



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/236177/>

Version: Published Version

Article:

Nash, H., Dixon, C., Clarke, P. et al. (2026) Predicting reading outcomes in school starters using a computer assisted dynamic assessment of decoding. Reading and Writing. ISSN: 0922-4777

<https://doi.org/10.1007/s11145-025-10758-1>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Predicting reading outcomes in school starters using a computer assisted dynamic assessment of decoding

Hannah Nash¹ · Chris Dixon¹ · Paula Clarke² · Emily Oxley¹ · Anna Weighall³ · Anna Steenberg Gellert⁴

Received: 11 June 2025 / Accepted: 22 December 2025
© The Author(s) 2026

Abstract

Early and accurate identification of children at risk of reading difficulties is essential for ensuring timely and appropriate educational support. Traditional screening tools are static, measuring only current knowledge. As a result, poor performance may reflect either a specific learning difficulty such as dyslexia or limited prior learning opportunities, making accurate identification challenging. This study examined the potential of a computer-assisted Dynamic Assessment (DA) of decoding, which measures learning potential through responsiveness to training, as an alternative approach. In a longitudinal study, 317 reception-aged children (mean age 5 years 3 months) completed the DA of decoding along with static assessments of reading, related skills, and non-verbal ability. Ten months later, 286 of these children were reassessed on the measures of word reading. The DA of decoding accounted for significant unique variance in later word reading and demonstrated excellent classification accuracy in identifying children at risk of dyslexia. For children with English as an Additional Language (EAL), combining the DA with static measures resulted in perfect classification accuracy. Further analysis of a sub-sample of non-readers suggests that administering the DA earlier in the school year could increase its predictive value for all learners. These findings support the use of DA of decoding as a promising early screening tool, particularly for culturally and linguistically diverse children who may be misidentified by traditional static assessments.

Keywords Dynamic assessment · Decoding · Word reading · Dyslexia

✉ Hannah Nash
h.nash@leeds.ac.uk

¹ School of Psychology, University of Leeds, Leeds, UK

² School of Education, University of Leeds, Leeds, UK

³ School of Education, University of Sheffield, Sheffield, UK

⁴ Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark

Reading proficiency is closely linked to later life chances, including employment, health and economic and social disadvantage (Morrisroe, 2014), yet international data from 81 countries reveals declining reading standards since 2012, with approximately 26% of young people leaving school as poor readers (OECD, 2023). Evidently, a substantial proportion of children require additional reading support, ideally provided as early as possible (Lovett et al., 2017). Measures of Phonological Awareness (PA) and Letter Knowledge (LK) reliably predict early word-reading performance across various orthographies (Caravolas et al., 2013; Puolakanaho et al., 2007), and screening based on these measures at school entry can achieve acceptable clinical accuracy (Thompson et al., 2015). However, such assessments are static, evaluating only the child's existing knowledge at a single point in time. Consequently, they can lack sensitivity to variations in children's learning environments (Peña & Halle, 2011), potentially confounding the causes underlying poor reading performance. Moreover, static measures frequently suffer from floor effects when used with young children (Catts et al., 2009).

Dynamic Assessment (DA) represents an alternative approach to screening, which may help distinguish between limited learning opportunities and intrinsic learning difficulties by evaluating a child's capacity to learn through feedback (Grigorenko & Sternberg, 1998). DA therefore has the potential to provide a more accurate assessment for children from diverse backgrounds who may have experienced fewer opportunities to develop the foundational language skills necessary for reading (Grigorenko, 2009). In this study, we developed a computer-assisted DA of decoding and investigated its effectiveness in predicting later reading achievement and identifying children at risk of word-reading difficulties within a diverse cohort of children starting their first year of formal education in England.

The cognitive capacities underlying word recognition are detailed in the Cognitive Foundations Framework (Hoover & Tunmer, 2020) which defines word recognition as the ability to read words accurately and quickly. The basis of word recognition is alphabetic decoding, mapping letters or groups of letters (graphemes) to spoken sounds (phonemes). This relies on knowledge of the alphabetic principle, underpinned by LK and phonemic awareness, evidenced across alphabetic orthographies (Caravolas et al., 2013; Melby-Lervag et al., 2012). Successful decoding of written words provides the foundation for the creation of orthographic representations or sight words (Share, 1995), which are necessary for fluent word recognition (Ehri, 2005).

Difficulties in word recognition are characteristic of the dyslexic reading profile, which has a pooled prevalence of 7.10% in primary school aged children (Yang et al., 2022). Dyslexia is defined as a set of processing difficulties that affect the acquisition of reading and spelling, such that literacy attainment is not in line with age and cannot be explained by a lack of instruction (Carroll et al., 2024). Difficulties in phonological processing have consistently been found in dyslexic readers at the group (Melby-Lervag et al., 2012) and individual level (Saksida et al., 2016) and shown to be present before formal literacy instruction begins (Snowling & Melby-Lervag, 2016). Furthermore, interventions targeting the foundation skills of PA, phonics, reading, spelling and word meaning contribute to significant improvements in reading outcomes in children with or at-risk for dyslexia (Hall et al., 2023).

Commercially available screeners (e.g., GL Assessment Dyslexia screener, Horne, 2020; Nessy Learning Dyslexia Quest, Bristol Dyslexia Centre, 2011) typically comprise standardised measures of phonological processing and LK. The accuracy of a screener is determined by examining the agreement in classification between a reference or gold standard such as a standardised reading assessment (administered later in development) and the screener. Metrics of classification accuracy are derived from this process, including sensitivity (proportion of correctly identified cases or true positives) and specificity (proportion of correctly rejected non-cases or 'true negatives'). The trade-off between sensitivity and specificity is quantified by the Area Under the Receiver Operating Characteristic Curve (AUC), which represents the probability of correct classification for any randomly drawn pair of individuals (Streiner & Cairney, 2007).

In measuring acquired knowledge, static assessments often fail to account for variation in cultural and linguistic environments, risking misclassification of children from socioeconomically disadvantaged or bilingual backgrounds (Peña & Halle, 2011). By focusing on learning potential rather than current attainment, DA offers a more context-sensitive approach, less confounded by differences in early educational exposure. Children acquiring the majority language as an additional language and/or from socio-economically disadvantaged backgrounds (collectively termed culturally and linguistically diverse; CALD) are likely to be at elevated risk of misclassification, particularly being overidentified as having reading difficulties. Crucially, these children may score poorly on static screeners due to reduced English language learning opportunities rather than an underlying specific learning difficulty such as dyslexia (Tzuriel, 2000). This may lead to the misallocation of school resources, with children receiving targeted and intensive interventions they do not need. A further issue is that many children may find these measures too difficult, leading to floor effects (Catts et al., 2009) and data that cannot accurately differentiate those requiring additional support.

DA measures a child's potential to learn from feedback during a task rather than their than developed ability (Grigorenko & Sternberg, 1998). Unlike static measures that capture only actualised development, DA reveals readiness to benefit from instruction, a function particularly valuable in diverse early years classrooms. DA may provide an alternative to static screeners for reading related skills, particularly for children from CALD backgrounds. Although not novel, there has been increasing interest in DA of reading related skills in recent years. Dixon et al. (2023a) reviewed 18 studies on the ability of DA to predict growth in children's reading ability over and above established static predictors of reading. A subset of studies ($N=13$) focussed on PA and decoding to predict word reading accuracy and/or fluency. Six studies used DAs of PA (Bridges & Catts, 2011; Coventry et al., 2011; Gellert & Elbro, 2017a; Krenca et al., 2020; Spector, 1992). Of these, the four that measured word reading accuracy found that the DA predicted 4–21% unique variance, but that this predictive power diminished over time. The two studies of word reading fluency found that DA of PA predicted 1–9% unique variance. Seven studies used DAs of decoding (Aravena et al., 2016; Cho et al., 2014, 2017, 2020; Fuchs et al., 2011; Horbach et al., 2015, 2018). In terms of word reading accuracy (included in all studies), DA accounted for 1–17% unique variance. For word reading fluency this ranged from 4 to 8%, but this

was not significant in two studies. Only three studies reported recruiting monolingual and bilingual children, but none explored the predictive validity of DA separately in these two groups. Likewise, five studies reported the socio-economic status (SES) of their participants, but the impact of SES on predictive validity was not explored.

