the combined L1-Con and L2 group there was a significant correlation in Read versus Voc, WM and RAN ($p < 0.05$), and in Spell versus Voc ($p < 0.001$) and STM and RAN ($p < 0.05$). The combined L1-Con and L1-Dys group showed a relationship between Read and STM, RAN, DL Re and VS1 ($p < 0.05$), and between Spell and Voc ($p < 0.05$), STM, VS1 and VS2 ($p < 0.001$).

Figure 3 shows the profiles of the *z*-scores of the three groups (L2, L1-Con, L1-Dys) with the scores of L1-Con as baseline. Three separate analyses were performed for (1) the literacy variables; (2) the domain-specific variables; (3) the domain-general variables.

Literacy (Read, Spell). Repeated measures showed an effect of Group, $F_{2,80} = 17.77$, $p < 0.0001$), but no effect of repeated measures or interaction. The effect of Group was due to lower scores in both L1-Dys and L2 compared to L1-Con ($p < 0.0001$) with no difference between L1-Dys and L2.

Domain-specific tasks (Voc, Comp). Repeated measures showed an effect of Group, $F_{2,121} = 26.95$, $p < 0.0001$, of Repeated measures, $F_{1,121} = 926$, $p < 0.05$, and of Interaction, $F_{2,121} = 17.95$, $p < 0.0001$. The effect of Group was that both L1-Dys and L2 showed significantly lower scores than L1-Con ($p = 0.002$ and $p < 0.0001$, respectively) and with a difference between L1-Dys and L2 ($p = 0.02$). The effect of Repeated measures did not reach significance ($p = 0.09$). The effect of interaction within each group showed no difference between the two tests in L1-Con, significantly higher scores in Voc versus Comp in L1-Dys, and significantly lower scores in Voc versus Comp in L2. Further, between groups L2 Voc was significantly lower than the three other scores in L1-Con and L1-Dys, and L2 Comp was significantly lower compared to L1-Con ($p < 0.05$), but with no difference to L1-Dys.

Domain-general tasks (STM, WM, RAN, DL Re, VS1, VS2). Repeated measures showed an effect of Group, $F_{2,136} = 7.897$, $p = 0.001$ of Repeated measures, $F_{5,680} = 13.243$, $p < 0.0001$, and of Interaction, $F_{10,680} = 6.851$, $p < 0.0001$. The effect of Group was due to significantly higher scores in L1-Con than in L1-Dys ($p = 0.005$), and that both L1-Con and L1-Dys had significantly lower scores than L2 ($p = 0.04$ and $p = 0.000$, respectively). The effect of Repeated measures was that STM and RAN were significantly lower than the DL Re, VS1 and VS2 ($p < 0.01$), and WM significantly lower than VS2 ($p = 0.03$). As to Interaction the main overall picture was that the L2 scores in DL Re, VS1 and VS2 were significantly higher than all scores in L1-Con and L1-Dys ($p < 0.01$). Within-group analyses showed: In L1-Con, no within-group differences; in L1-Dys, significantly lower STM than WM, DL Re, VS1 and VS2 ($p < 0.05$); in L2, significantly lower scores in STM, WM, RAN than DL Re, VS1, VS2 ($p < 0.001$) and significantly lower scores in DL Re versus VS2 ($p = 0.004$).

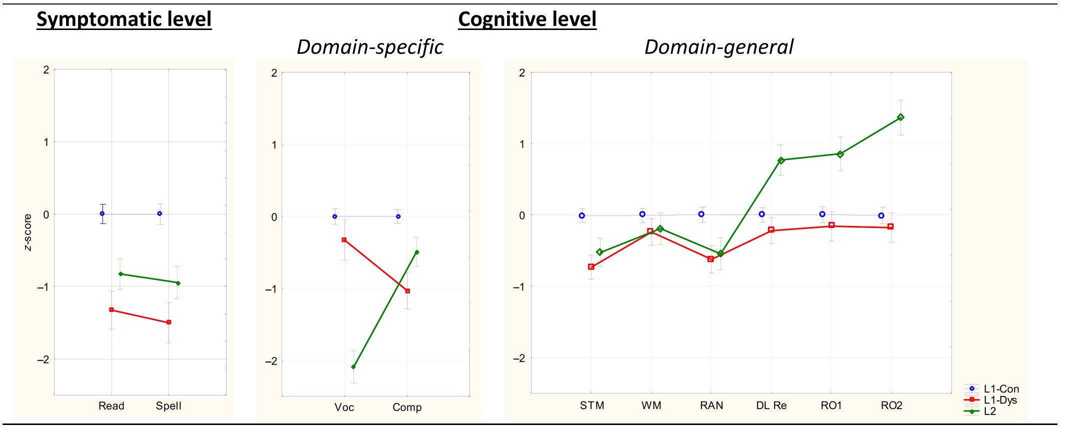

FIGURE 3 *Z*-scores by levels and groups.