A complementary review by the same authors (Dixon et al., 2023b) investigated the accuracy of DA for identifying reading difficulties. Most reviewed studies (13/15) focussed on DA of PA and decoding. The criteria for reading difficulties varied but overall, studies aimed to identify children at risk of the dyslexic reading profile. In the five studies that used DA of PA (Bridges & Catts, 2011 study 1 and 2; O'Connor & Jenkins, 1999; Krenca et al., 2020; Gellert & Elbro, 2017a), AUC estimates ranged from 0.5 to 0.7, indicating acceptable levels of identification accuracy. When added to static predictors, DA increased identification accuracy. The remaining eight studies used DA of decoding (Cho et al., 2020; Compton et al., 2010; Gellert & Elbro, 2017b, 2018; Petersen et al., 2016, 2018; Petersen & Gillam, 2015) and reported that when DA was entered alone or with static predictors, models yielded AUCs of 0.8 or above, which indicates excellent identification accuracy. Furthermore, four of the studies reported higher classification accuracy when DA was entered into a model after static measures of LK, naming fluency and PA, with AUCs in the 0.8 to 0.9 range. There was some evidence that identification accuracy was developmentally constrained, with one study finding that predictive power decreased with increasing grade level (Petersen et al., 2018) and another finding better prediction from DA of decoding when it was administered in kindergarten compared to grade 1 (Gellert & Elbro, 2018). Two studies compared identification accuracy in a subgroup of bilingual students to that in monolinguals. Petersen et al. (2016) found that a static measure of reading administered in kindergarten correctly identified fewer poor readers in grade 1 in the bilingual sample. In contrast, their DA of decoding correctly identified more poor readers in the bilingual sample compared to the main sample and adding it to static measures increased specificity. The second study followed these learners to grade 5 (Petersen et al., 2018) and found that the DA of decoding was a significant and unique predictor of reading status up to grade 5 in the bilingual group, but only up to grade 2 in the monolingual group.

To quantify the relationship between DA and reading, Wood et al. (2024a) conducted a meta-analysis of DAs of PA, symbol-sound learning and decoding in typical and at-risk readers, including both concurrent and longitudinal studies. Across 192 effect sizes, the overall effect size was $g = 0.49$ indicating a strong correlation between DA and reading. The strength of this did not differ by reader status or language group, but it was stronger for younger children. A follow-up review investigated the effect of DA characteristics (Wood et al., 2024b). Whether the DA was administered in person or via a computer, or whether the symbols were novel or familiar did not significantly alter the effect size. However, DAs of PA and decoding had stronger correlations with reading than those focusing on solely on symbol-sound learning.

DA of reading related skills, particularly decoding, shows promise for the early identification of children at risk of reading difficulties, obviating the need to wait until later in development when static measures of reading related skills become more reliable. Few studies have focused on CALD children, but there is some evidence to suggest that DA of decoding could improve classification accuracy in CALD

children. Further research is needed in the form of well powered studies with diverse samples of children.

Present study

In the current study we created an adapted, computer assisted version of the DA decoding developed by Elbro et al. (2012). We tested its ability to predict later reading outcomes and identify children at risk of word reading difficulties in a CALD sample of children entering their first year of formal schooling (Reception) in England. We posed three research questions and corresponding hypotheses:

1. What is the correlation between performance on a computer assisted DA of decoding and SES and English language proficiency? The relationship between DA of decoding and background variables has not been explored in previous studies (Dixon et al., 2023a), but based on the aim of DA being to reduce the impact of educational inequalities (Grigorenko, 2009) we hypothesised that there would not be strong relationships between performance on our DA of decoding and SES or English language proficiency (hypothesis 1).
2. Does learning performance on a computer assisted DA of decoding predict later word reading ability? We hypothesised that the DA of decoding would predict unique variance in later word reading as found previously (Dixon et al., 2023a) (hypothesis 2). In addition to planned analyses for the whole sample, we conducted exploratory analyses for monolingual (ML) and children learning English as an Additional Language (EAL) separately, to test whether any predictive relationship differed between the two groups.
3. Can a computer assisted DA of decoding accurately screen for word reading difficulties, specifically the dyslexic reading profile? Given that our DA of decoding was modelled on the DA from Elbro et al. (2012) we hypothesised that it would broadly replicate the acceptable levels of screening accuracy previously found (hypothesis 3). Gellert and Elbro (2017b), 2018) also found that this type of DA of decoding significantly improved accuracy of identification when entered after established static predictors. However, some caution is required in hypothesising a similar finding as our sample differs from theirs, we sampled children from CALD backgrounds learning to read in the English educational system, whereas they oversampled children at risk for later reading difficulties from Danish kindergartens. Because we had purposely sampled children from both English ML and EAL backgrounds, we explored classification accuracy for each group separately, as DA of decoding may improve classification accuracy more in bilingual learners (Petersen et al., 2016, 2018).

Method

Design

286 children were followed longitudinally from the summer term of their reception year to the spring term of Year 1. Data were collected at two time points approximately 10 months apart. Following pilot work, it was established that the end of the winter term through to the beginning of the spring term would be optimum for the first timepoint. This was based on children's performance on the DA of decoding, and recognition of the need for the reception children to have sufficient time to settle into school life. However, due to COVID-19 restrictions and significant school disruption, this timepoint was delayed by 5 months. This meant that the children had had more formal literacy instruction than originally intended.

Participants

We aimed to recruit a large enough sample to detect a 10% prevalence rate of the dyslexic reading profile with at least 80% sensitivity (Bujang & Adnan, 2016). Local government-funded primary schools were stratified into high/low SES and high/low EAL enrolment according to Department for Education statistics available at the time. Schools were asked to share data on the demographic characteristics, EAL status and Special Educational Needs (SEN) status (yes/no) of participating pupils. SES was measured using the Income Deprivation Affecting Children Index (IDACI) ranks derived from participants' home post codes (Ministry of Housing, Communities & Local Government, 2019). These range from 1 to 32,844, where the lower ranks indicate a higher proportion of children in that area living in low-income families. The sample at the first time point (t1) consisted of 317 children ($n=166$ male) with a mean age of 63.00 months ($SD=3.82$), or 5 years 3 months, recruited from six schools in Leeds. The sample included 47 SEN children and 75 learning EAL. IDACI ranks ranged from 261 to 32,510 (median = 13,615). EAL learners spoke a total of 26 different languages, primarily Urdu (19.4%), Bengali (16.7%), and Arabic (9.7%). At t1 a small number of children were excluded for being very new to English ($n=8$), SEN-related behavioural issues ($n=4$), or insufficient vocabulary knowledge ($n=1$). By t2, 286 children remained in the sample, representing an attrition rate of 9.8% ($n=146$ male; $n=41$ SEN; $n=65$ EAL; mean age in months: 72.94 [$SD=3.78$], or 6 years 0 months). The proportion of our sample who had EAL is in line with the latest estimate of the proportion of EAL pupils in primary schools in England (23.4%, DfE, 2025). Common R reasons for attrition included absence on the day of assessment ($n=19$; 61.3%), having moved school ($n=11$; 35.4%), and unavailability for assessment ($n=1$; 3.2%). We compared the demographic profiles of children missing at the second time point (t2) to children who remained in the study (missingness data can be found in Supplementary material, Table 1). The two groups did not differ significantly in terms of EAL status, SEN status, gender, or age. However, as a group, children who were not available for assessment at t2 came from significantly more deprived neighbourhoods in terms of IDACI scores ($p=.021$).

Measures and procedures

At t1 children were assessed on the DA of decoding as well as static measures of nonverbal reasoning, language, reading, and reading related skills. Assessments were split over two or three sessions. The assessment battery at t2 consisted of the static single word reading accuracy measures only. All assessments were carried out individually in quiet areas of the school by members of the research team accompanied by research assistants who had received 2.5 h of training. Standardised assessments were administered according to their manuals and research assistants were observed during data collection to ensure the fidelity of data collection. The scoring of assessments administered by research assistants was double-checked by a member of the research team prior to entry. Ethical approval was obtained from The University of Leeds School of Psychology Research Ethics Committee. Consent was provided by each headteacher, and parents were informed about the research and had the opportunity to opt-out, in accordance with the guidelines of the British Psychological Society (BPS, 2021). Ongoing child assent was monitored by the research assistants during data collection.

Static measures

Nonverbal reasoning. Raven's Coloured Progressive Matrices (CPM; Pearson, 2008) are a series of diagrammatic puzzles, in which children are asked to select one of six solutions to 'fit' the puzzle. The CPM was standardised using a UK sample of 4- to 11-year-olds and reports a split-half reliability of 0.97.

Vocabulary. The British Picture Vocabulary Scale-III (BPVS-3; Dunn et al., 2009) is a measure of receptive vocabulary knowledge, in which children are asked to match pictures to words spoken by the examiner. The BPVS-3 was standardised using a UK sample of 3- to 16-year-olds. No statistics are reported for reliability, however the PPVT-4, the assessment on which the BPVS-III is based, reports a high split-half reliability coefficient of $r=0.94$.

Phonological awareness (PA). Two subtests of the Phonological Assessment Battery-II (PhAB-2; Gibbs & Bodman, 2014) were administered. In Alliteration, children are presented with images depicting three objects. After the name of each object is said by the examiner (e.g., well, peg, pot), children are asked to indicate which two objects share the same initial phoneme. Parts 1 and 2 (up to age 6) were administered, though part 2 is only attempted if a child scores three or more correct on part 1. Internal reliability (alpha) for parts 1 and 2 is 0.86. In Blending, the children hear phonemes said by the examiner in a serial fashion (e.g., /n/ /i/ [knee]) and are asked "what word do we get when we put these sounds together?" The test consists of two parts, though part 2 is only attempted if at least five items are answered correctly or there are no more than four consecutive incorrect responses on part 1. Internal consistency (alpha) for Blending is 0.96. The PhAB-2 is standardised on a sample of 773 children aged 4;7 to 11;11 in England, Scotland, and Wales (8% with EAL).

Rapid Automatisated Naming (RAN). We administered the Picture Naming subtest of the PhAB-2, in which children were presented with a matrix of randomly repeating rows of five objects (ball, hat, door, table, box) and asked to name all items as quickly

as possible. To reduce the length and attentional demands of the battery, only part 1 was administered. Time taken to name all items was recorded in seconds. Internal reliability (alpha) is 0.89.

Letter-sound knowledge (LSK). In the LSK subtest of the York Assessment of Reading for Comprehension-Early Reading (YARC-ER; Hulme et al., 2012) children are shown a series of printed letters and digraphs and asked, “What sound does this letter make?” We administered the Extended assessment, consisting of 26 letters and 6 digraphs. Children attempted all 32 items. The LSK Extended assessment has an internal reliability of 0.98. The YARC-ER was standardised using a nationally representative sample of 662 children in the UK, 16.3% of whom were EAL learners.

Word reading. Word reading ability was measured using two different assessments. In the Early Word Recognition (EWR) subtest of the YARC-ER, children attempt to read aloud a series of 15 regular words and 15 exception words of increasing difficulty (e.g., *cat*, *in*, ... *biscuit*, *giant*). Given the fairly high performance by some children that we observed on the EWR subtest during time 1 data collection, we decided to administer the Diagnostic Test of Word Reading Processes (DTWRP; Forum for Research in Literacy and Language, 2012) as a second, more challenging measure capable of capturing growth between time points. The DTWRP measures recognition of regular words, exception words, and nonwords, with 30 items each. Administration of each of the three lists discontinues after five consecutive errors. The DTWRP was standardised on a representative sample of 1,125 children in England aged 5 to 12 years, 12.4% of whom were EAL learners. Internal reliability (alpha) for the complete test of 90 items is 0.99.

DA of decoding

We designed a computer assisted DA of decoding based on Elbro et al. (2012), operationalising DA as learning within the assessment, using a pre-test – teach – post-test format and novel symbols. Few studies have used computer assisted DAs of decoding; however, advantages may include automating administration and scoring, standardising feedback for learning and reducing the length of administration (Dixon et al., 2023a, 2023b). The task was adapted from the original version, by adopting a story format intended to maximise appeal and engagement, in which ‘Chompy the Crocodile’ asks the children for help to crack an ancient code and find some treasure. The task was structured to mimic the process of learning to decode across three tasks. First, children are taught three novel symbols and their sounds (η = /s/, $\diamond$ = /m/, and $\bullet$ = /b/) to criterion. Gellert and Elbro (2017b) used the sound /a/ as in *calm* in their Danish study. Unfortunately, this resulted in too many real words in English, so it was changed to /b/ as in *lock*. The consonants /s/ and /m/ are common across languages, the vowel /b/ is somewhat less common across languages, but frequent in English (Mielke, 2018). One of the symbols $\blacksquare$ was also changed to the alternative non-nameable symbol $\bullet$. Second, children are taught how to blend these sounds together to form CV and VC syllables, following Gellert and Elbro (2017b). Finally, they are asked to read pseudowords written using the symbols. The task was designed in PsychoPy (Peirce, 2007) and administered on a laptop, with the examiner indicating correct and incorrect responses with key presses. The script, including instructions and

feedback, was recorded by a female native English speaker, and was played to the children through headphones. The individual phases of the task are described below. Reliability of each phase represents the omega coefficient, and its 95% confidence interval computed using the MBESS package in R (Kelley, 2022).

Symbol-Sound Learning. After an initial presentation of the three sound-symbol pairs, a training phase began in which each symbol was presented, and children were asked to provide the corresponding sound. There was a maximum of 10 blocks, where the symbols were presented in a random order. Corrective feedback was provided (e.g., “not quite, it’s /s/” or “that’s right, it’s /s/”); training discontinued after three consecutive correct blocks or when 30 trials had been attempted. Reliability of sound-symbol learning (omega) was 0.90 [0.87, 0.93]. The score for this part of the task was the number of correctly produced sounds out of 30 (if children reached criterion they were credited with the remaining trials).