3.3 | Data analyses, part 2

The *z*-scores were used in case evaluations and tentative subgrouping of the L2 scores. Based on the BDA dyslexia definition (BDA, 2007), the multi-factorial characteristics of dyslexia (Helland, 2022b; Pennington, 2006; Pennington & Bishop, 2009; Peterson & Pennington, 2015) and the findings in Part 1, individual cases were assessed using the domain-general scores only, searching for scores of − 0.5 SD or below compared to Control *z*-scores (baseline). After evaluations of all individual scores three subgroups were identified. Subgroup L2/A: no or only one *z*-score at − 0.5 SD; subgroup L2/B: one *z*-score at − 1 SD and one or more *z*-scores at − 0.5; subgroup L2/C: two or more *z*-scores at − 1 SD. Due to small groups Part 2 was not subjected to any statistical analyses.

3.4 | Results, part 2

Table 4 shows the outcome of the L2 subgrouping (L2/A, L2/B, L2/C) as to the information given by the parents; RI-5/8 scores, counts of RI-5/8 above cut-off, support in class, home language and biological information (gender, heredity, handedness).

As the table shows, the biological information on gender, heredity and handedness does not give any strong associations to dyslexia in any of the subgroups. However, a tendency is seen in L2/C: the RI-5/8 is above cut-off in three cases, four of the five participants are boys, which gives associations to earlier research pointing to a gender difference (Berninger et al., 2008a; Snowling et al., 2015; Zambrana et al., 2015), and two are left-handed, which is in line with research on left-handedness in dyslexia (Bishop & Bates, 2020; McManus, 2019).

Test scores by L2 subgroups of all 10 tests are shown in Table 5. In general, the scores show a gradual decrease: L2/A > L2/B > L2/C not only in the domain-general scores, which were used for subgroup identification, in both levels and domains. Compared with the group scores shown in Table 2, the L2/A-scores are close to the L1-Con-scores, and the L2/C-scores are close to the L1-Dys-scores with the L2/B-scores in a mid-position.

Figure 4 compared to Figure 3 illustrates this tendency. In accordance with the profile of L1-Con in Figure 3 the zero *z*-score is used as baseline in Figure 4. Distinct differences to the profiles of L1-Con and L1-Dys are shown in the domain-specific scores with lower scores in Voc versus Comp in all three subgroups.

TABLE 4 Tentative L2 subgrouping from domain-general scores, information on environmental and biological levels.

L2, $N = 20$				Environmental level		Biological level		
Subgroup	Criteria score >0.5 SD	RI-5/8 mean (SD) range	RI-5/8 > cut-off, *n*	Support in class, *n*	Home language (*n*)	Gender M/F	Heredity	Hand Right/left
L2/A, $n = 6$	One or none below 0.5 SD	6.12 (6.09) 0–15.4	1	0	Slovenian (1) Tamil (1) Somali (1) Polish (1) Moroccan (1) Spanish (1)	3/3	0	6/0
L2/B, $n = 9$	One below 1 SD, one or more below 0.5 SD	4.51 (3.73) 0–9.0	0	5	English (1) Hungarian (1) Kurdish (1) Arab (3) Somali (2) Arab(1)	5/4	1	9/0
L2/C, $n = 5$	Two or more below 1 SD	14.86 (7.70) 3.6–20.7	3	2	Serbian (1) Somali (2) Arab (1) unknown (1)	4/1	0	3/2

TABLE 5 Scores by L2 subgroups.

Level	Symptomatic		Cognitive							
	Literacy		Domain-specific		Domain-general					
Groups	Read	Spell	Voc	Comp	STM	WM	RAN	DL Re	VS1	VS2
L2/A	127 (65.0)	23.0 (5.6)	70.5 (13.5)	81.5 (13.4)	6.17 (1.5)	4.83 (1.3)	47.2 (9.3)	15.0 (3.6)	25.6 (7.4)	19.8 (8.8)
L2/B	83 (28.7)	20.8 (6.7)	68.0 (11.5)	80.9 (14.3)	6.78 (1.6)	3.00 (0.5)	60.3 (11.3)	14.1 (4.4)	22.6 (6.9)	13.5 (6.0)
L2/C	69 (39.1)	16.4 (10.3)	51.2 (19.8)	71.4 (10.6)	5.60 (1.5)	2.60 (0.9)	73.0 (23.9)	14.8 (1.5)	18.6 (11.3)	11.7 (7.7)

Note: Test scores, mean (SD) by L2 subgroups. Abbreviations as in Table 1.