Blending. The blending phase began with four practice trials in which the examiner demonstrated synthesis of sounds (order presented in Fig. 1). In the first trial, two symbols were presented apart, and each sound was provided with a one second gap (e.g., /s/ ... /v/). Then the symbols moved closer, and each sound was provided with a 500-millisecond gap. Next the two symbols were joined together, and the blend was produced (e.g., /sv/). The remaining three training trials incorporated a gradual release of autonomy, with the children initially being asked to repeat what they heard and by the final trial to produce the blend independently. The subsequent test phase consisted of 20 trials presented in five blocks of four items (there were four possible combinations of the three symbols in pairs) in which children were asked to blend symbols into CV and VC structures. The order of test trials was pseudorandomised and presented in this fixed order. Test trials contained two stages: first, the two symbols were presented and the children were invited to repeat the sounds individu-

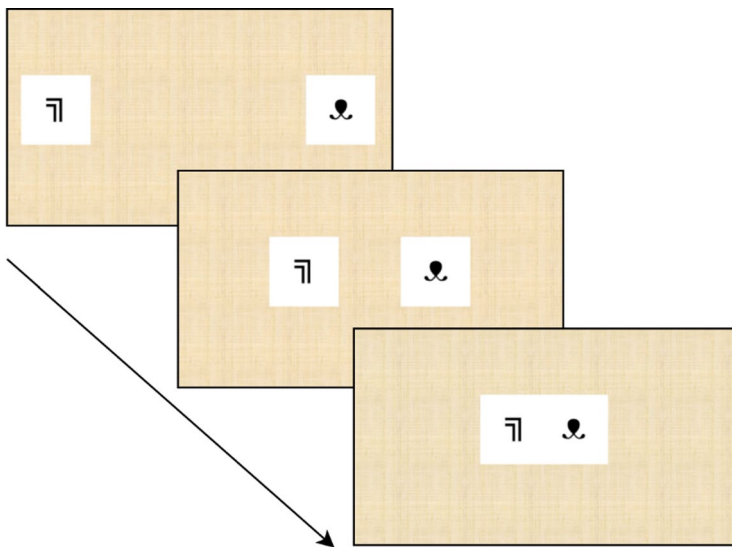


Fig. 1 Blending phase of the DA of decoding

ally, receiving corrective feedback from the examiner if they could not remember the sounds; secondly, the two symbols were brought together and the children were asked to produce the blend. Corrective feedback was provided, e.g. “not quite, it’s /sp/”. Training discontinued after two consecutive correct blocks or when 20 trials had been attempted. The score is the number of correct trials out of a maximum of 20, (if children reached criterion they were credited with the remaining trials). Reliability of the blending test phase was 0.97 [0.97, 0.98].

Reading. The final reading phase was attempted only if the child scored a total of 16 out of 20 or two consecutive blocks correct on blending trials. If the child did not achieve either criterion, they did not attempt the reading trials and were awarded a score of 0. In the reading phase, the children were asked to read increasingly more structurally complex pseudowords written in the novel orthography (e.g., /mɒm/, /ɒsmɒ/, /smɒms/). After an initial practice item, the children were asked to decode a maximum of 12 pseudowords presented one at a time. Reading discontinued after three consecutive incorrect trials or when the final trial was attempted. Reliability of the reading phase was 0.96 [0.95, 0.96]. The DA took on average 15–20 minutes to administer.

Analysis

In Gellert and Elbro (2017b) the scores from the blending and reading phases of the task were summed to form a composite decoding score, resulting in two scores from the task; dynamic letter learning and dynamic decoding. To explore whether our DA of decoding was similarly multidimensional, we conducted principal components analysis (PCA) on the phases of the task: sound-symbol learning, blending, and reading. These variables indicated an acceptable level of sampling adequacy, with Keyser-Meyer-Olkin values between 0.55 and 0.62. and an R-matrix determinant of 0.71. A two-factor solution with orthogonal (varimax) rotation was preferred, yielding eigenvalues of 1.36 and 1.12 for the first and second components, respectively, explaining 83% of the variance. Blending and reading loaded highly onto a first ‘DA synthesis’ component (0.92 and 0.71 respectively) while sound-symbol learning loaded highly onto a second ‘DA learning’ component (0.96). These component scores were entered separately into statistical models as measures of dynamic decoding as described below. We carried out the same PCA for both language subgroups and replicated the two-factor structure for each (the analyses can be found in Supplementary materials, Table 3).

To address hypothesis one, we carried out correlation analysis to examine relationships between covariates, static, and dynamic variables.

To address hypothesis two, we used linear mixed effects (LME) models to predict performance on a word reading composite measure at the second timepoint. This composite represented the sum of YARC EWR and DTWRP total score (regular, irregular, & nonwords) divided by two, with sample statistics of $M=12.54$ ($SD=9.72$) at t1 and $M=26.87$ ($SD=13.40$) at t2. Modelling was carried out using the lme4 package (Bates et al., 2015) in R (R Core Team, 2024) and all data, analysis code, and the study’s preregistration are available online via the following link: https://osf.io/agr6b/overview?view_only=3ca0bfa8e0b14b4c86878b85bd217190). The

addition of random intercept terms for schools ($n=7$) caused some issues with convergence and therefore LME models reported here contain only a single intercept for subject (note that a random slope for time was inappropriate due to only two repeated measurements). Maximum likelihood estimation was used to allow comparison of nested models, and diagnostics included model convergence status, variance inflation factors, and plots of standardised residuals. Continuous predictors were scaled and centred, while time, EAL status, and dyslexic profile at-risk status were dummy coded with reference categories of t1 (0), monolingual (ML) and non-dyslexic reading profile, respectively. We employed a step-up model building process beginning with covariates, followed by either dynamic variables or static assessments such that these were allowed to compete evenly. In a final exploratory step, we fitted ‘final’ models with dynamic variables entered after covariates and static predictors. Statistical inference of predictors at each step was estimated using the results of likelihood ratio tests and pseudo R-squared was estimated using the MuMIn package (Bartón, 2019).

To address hypothesis three, we defined dyslexic profile at-risk status as a t2 word reading composite score at or below the 15th quantile for the sample (i.e. -1SD or a score of 11.5). We used the `glm()` function in R to run logistic regression models predicting binary risk status at t2 as a function of the same covariates, static, and dynamic variables used in LME models. The `reportROC` package (Du & Hao, 2020) was used to compute measures of classification accuracy including AUC, sensitivity, specificity, false positive rate (FPR) and false negative rate (FNR). We also statistically compared the classification accuracy of dynamic and static models at the final step using DeLong’s method with the `roc.test()` function of the `pROC` package (Robin et al., 2011). Although there are no absolute thresholds, an AUC of 0.7 to 0.8 represents acceptable discrimination, 0.8 to 0.9 represents excellent discrimination and above 0.9 outstanding discrimination (Hosmer et al., 2013). The total sample size in all statistical models presented here is $n=286$ due to attrition at t2.

Results

Descriptive statistics and intercorrelations are presented in Table 1. Children’s word reading scores increased significantly over time ($t(285) = -30.63, p < .001$), and this pattern applied equally to ML and EAL subgroups. The largest standardised mean differences between the language subgroups were in vocabulary ($d=1.27$), LSK ($d=0.85$), and early word reading ($d=0.85$). No significant group difference was found on the DA Learning component, though ML children performed significantly better on DA Synthesis ($d=0.70$). Descriptive data and correlations for the raw scores from the DA task phases can be found in supplementary materials (Table 2).

Correlations between the DA of decoding, English language proficiency, and SES

In terms of the first hypothesis, the DA component scores correlated only weakly with SES (Learning, $r=.06$; Synthesis, $r=.33$) and weakly to moderately with English receptive vocabulary at t1 ($rs=.28-.46$). The static predictors (LSK, PA & RAN)

Table 1 Descriptive and intercorrelations for the static measures and DA at t1 and t2

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. SES													
2. Matrices	.17**												
3. DA: Learning	.06	.18**											
4. DA: Synthesis	.33**	.31**	.00										
5. Vocabulary	.43**	.45**	.28**	.46**									
6. Alliteration	.35**	.38**	.27**	.61**	.58**								
7. Blending	.35**	.26**	.21**	.72**	.48**	.60**							
8. LSK	.41**	.33**	.40**	.58**	.62**	.57**	.63**						
9. RAN	.21**	-.14*	-.16**	-.06	-.04	-.05	-.05	-.01					
10. EWR t1	.39**	.42**	.20**	.75**	.53**	.58**	.69**	.71**	-.13*				
11. EWR t2 ^a	.38**	.39**	.30**	.68**	.54**	.59**	.68**	.80**	-.01	.78**			
12. DTWRP t1	.36**	.41**	.14*	.70**	.47**	.53**	.61**	.56**	-.16**	.90**	.67**		
13. DTWRP t2 ^a	.32**	.42**	.17**	.75**	.51**	.59**	.68**	.66**	-.04	.82**	.88**	.79**	
Mean	16303.15	14.21	0.00	0.00	70.3	5.59	9.25	26	108.91	11.71	21.17	12.92	32.4
Standard deviation	10779.54	4.27	1.00	1.00	20.02	2.84	6.49	7.83	50.44	7.95	8.26	12.18	19.25
Minimum	261	4	-5.53	-1.69	9	0	0	0	42	0	0	0	0
Maximum	32,510	28	0.94	2.04	114	10	18	32	358	30	30	63	81
Skewness	0.18	0.51	-2.5	0.35	-0.77	-0.26	-0.21	-1.86	1.74	0.35	-1	1.63	0.24
N	317	317	317	317	317	317	317	317	317	317	286	317	286
ML vs EAL (D)	1.14**	0.38	0.20	0.70**	1.27**	0.72**	0.48**	0.85**	0.06	0.85**	0.69**	0.67**	0.59**

SES=IDACI rank; Matrices=CPM; DA=dynamic decoding component scores; Vocabulary=British Picture Vocabulary Scale-3; LSK=YARC Letter Sound Knowledge; RAN=rapid automatized naming; EWR=YARC Early Word Reading; DTWRP=Diagnostic Test of Word Reading Processes; $n=317$ unless stated; ^a $n=286$; ** $p<.01$, * $p<.05$

were weakly to moderately correlated with SES ($r_s=.21-.41$) and moderately to strongly correlated with vocabulary knowledge ($r_s=.48-.62$).

In order to contrast the strength of relationships between SES, vocabulary, and DA component scores relative to those with LSK, PA and RAN, zero-correlations were compared statistically using Williams' t-test (see supplementary material, Table 4). Results provide further nuance to the above in showing that it was the DA Learning component specifically that showed significantly weaker relationships than static measures with SES and vocabulary.

Together, results suggest that the DA of decoding (and specifically the learning component) was less strongly related to prior learning opportunities than static measures of language and reading. Also of note was the high degree of stability between the two time points in the measures of reading, YARC EWR ($r=.78$) and DTWRP ($r=.79$).

Predictive validity of static measures and DA of decoding components

LME models predicting the word reading composite score at t2 can be found in Table 2. Regarding the full sample, likelihood ratio tests indicated that the covariates were all significant predictors, accounting for just under 25% of the variance. In a subsequent model, the autoregressor and all other static predictors explained an additional 47% of variance, though RAN did not make a statistically significant contribution. When added after the static predictors, both the DA Learning and DA Synthesis components contributed significantly to later reading skill, explaining 0.3% and 4% of the variance in steps 10 and 11, respectively. These patterns played out similarly for the ML subgroup, with DA Synthesis accounting for relatively more variance than DA Learning (0.6% and 4.5%, respectively). For EAL learners, only DA Synthesis was a significant predictor, accounting for 1.6% of variance in later reading skills.

Classifying risk status for the dyslexic reading profile

We identified 47 children at risk of the dyslexic reading profile, representing an overall prevalence rate of 16.4%. EAL learners were significantly more likely to be at risk ($n=24$; 37%) (vs. ML: $n=23$; 10.4%) ($\chi^2(1)=23.82$, $p<.001$). As expected, based on the classification criteria, the 47 children in the at-risk group obtained significantly lower word reading composite scores at t2 (at-risk: $M=5.89$ [range: 0–11.5]; typically developing: $M=30.89$ [range: 12–55.5]; $t(173)=27.35$, $p<.001$, Hedge's $g=3.14$). Logistic regression models predicting dyslexia risk status for the full sample are presented in Table 3, while separate models for the ML and EAL subgroups are presented in Table 4. ROC curves are presented in Fig. 2. Additionally, Table 6 presents logistic models containing only the dynamic components for the full sample as well as subgroups.

For the full sample, both DA Learning and DA Synthesis were significant predictors of later risk when entered after the covariates (after step 4), yielding excellent classification accuracy at step 6 (Table 4; $AUC=.945$). However, the static model at step 9 resulted in even higher accuracy ($AUC=.973$) which represented a significantly better fit according to DeLong's test ($\chi^2(7)=31.17$, $z=2.26$, $p=.024$). Among

Table 2 Linear mixed effects models predicting time 2 word reading for the full sample and Language subgroups

	Step	Predictor	Full sample (<i>n</i> =286)			Step	ML group (<i>n</i> =221)			Step	EAL group (<i>n</i> =65)		
			χ^2	R^2	ΔR^2		χ^2	R^2	ΔR^2		χ^2	R^2	ΔR^2
Covariates	1	SES	44.61**	0.094	0.094	1	18.80**	0.050	0.050	1	11.31**	0.108	0.108
	2	EAL	7.54**	0.108	0.014								
	3	Matrices	50.75**	0.196	0.088	2	27.72**	0.117	0.067	2	28.93**	0.313	0.204
Static	4	BPVS	34.41**	0.247	0.051	3	23.65**	0.167	0.050	3	10.12**	0.365	0.053
	5	Word composite t1	416.68**	0.518	0.271	4	347.66**	0.474	0.307	4	74.14**	0.587	0.222
	6	Alliteration	48.63**	0.580	0.063	5	32.03**	0.534	0.060	5	21.96**	0.676	0.089
	7	Blending	96.25**	0.677	0.097	6	65.4**	0.633	0.099	6	28.95**	0.755	0.079
	8	LSK	50.21**	0.716	0.039	7	45.17**	0.686	0.053	7	9.76**	0.775	0.020
Dynamic	9	RAN	2.10	0.717	0.002	8	3.71	0.690	0.004	8	0.42	0.776	0.001
	10	DA Learning	3.99*	0.720	0.003	9	5.72*	0.696	0.006	9	0.00	0.776	0.000
	11	DA Synthesis	64.54**	0.760	0.040	10	49.15**	0.741	0.045	10	9.18**	0.791	0.016

* $p < .05$; ** $p < .01$; R^2 =marginal pseudo R-squared; ΔR^2 =change in pseudo R-squared

the static variables, only the autoregressor (early word reading at t1) and LSK were significant predictors. Relative to the static model, the dynamic model identified 7 fewer true positive cases (sensitivity=.830) but did correctly reject an additional 23 true negative cases (specificity=.937). When entered in a 'final model' after covariates and static predictors (steps 10 and 11), neither dynamic component was a significant predictor of at-risk status.

A similar pattern was found for ML learners (Table 4). Both dynamic components were significant predictors of at-risk status when entered after covariates, yielding an AUC of .925 and correctly identifying 21 out of 23 true positive cases. Again, the only significant static predictors were the autoregressor and LSK knowledge, and the static model resulted in significantly better fit than the dynamic model ($AUC=.962$; $\chi^2(6)=24.23$, $z=1.98$, $p=.047$). Neither dynamic component made a significant contribution to classification when entered after covariates in steps 9 and 10.

For EAL learners, both dynamic components made a significant contribution to classification when entered after covariates, with the model at step 5 (DA Synthesis) accounting for nearly double the amount of variance (30.4%) than for ML learners (17.3%). This dynamic model yielded outstanding classification accuracy with an AUC of .976 and 100% sensitivity (i.e. identifying all 24 true positives). The fit of this model was not significantly different from that of a static model ($AUC=.985$; $\chi^2(6)=6.64$, $z=0.61$, $p=.545$). DA Learning was significant when entered after covariates and static predictors, resulting in an AUC of 1.00 and hence perfect classification accuracy. However, some caution is warranted in this interpretation given the relatively small EAL sample size and the likelihood of overfitting.