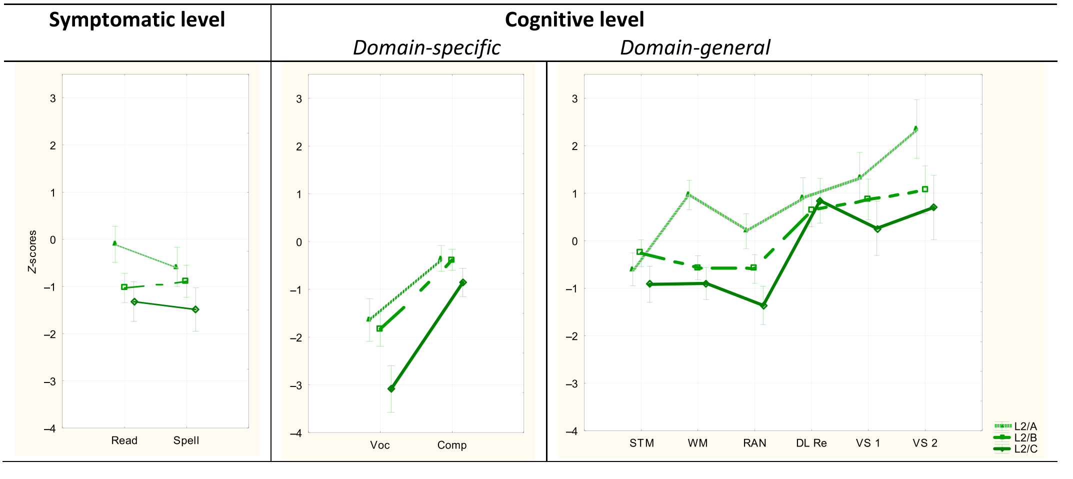

FIGURE 4 *Z*-scores, domain general scores by L2 subgroups.

Further, two striking subgroup differences are seen in the domain-general profiles. First, L2/A has a high *z*-score compared to the *z*-scores in L2/B and L2/C, which are below baseline. Second, the profiles in all three subgroups are above baseline the three *z*-scores DL Re, VS1 and VS2, however with a gradual decrease in VS1 and VS2: L2/A > L2/B > L2/C. In short, the profiles underline the order of strongest overall *z*-scores in L2/A, weakest in L2/C and a middle position in L2/C.

4 | DISCUSSION

This study focused on how to disentangle dyslexia in L2 from typical L2 problems in a group of children at the emergent literacy stage, all attending regular Norwegian primary schools. Dyslexia was defined as a multifactorial impairment (BDA, 2007; Helland, 2022a; McGrath et al., 2020; Morton & Frith, 1995; Peterson & Pennington, 2015) which resulted in a variety of dyslexia profiles. The combination of multilingualism and dyslexia adds a double layer to identification and diagnosis. Conducting research on L2 development can be done in a variety of ways, from addressing monolingual and monocultural cohorts to cohorts with a pluralistic background, as in the present study. By applying the causal four-level model by Morton and Frith (1995) the results indicate that dyslexia in L2 can be disentangled from typical L2 literacy problems in children at an emergent literacy stage. These findings should be considered tentative and should challenge further research on dyslexia in L2.

The assessments were done in two steps. First, by comparing data from three age-matched groups: an L1 control group (L2-Con), an L1 dyslexia group (L1-Dys), and a mixed group of L2 children; second, by case-wise inspection of the scores of the L2 group.

Since language and literacy scores are not seen as valid dyslexia indicators of L2 problems (Ramírez & Kuhl, 2016), variations within the cognitive level (cp Figure 3) were focused (Gray et al., 2019).

4.1 | Part 1

4.1.1 | Environmental level

Part 1 of the study assessed the relationships between the L2 group and the two L1 groups L1-Con and L1-Dys, all in the emergent literacy stage. All children had received literacy training in accordance with the national curriculum (The Norwegian Directorate for Education and Training, 2022) irrespective of their linguistic and cultural backgrounds. The collected information from the parents revealed few at-risk factors in L1-Con and L2. In contrast, L1-Dys had a high-risk score as defined by the RI-5/8. This does not necessarily mean that there were no risk-factors in the L2 group. Rather, one should keep in mind cultural and conceptual diversities, which demands a different focus and research design than was the case in the present study. However, individual assessments were further elaborated by testing at the symptomatic and cognitive levels.

4.1.2 | Symptomatic level

Assessments at the symptomatic level focused on word reading and spelling. As expected, and in line with other studies (Peterson & Pennington, 2015; von Hagen et al., 2020) the literacy scores in L2 were low compared to L1-Con, but comparable to L1-Dys. However, these symptoms may be understood in different ways in the two groups, and assessment at the cognitive level is needed for further identification (Frith, 1999).

4.1.3 | Cognitive level

According to the definition by BDA (2007), cognitive benchmarks of dyslexia are impairments within phonological awareness, processing speed, working memory (auditory, visual) or rapid naming that do not match the individual's other cognitive abilities. Comorbidity with DLD is often seen, where either language comprehension or production, or both, are affected (Pennington & Bishop, 2009). In this respect an important factor within this level is the continuum from high to low linguistic task demands. Therefore, in line with recent research, the cognitive level in this study was subdivided into domain-specific (high linguistic demands) and domain-general (low linguistic demands) factors (Hachmann et al., 2020; Liu & Xu, 2022; O'Brien et al., 2014).