Non-readers

To explore the potential impact of the DA of decoding being administered later than planned, we focused on a sub-sample of children still in the very earliest stages of learning to read. Non-readers were defined as children scoring ≤ 3 on YARC EWR at t1 ($n=51$; 17.8%). As a group, the non-readers came from lower SES backgrounds ($d=-1.09$), had poorer English vocabulary knowledge ($d=-1.31$), non-verbal ability ($d=-0.85$), letter-sound knowledge ($d=-2.32$) and PA (alliteration $d=-1.54$, blending $d=-2.39$) than the remainder of the sample. There was no significant difference between groups on the RAN task. A greater proportion of non-readers had EAL (41.5%) than in the whole sample (23%). The proportion of males across groups was similar (non-readers 53%, main sample 51%). Of the 51 children who met the non-reader criterion, 38 (75%) also met criteria for being at-risk for the dyslexic reading profile at t2. Given that 71% and 65% of these children scored zero on the YARC EWR and PhAB Blending tests, respectively, at t1 we did not include these measures in the logistic models. Results should be interpreted with caution given the low performance of these children and small sample size.

As shown in Table 5, both dynamic components made a significant contribution when entered after covariates, resulting in excellent classification accuracy ($AUC=.852$). The dynamic model at step 6 correctly identified 28 out of 38 true positive cases. Although the static model at step 7 resulted in higher accuracy ($AUC=.907$), the fit of this model was not significantly different from that of the

Table 3 Logistic models predicting t2 at-risk status for the full sample ($n=286$)

Model	Step	Predictor	χ^2	R^2	ΔR^2	AUC	Sen	Spec	TN	TP	FN	FP	FPR	FNR
Covariates	1	SES	41.81**	0.230	0.230	0.761	0.872	0.678	162	41	6	77	0.322	0.128
	2	EAL	6.13*	0.261	0.031	0.783	0.936	0.636	152	44	3	87	0.364	0.064
	3	Matrices	14.35**	0.331	0.070	0.830	0.851	0.736	176	40	7	63	0.264	0.149
	4	BPVS	13.60**	0.395	0.063	0.858	0.809	0.736	175	38	9	64	0.268	0.191
Static	5	EWR t1	76.28**	0.698	0.304	0.964	0.894	0.895	214	42	5	25	0.105	0.106
	6	Alliteration	1.37	0.703	0.005	0.965	1.000	0.803	191	47	0	48	0.201	0.000
	7	Blending	2.93	0.713	0.010	0.966	0.872	0.929	222	41	6	17	0.071	0.128
	8	LSK	12.97**	0.757	0.043	0.973	0.979	0.849	203	46	1	36	0.151	0.021
	9	RAN	0.39	0.758	0.001	0.973	0.979	0.841	201	46	1	38	0.159	0.021
Dynamic	5	DA learning	14.89**	0.460	0.066	0.886	0.830	0.778	186	39	8	53	0.222	0.170
	6	DA synthesis	47.88**	0.650	0.190	0.945	0.830	0.937	224	39	8	15	0.063	0.170
Dynamic	10	DA learning	1.23	0.762	0.004	0.973	0.979	0.845	202	46	1	37	0.155	0.021
After static	11	DA synthesis	0.39	0.763	0.001	0.972	0.979	0.849	203	46	1	36	0.151	0.021

AUC=area under receiver operating curve; TN=true negatives ($n=239$); TP=true positives ($n=47$); FN=false negatives; FP=false positives; FPR=false positive rate; FNR=false negative rate; * $p < .05$; ** $p < .01$; R^2 =Nagelkerke

dynamic model at step 6 ($\chi^2(7)=7.26$, $z=0.82$, $p=.411$). When the dynamic components were added after the static predictors, DA Learning increased accuracy (AUC=.923), which approached significance, by identifying 33 true positives and one fewer false positive. Adding DA Synthesis significantly improved accuracy further (AUC=.953), by achieving perfect specificity.

Classification accuracy of DA of decoding only

To explore the classification accuracy of the DA of decoding when no other information about a child is available, we ran logistic models including only the DA components (Table 6). In all models, both dynamic components were significant predictors of future at-risk status and in every case DA Synthesis accounted for the most variance. Classification accuracy was excellent in every model, with the highest AUC found for the EAL group (.971) and the lowest for the non-reader subgroup (.800).

Discussion

Recognising the limitations of static assessments, particularly when used with CALD children and young learners who have yet to receive formal reading instruction, this study examined the potential of a DA of decoding as a more inclusive screening tool. Specifically, we explored whether a DA of decoding could more accurately identify the dyslexic reading profile and better differentiate learners' needs. We posed three main research questions and corresponding hypotheses.

Supporting our first hypothesis, the DA Learning and DA Synthesis components of our DA of decoding were only weakly correlated with socioeconomic status (SES) and weak-to-moderately correlated with English vocabulary knowledge. By contrast, traditional static measures of pre-reading skills showed stronger correlations with these background factors. Comparing the strength of these relationships, the DA Learning component showed statistically weaker relationships with SES and vocabulary.

Supporting our second hypothesis, the DA of decoding components predicted significant unique variance in word reading at t2, after accounting for the autoregressor, nonverbal ability, vocabulary, LSK, PA and RAN. Of the two components, DA Synthesis explained more variance and the models including the demographic variables, static predictors and DA components provided good fits to the data.

Our main aim was to explore whether our DA of decoding could accurately screen for later word reading difficulties, specifically the dyslexic reading profile, in a CALD sample. When we identified children who were at risk for the dyslexic reading profile (16% of the full sample), a higher proportion were children with EAL (37%), compared to monolinguals (10%). For the full sample, our DA of decoding resulted in outstanding classification accuracy when used alone and when entered after the demographic variables, however a model with the demographic variables and static predictors resulted in slightly better accuracy. Adding the dynamic components did not significantly improve classification accuracy further, although it did improve specificity. While a similar picture emerged for the ML subsample, the findings for

Table 4 Logistic models predicting t2 at-risk status for ML and EAL subgroups

Model	Step	Predictor	χ^2	R^2	ΔR^2	AUC	Sen.	Spec.	TN	TP	FN	FP	FPR	FNR
ML														
Covariates	1	SES	22.53**	0.199	0.199	0.754	0.870	0.747	148	20	3	50	0.253	0.130
	2	Matrices	3.06	0.224	0.025	0.804	0.826	0.763	151	19	4	47	0.237	0.174
	3	BPVS	7.21**	0.283	0.059	0.840	0.870	0.732	144	20	3	54	0.273	0.130
Static	4	EWR t1	45.12**	0.610	0.327	0.956	1.000	0.793	157	23	0	41	0.207	0.000
	5	Alliteration	0.71	0.615	0.005	0.955	1.000	0.823	163	23	0	35	0.177	0.000
	6	Blending	1.87	0.627	0.012	0.961	0.957	0.854	169	22	1	29	0.146	0.043
	7	LSK	6.79**	0.670	0.043	0.964	0.957	0.864	171	22	1	27	0.136	0.043
	8	RAN	2.19	0.683	0.014	0.962	0.957	0.864	171	22	1	27	0.136	0.043
Dynamic	4	DA learning	8.81**	0.352	0.069	0.870	0.913	0.717	142	21	2	56	0.283	0.087
	5	DA synthesis	23.64**	0.525	0.173	0.925	0.913	0.778	154	21	2	44	0.222	0.087
Dynamic	9	DA learning	0.13	0.684	0.001	0.963	0.957	0.864	171	22	1	27	0.136	0.043
After static	10	DA synthesis	0.02	0.684	0.000	0.963	0.957	0.864	171	22	1	27	0.136	0.043
EAL														
Covariates	1	SES	3.43	0.070	0.070	0.616	0.750	0.488	20	18	6	21	0.512	0.250
	2	Matrices	14.76**	0.333	0.263	0.808	0.750	0.780	32	18	6	9	0.220	0.250
	3	BPVS	6.69**	0.434	0.101	0.865	0.917	0.756	31	22	2	10	0.244	0.083
Static	4	EWR t1	32.98**	0.805	0.371	0.977	0.875	0.976	40	21	3	1	0.024	0.125
	5	Alliteration	0.13	0.806	0.001	0.977	0.875	0.976	40	21	3	1	0.024	0.125
	6	Blending	0.34	0.809	0.003	0.978	0.958	0.902	37	23	1	4	0.098	0.042
	7	LSK	6.42*	0.862	0.052	0.986	1.00	0.902	37	24	0	4	0.098	0.000
	8	RAN	2.37	0.879	0.018	0.985	0.958	0.951	39	23	1	2	0.049	0.042
Dynamic	4	DA learning	6.52*	0.523	0.089	0.888	0.958	0.707	29	23	1	12	0.293	0.042
	5	DA synthesis	29.09**	0.827	0.304	0.976	1.00	0.902	37	24	0	4	0.098	0.000
Dynamic	9	DA learning	18.49**	1.00	0.121	1.00	1.00	1.00	41	24	0	0	0.000	0.000
After static	10	DA synthesis	0.00	1.00	0.000	1.00	1.00	1.00	41	24	0	0	0.000	0.000

AUC=area under receiver operating curve; TN=true negatives (ML $n=198$; EAL $n=41$); TP=true positives (ML $n=23$; EAL $n=24$); FN=false negatives; FP=false positives; FPR=false positive rate; FNR=false negative rate; * $p<.05$; ** $p<.01$; R^2 =Nagelkerke

children with EAL were different: on their own the dynamic components achieved outstanding classification accuracy and when added to the demographic variables all at-risk children were correctly identified. Finally, adding the DA of decoding after the static predictors resulted in perfect classification accuracy. In a sample of children who were non-readers at the first timepoint, the dynamic components achieved excellent classification accuracy on their own, identifying 36/38 of the non-readers classified as at-risk for dyslexia at t2. When added to the demographic variables, overall accuracy increased but sensitivity reduced. The best overall accuracy was seen when the dynamic components were added after the static predictors, where specificity was increased.

We hypothesised that our DA of decoding would not be strongly related to children's backgrounds (SES & English vocabulary knowledge) because DA was developed, in part, for children who haven't been able to access or take advantage of learning experiences (Grigorenko, 2009). Static measures of reading related skills are typically developed for monolingual children, who have had experience in that language. In our sample, scores on LSK, PA and early reading were moderately to strongly correlated with English vocabulary knowledge. That the relationships between the DA Learning component and SES and English vocabulary were significantly weaker than those with between the static measures and SES and English vocabulary lends support to this central tenet of DA and suggests that our task may be a less biased and more accurate screener for children from CALD backgrounds.

A recent review found that DAs of decoding predicted 1–17% of unique variance in word reading accuracy (Dixon et al., 2023a). Our finding that the two components combined accounted for 4.3% unique variance in word/nonword reading at t2 across the whole sample, is towards the lower end of this range. However, this was after an extensive battery including the autoregressor (t1 early reading), three demographic variables and four static predictors. Indeed, before the addition of the DA components, the model accounted for almost 72% of the variance in reading. This increased to 76% after the DA, showing that this combination of predictors accounts for a large proportion of variance in later reading. This pattern held for the ML subgroup, but in the analysis for the children with EAL, only DA Synthesis explained unique variance (1.6%). Ours is the first study to show that DA of decoding predicts significant unique variance in later reading in children with EAL, even after accounting for a range of common static predictors, as although previous studies recruited samples including children from CALD backgrounds, they did not assess the differential predictive validity of DA for these children (Dixon et al., 2023a).

The main aim of this study was to establish whether our DA of decoding could accurately identify children at risk of later reading difficulties, specifically the dyslexic reading profile. Early identification is vital to ensure that additional support is in place for children who need it (Lovett et al., 2017). In England, as in many countries across the world, children entering formal education are increasingly from CALD backgrounds. This denotes the need for a method of screening for reading difficulties that is minimally affected by children's existing knowledge of foundational language skills, which are a product of learning ability and the opportunity to learn. DA of the foundational skills involved in reading (symbol sound learning and decoding) offers a potential screening approach suitable for diverse populations. Previous research

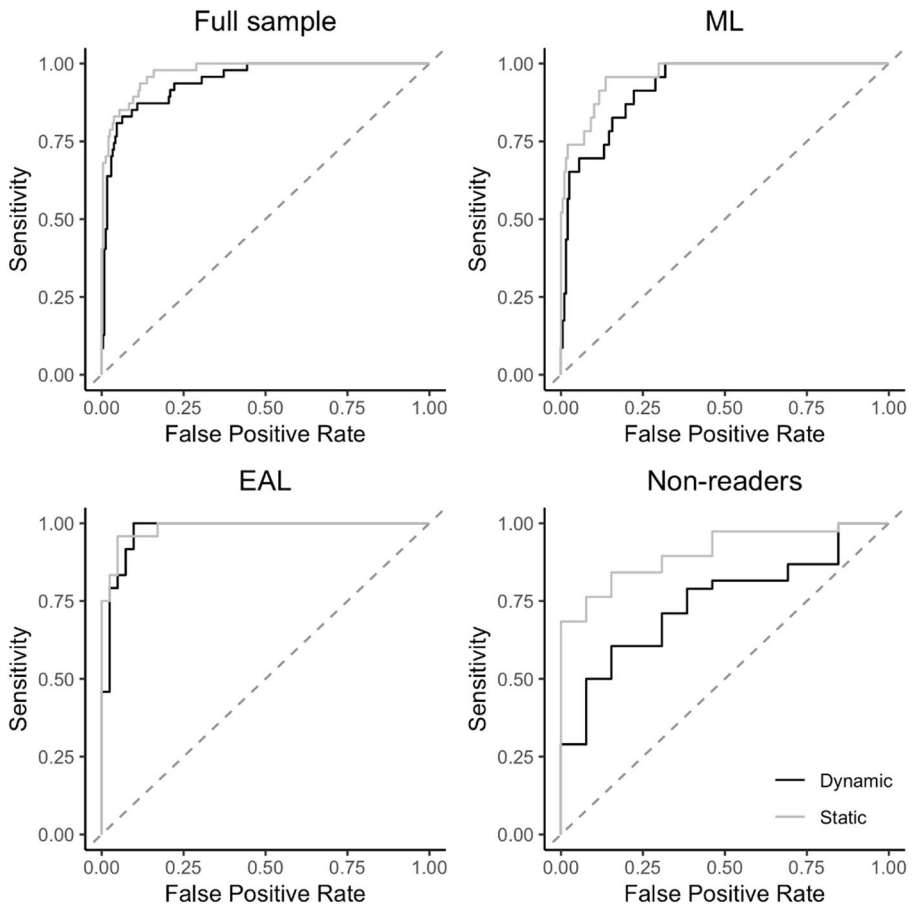


Fig. 2 ROC curves predicting risk for the dyslexic reading profile in the full sample and subsamples

has found that DAs of decoding result in excellent classification accuracy when used alone or with static predictors. Our DA of decoding produced similar results, outstanding classification accuracy when used alone and when entered after the demographic variables. Adding the dynamic components to the static predictors did not significantly improve classification accuracy further for the whole sample, unlike in some previous studies, but it did improve specificity. This is important because wrongly identifying children as at-risk will absorb precious resources for additional support. That the addition of the DA of decoding did not further improve classification accuracy is likely due to accuracy already being near perfect. That static measures of LSK and early word reading at t1 were able to accurately predict later reading status may be because our first testing timepoint was delayed, meaning children had already received around 8 months of formal literacy instruction. This contrasts with previous studies, such as Gellert and Elbro (2017b), where the children were in kindergarten and in Denmark children are not exposed to formal literacy instruction until grade 1. To explore the issue of timing we identified a group of non-readers at t1. With this sub-