Domain-specific skills were assessed by tests of vocabulary (Voc) and sentence comprehension (Comp). Although both L1-Dys and the L2 group showed lower scores within these two variables compared to L1-Con, the profiles of L1-Dys and L2 were unexpectedly different: lower scores on comprehension versus vocabulary in L1-Dys, and higher scores on comprehension versus vocabulary in L2. Sentence comprehension is more demanding on working memory than is single word comprehension, and in both dyslexia and DLD language comprehension difficulties are associated with impairments within short term and/or working memory (Pennington & Bishop, 2009). However, low vocabulary skills in L2 are in line with other studies (Fox et al., 2019; Karlsen et al., 2017). Since the comprehension scores were significantly higher than the vocabulary scores in L2, one may speculate that the L2 pattern of low vocabulary is compensated for when used in a context. In sum, this variability underlines earlier findings that the domain-specific scores cannot be seen as valid indicators of dyslexia in L2 (Bialystok et al., 2010; Luk & Bialystok, 2013). However, the significant correlations between literacy scores and vocabulary scores signal that vocabulary is an important factor in L2 literacy. Hence, the value of these findings is important in an L2 classroom setting. Vocabulary is essential to literacy and should be trained in context, and not as isolated words.

Domain-general skills were assessed by tests of short-term memory (STM), working memory (WM), rapid naming (RAN), dichotic listening (DL), visuo-spatial copy (VS1) and visuo-spatial recall (VS2). Combinations of low scores within some of these areas are typical dyslexia benchmarks (BDA, 2007; Hachmann et al., 2020; Helland & Asbjørnsen, 2000, 2001, 2003, 2004). Also, they require minimal language skills and should be equally demanding for the children in this study irrespective of their language background.

Again, an unexpected pattern emerged: the scores in dichotic listening (DL Re) and the two visuo-spatial tests (VS1, VS2) in L2 were significantly higher compared to both L1-Con and L1-Dys. As described earlier, several cognitive functions, as auditive and visual working memory, executive functions, attention, and brain lateralisation for language can be identified by these tests. In addition, these three tests relate to perception: DL in the non-forced condition assesses auditory perception (Hugdahl, 2009), and the visuo-spatial tests are both tests of visual perception and long term visual memory (Canham et al., 2000). A parallel finding is reported in hearing impaired individuals, showing cross-modal activations visually and auditorily in Cochlear Implants users (Chen et al., 2016). Hence, one may speculate that the high scores in L2 reflect a strategy to compensate for limited language and literacy skills. Also, it could be stipulated that rather than compensation, these are indicative of exposure to more than one language, consistent with advantages in, that is working memory, speech perception and phonological awareness, as discussed in Bialystok (2017).

In contrast to the L2 group, there were significant correlations in the combined L1-Con/L1-Dys group between these scores and the literacy scores. This indicates that a limited domain-general area, more specifically consisting of STM, WM and RAN, should be seen as valid dyslexia identifiers in L2, while a broader spectrum within the domain-general area is applicable in L1 dyslexia. It should be noted, however, that the effect of these factors on literacy may change by age (Helland & Morken, 2016).

In sum Part 1 of this study separated the variables in three main groups: variable group (1) emergent *literacy* (single word reading and spelling); variable group (2) *domain-specific* cognitive factors (vocabulary, sentence comprehension) and variable group (3) *domain-general* cognitive factors (STM, WM, RAN, DL Re, VS1, VS2). In accordance with recent research on multilingualism, variables within groups (1) and (2) are typically low in L2 children, while variables in group (1) and a variety within groups (2) and (3) are typically low in dyslexia. In L2 low literacy and domain-specific scores could be attributed to their language situation and are not necessarily signs of dyslexia. However, following the argument that strong auditory and visual perception may be compensatory for linguistic shortcomings in L2, the three domain-general variables STM, WM and RAN may be keys to disentangling typical L2 problems from signs of L2 dyslexia. This was further investigated in Part 2, which included the L2 group only.

4.2 | Part 2

As discussed earlier, *literacy scores* are typically low both in Dyslexia and in L2 at the emergent literacy stage and cannot be used for dyslexia identification alone. In L1 dyslexia, literacy and variations within cognitive domain-specific and domain-general scores constitute the diagnosis (BDA, 2007). Low L2 literacy scores are seen in many L2 populations (O'Brien et al., 2014), which was also the case within the L2 group in the present study.

The *domain-specific* scores (Voc, Comp) were expected to be low in L2. This is also often, but not always, seen in dyslexia, where low scores can be explained by DLD comorbidity (Pennington & Bishop, 2009). Due to these expected low literacy and domain-specific scores, only a variation of low *domain-general* scores (STM, WM, RAN, DL Re, VS1, VS2) could be seen as valid dyslexia identifiers in L2. Hence, the cognitive domain-general scores in L2 were tentatively used for single case analyses. By counting *z*-scores below −0.5 and −1.0 standard deviations below basic, three L2 profiles emerged, L2/A, L2/B and L2/C. In the *domain-general* variables the pattern was again higher scores in subgroup L2/A and lower scores in subgroups L2/B and L2/C, however with the unexpected pattern of higher scores in the three 'perceptual' areas (DL Re, VS1, VS2). By these findings, it seems reasonable to emphasize STM, WM and RAN as valid identifiers of dyslexia in L2 and using the perceptual scores for deeper insight and understanding of individual variations. Four L2 cases showed low perceptual scores, indicating that high scores within these variables are not a universal in L2, but that low scores could be understood as dyslexia benchmarks also in L2. Hence, age and development should always be considered when evaluating the independent auditory and visual modalities. The participants in the present study were still in the emergent literacy stage, and the two modalities develop at different scales across age and literacy development.

Applied to all tests, the following pattern emerged: The L2/A scores were close to the basic L1-Con scores, while the L2/C score were close to the L1-Dys scores, leaving the L2/B scores in a middle position. This was underlined by the test scores (cp. Table 5).

As Table 4 shows, there is little or nothing in the available background information (at-risk factors, support in class or home language) that is especially alarming or different from what can be expected from a typical L2 group. However, when looking at the subgroups, the number of RI-5/8 scores above cut-off is disproportionally large in subgroup L2/C, and the frequency of support in class is higher in L2/B and L2/C compared to L2/A. At the biological level, there are no alarming differences from norm as to gender distribution, heredity, or handedness. However, the gender distribution in L2/C is in line with findings that boys are either more affected (Snowling et al., 2015) or more easily identified by teachers (Peterson & Pennington, 2015; Shaywitz et al., 1990).

To sum up it is reasonable to conclude that subgroup L2/A has no sign of dyslexia, despite their low literacy and domain-specific scores. The main point is that their domain-general scores are comparable to norm. In contrast, subgroup L2/C showed a variety of low scores within the domain-general area, which are typical signs of dyslexia. Subgroup L2/B appeared to be in an in-between position, with a variety of domain-general scores slightly below baseline. This leaves two interpretations; that the L2/B-subgroup represents an intermediate position of either resolving or emerging problems (Helland et al., 2021; Snowling et al., 2015). To examine this further, an intensive individual training program with a pre/post-test design would reveal responsiveness to intervention. A positive response could indicate resolving problems, while a negative response could indicate emerging problems or dyslexia.

4.3 | Limitations and further research

For future studies a larger L2 group with more background information concerning years abroad, years in kindergarten, number of siblings, and home literacy would strengthen the validity of any L2-study. A substantial variation in home languages was seen in all three L2 subgroup in this study. Separate studies of participants from homogenous language groups may give more information as to the effect of language typology on L2 learning.

The L2 children in this study were given equal opportunities through an egalitarian school system where literacy training was integrated in line with their L1 school mates from day one at school. Although language typology did not seem to make any difference, valuable information as to L2 learning would be given by a closer focus on how L1 typology interferes with L2. This should include development in single word reading and spelling, text writing, pragmatics and discourse using dynamic assessment as a method (Orellana et al., 2019). Similar studies from other societies may find strengths and weaknesses that did not come forth in this study.

Due to the unexpected discrepancies within the domain-specific area (Voc < Comp) and the domain-general areas (STM, WM, RAN < DL Re, VS1, VS2) one may speculate if these functions could be explained by compensatory mechanisms. Unfamiliar words (as in Voc) can gain meaning when put into context (as in Comp), aided by attention to phonological and visual cues (as in DL Re, VS1 and VS2, respectively). Moreover, following-up discussions on executive functions in L2 (Diaz et al., 2021), would be in line with the findings of this study. The non-forced attention paradigm in dichotic listening is foremost a source of assessing language lateralisation in the brain, while the forced attention paradigm assesses attention and executive functions (Helland et al., 2018; Helland & Asbjørnsen, 2000). Hence, further studies could apply all three paradigms in dichotic listening. Likewise, the visuo-spatial test used in the present study has been used as measures of executive functions in L1 children with dyslexia (Helland & Asbjørnsen, 2003) and in children diagnosed with various neurological disorders (Watanabe et al., 2005).

Pursuing these aspects would give further insight into the relationship between the cognitive factors in L2 dyslexia. According to Bialystok (2017) bilingualism changes the way both verbal and nonverbal cognitive processing is carried out. Therefore, a longitudinal design following the children through the three literacy stages should give important insight into the development of the three L2 subgroups that emerged in the present study. Unfortunately, due to the covid-19 pandemic a reliable and valid follow-up study was not possible.

4.4 | Practical implications and concluding remarks

All tests in the present study are well known in dyslexia assessment. Since literacy scores and verbal, domain-specific scores in general are low in L2 in the emergent literacy stage, focus should be on the domain-general factors with scores below norm. As dyslexia is a multifactorial impairment, combinations of two or more low scores could indicate dyslexia also in L2. Our main findings were that the typical dyslexia benchmarks STM, WM and RAN are valid indicators of dyslexia in L2, while other dyslexia benchmarks are hard to separate from typical L2 literacy problems. This goes for low reading, writing, vocabulary, and comprehension skills. Three L2 profiles emerged: one with no signs of dyslexia, one with typical signs of dyslexia, and one in a middle position, indicating either resolving or emerging literacy problems. In sum we suggest that low literacy and language scores are not sufficient to identify dyslexia in L2, but multiple low scores within the domain-general cognitive area reflecting dyslexia benchmarks could be a reliable and valid way to identify dyslexia in L2.

ACKNOWLEDGEMENTS

We thank our colleague Professor Janne Torkildsen, Department of Special Needs Education, University of Oslo, Norway for her valuable contribution to this project.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Turid Helland https://orcid.org/0000-0002-4535-1870

REFERENCES

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

BDA. (2007). British dyslexia associtation: Definition of dyslexia. http://www.bdadyslexia.org.uk/about-dyslexia/faqs.html

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008a). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology*, *46*(2), 151–172. https://doi.org/10.1016/j.jsp.2007.02.007

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008b). Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology*, *46*(1), 1–21. https://doi.org/10.1016/j.jsp.2006.11.008

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233–262.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and cognition*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bishop, D. (2009). *Test for reception of grammar* (2nd ed. Norwegian version). Pearson Assessment.

Bishop, D. V. M., & Bates, T. C. (2020). Heritability of language laterality assessed by functional transcranial doppler ultrasound: A twin study [version 3; peer review: 3 approved]. *Wellcome Open Research*, *4*, 161. https://doi.org/10.12688/wellcomeopenres.15524.3

Bless, J. J., Westerhausen, R., Kompus, K., Gudmundsen, M., & Hugdahl, K. (2014). Self-supervised, mobile-application based cognitive training of auditory attention: A behavioral and fMRI evaluation. *Internet Interventions*, *1*(3), 102–110. https://doi.org/10.1016/j.invent.2014.06.001

Bless, J. J., Westerhausen, R., von Koss Torkildsen, J., Gudmundsen, M., Kompus, K., & Hugdahl, K. (2015). Laterality across languages: Results from a global dichotic listening study using a smartphone application. *Laterality: Asymmetries of Body, Brain and Cognition*, *20*(4), 434–452. https://doi.org/10.1080/1357650X.2014.997245

Brinchmann, E. I., Hjetland, H. N., & Lyster, S.-A. H. (2015). Lexical quality matters: Effects of word knowledge instruction on the language and literacy skills of third- and fourth-grade poor readers. *Reading Research Quarterly*, *51*, 165–180. https://doi.org/10.1002/rrq.128

Canham, R., Smith, S., & Tyrrell, A. (2000). Automated scoring of a neuropsychological test: The rey osterrieth complex figure (Vol. 2).

Chen, L.-C., Sandmann, P., Thorne, J. D., Bleichner, M. G., & Debener, S. (2016). Cross-modal functional reorganization of visual and auditory cortex in adult Cochlear implant users identified with fNIRS. *Neural Plasticity*, *2016*, 4382656.

Clark, K. A., Helland, T., Specht, K., Narr, K. L., Manis, F. R., Toga, A. W., & Hugdahl, K. (2014). Neuroanatomical precursors of dyslexia identified from pre-reading through to age 11. *Brain*, *137*(12), 3136–3141. https://doi.org/10.1093/brain/awu229

Cooke, A., & Adams, J. O. (2007). *The dyslexia handbook 2007/8/a compendium of articles, checklists and resources for dyslexic people, their families and teachers*. British Dyslexia Association.

Cooksey, R. W. (2020). Descriptive statistics for Summarising data. In R. W. Cooksey (Ed.), *Illustrating statistical procedures: Finding meaning in quantitative data* (pp. 61–139). Springer Singapore.

Dewey, G. (1971). *English spelling: Roadblock to reading*. Teachers College Press.

Diaz, V., Borjas, M., & Farrar, M. J. (2021). Is there an association between executive function and receptive vocabulary in bilingual children? A longitudinal examination. *Children*, *8*(1), 44.

Dunn, L. M., Dunn, D. M., & Styles, B. (2003). *British picture vocabulary scale* (3rd ed.). PsychCorp, Pearson Assessment.

Ehri, L. C. (1987). Learning to read and spell words. *Journal of Literacy Research*, *19*(1), 5–31. https://doi.org/10.1080/10862968709547585

Fan, X., & Konold, T. R. (2010). Statistical significance versus effect size. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 444–450). Elsevier.

Fox, R., Corretjer, O., & Webb, K. (2019). Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012–2019. *Foreign Language Annals*, *52*(4), 699–726.

Franceschini, S., Bertoni, S., Puccio, G., Gori, S., Termine, C., & Facoetti, A. (2022). Hidden in the labyrinth: Visuo-spatial attention deficit in children with developmental dyslexia. *Research Square*, preprint. https://doi.org/10.21203/rs.3.rs-1558371/v1

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, *36*, 69–81.

Frith, U. (1999). Paradoxes in the definition of dyslexia. *Dyslexia*, *5*(4), 192–214.

Gardner, R. A. (1981). Digits forward and digits backward as two separate tests: Normative data on 1567 school children. *Journal of Clinical Child Psychology*, *10*(2), 131–135. https://doi.org/10.1080/15374418109533032

Glascoe, F. P. (1997). Parents' concerns about Children's development: Prescreening technique or screening test? *Pediatrics*, *99*(4), 522–528. https://doi.org/10.1542/peds.99.4.522

Golden, C. J., & Freshwater, S. M. (1978). Stroop color and word test.

Gray, S., Fox Annie, B., Green, S., Alt, M., Hogan Tiffany, P., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Hachmann, W. M., Cashdollar, N., Postiglione, F., & Job, R. (2020). The relationship of domain-general serial order memory and reading ability in school children with and without dyslexia. *Journal of Experimental Child Psychology*, *193*, 104789. https://doi.org/10.1016/j.jecp.2019.104789

Hagtvet, B., Helland, T., & Lyster, S. A. H. (2006). Literacy acquisition in Norwegian. In *Handbook of orthography and literacy* (p. 840). Lawrence Erlbaum Associates.

Helland, T. (2015). RI-5. Dyslexia Risk Index; a questionnaire for parents and pre-school teachers.

Helland, T. (2022a). *Different definitions of dyslexia*. Scholarly Community Encyclopedia. https://encyclopedia.pub/entry/33554

Helland, T. (2022b). Trends in dyslexia research during the period 1950 to 2020-theories, definitions, and publications. *Brain Sciences*, *12*(10), 1323. https://doi.org/10.3390/brainsci12101323

Helland, T., & Asbjørnsen, A. E. (2000). Executive functions in dyslexia. *Child Neuropsychology*, *6*(1), 37–48.

Helland, T., & Asbjørnsen, A. E. (2001). Brain asymmetry for language in dyslexic children. *Laterality: Asymmetries of Body, Brain and Cognition*, *6*(4), 289–301. https://doi.org/10.1080/713754422

Helland, T., & Asbjørnsen, A. E. (2003). Visual-sequential and Visuo-spatial skills in dyslexia: Variations according to language comprehension and mathematics skills. *Child Neuropsychology*, *9*(3), 208–220.

Helland, T., & Asbjørnsen, A. E. (2004). Digit span in dyslexia: Variations according to language comprehension and mathematics skills. *Journal of Clinical and Experimental Neuropsychology*, *26*(1), 31–42.

Helland, T., & Kaasa, R. (2005). Dyslexia in English as a second language. *Dyslexia*, *11*(1), 41–60. https://doi.org/10.1002/dys.286

Helland, T., & Morken, F. (2016). Neurocognitive development and predictors of L1 and L2 literacy skills in dyslexia: A longitudinal study of children 5–11 years old. *Dyslexia*, *22*(1), 3–26. https://doi.org/10.1002/dys.1515

Helland, T., Morken, F., Bless, J. J., Valderhaug, H. V., Eiken, M., Helland, W. A., & von Koss Torkildsen, J. (2018). Auditive training effects from a dichotic listening app in children with dyslexia. *Dyslexia*, *24*(4), 336–356. https://doi.org/10.1002/dys.1600

Helland, T., Morken, F., & Helland, W. A. (2021). Kindergarten screening tools filled out by parents and teachers targeting dyslexia. Predictions and developmental trajectories from age 5 to age 15 years. *Dyslexia*, *27*(4), 413–435.

Helland, T., Plante, E., & Hugdahl, K. (2011). Predicting dyslexia at age 11 from a risk index questionnaire at age 5. *Dyslexia*, *17*(3), 207–226. https://doi.org/10.1002/dys.432

Hirnstein, M., Westerhausen, R., Korsnes, M. S., & Hugdahl, K. (2013). Sex differences in language asymmetry are age-dependent and small: A large-scale, consonant–vowel dichotic listening study with behavioral and fMRI data. *Cortex*, *49*(7), 1910–1921. https://doi.org/10.1016/j.cortex.2012.08.002

Hofslundsengen, H., Gustafsson, J.-E., & Hagtvet, B. E. (2019). Contributions of the home literacy environment and underlying language skills to preschool invented writing. *Scandinavian Journal of Educational Research*, *63*(5), 653–669.

Hugdahl, K. (2009). Dichotic listening studies of brain asymmetry. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (pp. 517–522). Academic Press.

Hugdahl, K. (2011). Fifty years of dichotic listening research – Still going and going and…. *Brain and Cognition*, *76*(2), 211–213. https://doi.org/10.1016/j.bandc.2011.03.006

Hugdahl, K. (undated). Stroop test, Colour/word naming. Norwegian version.

Karlsen, J., Lyster, S.-A. H., & Lervåg, A. (2017). Vocabulary development in Norwegian L1 and L2 learners in the kindergarten–school transition. *Journal of Child Language*, *44*(2), 402–426. https://doi.org/10.1017/S0305000916000106

Kim, K., & Helphenstine, D. T. (2017). The perils of multi-lingual students: "I'm not LD, I'm L2 or L3". *Journal of International Students*, *7*(2), 421–428.

Klinkenberg, J. E., & Skaar, E. (2001). *STAS. Standardisert test i avkoding og staving [Standardised test in decoding and spelling]*. Pedagogisk-psykologisk tjeneste.

Liu, D., & Xu, Z. (2022). How to understand dyslexia: From domain-specific to domain general factors. In *Supporting diverse students in Asian inclusive classrooms* (pp. 73–89). Routledge.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. Defining dyslexia, comorbidity, Teachers' knowledge of language and Reading. *Annals of Dyslexia*, *53*, 1–14.

McGrath, L. M., Peterson, R. L., & Pennington, B. F. (2020). The multiple deficit model: Progress, problems, and prospects. *Scientific Studies of Reading*, *24*(1), 7–13. https://doi.org/10.1080/10888438.2019.1706180

McManus, C. (2019). Half a century of handedness research: Myths, truths; fictions, facts; backwards, but mostly forwards. *Brain and Neuroscience Advances*, *3*, 2398212818820513. https://doi.org/10.1177/2398212818820513

Meyers, J. E., & Meyers, K. R. (1995). *Rey complex figure test and recognition trial*. Psychological Assessment Resources, Inc.

Miles, T. R. (2004). Some problems in determining the prevalence of dyslexia.

Morton, J., & Frith, U. (1995). Causal modeling: A structural approach to developmental psychopathology. In D. J. C. Dante Cicchetti (Ed.), *Developmental psychopathology, Vol. 1: Theory and methods. Wiley series on personality processes* (pp. 357–390). John Wiley & Sons.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and Reading fluency: Implications for understanding and treatment of Reading disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

O'Brien, B. A., Yin, B., Li, L., Zhang, D., Chin, C. F., Zhao, S., & Vaish, V. (2014). *Bilingualism, literacy and reading achievement* (NIE Working Paper Series No.4). National Institute of Education.

Ocklenburg, S., & Gunturkun, O. (2017). *The lateralized brain: The neuroscience and evolution of hemispheric asymmetries*. Academic Press.

Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology*, *28*(3), 1298–1317. https://doi.org/10.1044/2019_AJSLP-18-0202

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and Reading disorders. *Annual Review of Psychology*, *60*(1), 283–306. https://doi.org/10.1146/annurev.psych.60.110707.163548

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*(1), 283–307. https://doi.org/10.1146/annurev-clinpsy-032814-112842

Pugh, K., & Verhoeven, L. (2018). Introduction to this special issue: Dyslexia across languages and writing systems. *Scientific Studies of Reading*, *22*(1), 1–6. https://doi.org/10.1080/10888438.2017.1390668

Ramírez, N. F., & Kuhl, P. K. (2016). *Bilingual language learning in children*. Institute for Learning & Brain Sciences, University of Wsahington.

Saiegh-Haddad, E. (2019). What is phonological awareness in L2? *Journal of Neurolinguistics*, *50*, 17–27. https://doi.org/10.1016/j.jneuroling.2017.11.001

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of Reading disability in boys and girls: Results of the Connecticut longitudinal study. *Journal of the American Medical Association*, *264*(8), 998–1002.

Snowling, M., Duff, F. J., Nash, H. M., & Hulme, C. (2015). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*, *57*, 1360–1369. https://doi.org/10.1111/jcpp.12497

Specht, K., Hugdahl, K., Ofte, S. H., Nygård, M., Bjørnerud, A., Plante, E., & Helland, T. (2009). Brain activation reveals at-risk for dyslexia in 6-year old children. *Scandinavian Journal of Psychology*, *50*, 79–91. https://doi.org/10.1111/j.1467-9450.2008.00688.x

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

The Council of Europe. (2007). From linguistic diversity to plurilingual education: Guide for the development of language education policies in Europe. http://www.coe.int/lang

The Norwegian Directorate for Education and Training. (2022). The education System in Norway. From Early Childhood to Young Adults. Information for newly arrived parents and guardians. udir.no: Ministry of Education.

Torkildsen, J., Morken, F., Helland, W. A., & Helland, T. (2016). The dynamics of narrative writing in primary grade children: Writing process factors predict story quality. *Reading and Writing*, *29*, 529–554. https://doi.org/10.1007/s11145-015-9618-4

van Setten, E. R. H., Tops, W., Hakvoort, B. E., van der Leij, A., Maurits, N. M., & Maassen, B. A. M. (2017). L1 and L2 reading skills in Dutch adolescents with a familial risk of dyslexia. *PeerJ*, *5*, e3895. https://doi.org/10.7717/peerj.3895 von Hagen, A., Kohnen, S., & Stadie, N. (2020). Foreign language attainment of children/adolescents with poor literacy skills: A systematic review and meta-analysis. *Educational Psychology Review*, *33*, 459–488. https://doi.org/10.1007/s10648-020-09566-6

Watanabe, K., Ogino, T., Nakano, K., Hattori, J., Kado, Y., Sanada, S., & Ohtsuka, Y. (2005). The Rey–Osterrieth complex figure as a measure of executive function in childhood. *Brain and Development*, *27*(8), 564–569.

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence™ – third edition (WPPSI™ – III). Norwegian adjustmeny*. Pearson.

Ylinen, S., Junttila, K., Laasonen, M., Iverson, P., Ahonen, L., & Kujala, T. (2019). Diminished brain responses to second-language words are linked with native-language literacy skills in dyslexia. *Neuropsychologia*, *122*, 105–115. https://doi.org/10.1016/j.neuropsychologia.2018.11.005

Zambrana, I. M., Dearing, E., Nærde, A., & Zachrisson, H. D. (2015). Time in early childhood education and care and language competence in Norwegian four-year-old girls and boys. *European Early Childhood Education Research Journal*, *1-14*, 793–806. https://doi.org/10.1080/1350293X.2015.1035538

How to cite this article: Helland, T., Morken, F., & Helland, W. A. (2023). Disentangling dyslexia from typical L2-learning in emergent literacy. *Dyslexia*, 1–22. https://doi.org/10.1002/dys.1753