Table 5 Logistic models predicting t2 at-risk status for non-readers

Model	Step	Predictor	χ^2	R^2	ΔR^2	AUC	Sen.	Spec.	TN	TP	FN	FP	FPR	FNR
Covariates	1	SES	0.59	0.017	0.017	0.397	0.974	0.231	3	37	1	10	0.769	0.026
	2	EAL	1.18	0.050	0.033	0.540	0.974	0.231	3	37	1	10	0.769	0.026
	3	Matrices	0.06	0.052	0.002	0.574	0.974	0.308	4	37	1	9	0.692	0.026
	4	BPVS	0.01	0.052	0.000	0.583	0.868	0.462	6	33	5	7	0.538	0.132
Static	5	Alliteration	2.21	0.113	0.060	0.692	0.711	0.692	9	27	11	4	0.308	0.289
	6	LSK	18.42**	0.525	0.413	0.895	0.684	1.00	13	26	12	0	0.000	0.316
	7	RAN	0.71	0.538	0.013	0.907	0.842	0.846	11	32	6	2	0.154	0.158
Dynamic	5	DA learning	7.54**	0.247	0.195	0.745	0.605	0.846	11	23	15	2	0.154	0.395
	6	DA synthesis	6.54*	0.395	0.148	0.852	0.737	0.923	12	28	10	1	0.077	0.263
Dynamic	8	DA learning	3.43	0.599	0.061	0.923	0.868	0.923	12	33	5	1	0.077	0.132
After static	9	DA synthesis	9.77**	0.751	0.152	0.953	0.789	1.00	13	30	8	0	0.000	0.211

AUC=area under receiver operating curve; TN=true negatives ($n=13$); TP=true positives ($n=38$); FN=false negatives; FP=false positives; FPR=false positive rate; FNR=false negative rate; * $p < .05$; ** $p < .01$; R^2 =Nagelkerke

sample, adding the dynamic components to the static predictors improved screening accuracy. In fact, when the DA of decoding was used alone it identified 95% of the at-risk readers.

Although DA has been suggested to be particularly useful as a screener in CALD populations, only two studies, based on the same sample, have tested this directly. Petersen et al. (2016) found that their DA of decoding correctly identified more poor readers in a bilingual sample compared to the main sample and adding it to static measures increased specificity. We also found this to be the case, with the DA components identifying all the at-risk children with EAL but 91% of the ML children, and when added to the static measures perfect classification accuracy was achieved. This finding suggests that the DA of decoding is especially valuable for identifying CALD children who may be at risk of reading difficulties, enabling targeted allocation of educational resources to those who most require support from the very start of schooling. The additional benefit of our DA of decoding for children with EAL and those in the earliest stages of learning to read may initially appear modest; however, our findings are derived from a small sample within a much larger population and were obtained using a conservative analytical approach. Specifically, our comprehensive battery of assessments represents an idealised research scenario rather than the limited information typically available to schools. When scaled to the broader population level, the statistical advantages are more substantial. Taking children with EAL in England as an illustrative example, adding the DA of decoding task to static measures identified all 24 (100%) of the EAL pupils who later became the poorest readers in our sample. By comparison, static measures alone identified 23 out of 24 pupils (96%). While in our smaller sample of 65 EAL pupils this difference amounts to only one child, scaling this finding to all reception-aged children with EAL in England (estimated at 99,619, DfE, 2024) of whom our data suggest 37% (36,859) will experience reading difficulties - a difference of 4% translates into approximately 1,475 children correctly identified using our DA of decoding who would otherwise have been missed by traditional screeners. Additionally, incorporating the DA reduced misidentification, preventing two EAL children in our sample from being incorrectly classified as at-risk readers. At a national scale, this corresponds to 3,138 reception-aged children with EAL potentially being spared unnecessary additional support. This highlights a significant limitation of static measures, which tend to overidentify CALD children as being at risk for dyslexia. Our DA of decoding task can pinpoint whether poorer letter-sound learning and blending being observed in the classroom is the result of a learning difficulty or less opportunity to learn, thus reducing the proportion of CALD children requiring further assessment and targeted support.

An additional consideration is whether administering the DA of decoding earlier in the school year could enhance its predictive value. In our study, combining a static measure of LSK with demographic variables correctly identified 26 out of 38 at-risk readers (68%). Adding the dynamic components increased accurate identification to 30 out of 38 children (79%). Although this difference corresponds to only four pupils in our sample, when scaled up to the total population of children currently in reception in England (approximately 618,397, DfE, 2024), around a quarter of whom (154,599 pupils, DfE, 2024) will leave school not meeting expected reading standards, the impact becomes substantial. Specifically, this improvement would

Table 6 Logistic models predicting at-risk status for the full sample and all groups using dynamic components only

	Step	Predictor	χ^2	R^2	ΔR^2	AUC	Sen.	Spec.	TN	TP	FN	FP	FPR	FNR
Full sample (<i>n</i> =286)	1	DA learning	25.4**	0.144	0.144	0.741	0.723	0.715	171	34	13	68	0.285	0.277
	2	DA synthesis	103.13**	0.613	0.469	0.931	0.851	0.908	217	40	7	22	0.092	0.149
ML group (<i>n</i> =221)	1	DA learning	10.55**	0.096	0.096	0.675	0.609	0.692	137	14	9	61	0.308	0.391
	2	DA synthesis	45.88**	0.463	0.367	0.890	0.739	0.919	182	17	6	16	0.081	0.261
EAL group (<i>n</i> =65)	1	DA learning	11.64**	0.224	0.224	0.817	0.833	0.829	34	20	4	7	0.171	0.167
	2	DA synthesis	45.41**	0.798	0.574	0.971	0.875	0.951	39	21	3	2	0.049	0.125
Non-reader (<i>n</i> =51)	1	DA learning	4.75*	0.131	0.131	0.704	0.684	0.769	10	26	12	3	0.231	0.316
	2	DA synthesis	8.10**	0.328	0.197	0.800	0.947	0.615	8	36	2	5	0.385	0.053

AUC=area under receiver operating curve; TN=true negatives; TP=true positives; FN=false negatives; FP=false positives; FPR=false positive rate; FNR=false negative rate; * $p < .05$; ** $p < .01$; R^2 =Nagelkerke

equate to an additional 17,006 pupils potentially identified if the DA of decoding were administered earlier. Moreover, considering that the sensitivity of the DA of decoding alone approached 95%, the estimated number of additional pupils identified increases to 41,742, underscoring the significant potential of early DA implementation. Our finding that the DA of decoding significantly predicts later word reading ability reinforces the crucial role of LSK and blending during the early stages of reading development, as outlined in the Cognitive Foundations Framework (Hoover & Tunmer, 2020). Importantly, our results indicate that not only static knowledge of letter sounds, but also the ability to learn these sounds effectively in response to instruction, is essential. The effectiveness of the DA of decoding in accurately identifying children at risk of dyslexia, particularly benefiting children from CALD backgrounds and those in the earliest stages of reading acquisition, aligns with the established view that dyslexia is characterised by difficulty learning foundational reading skills through typical instructional approaches (Carroll et al., 2024). A key feature of our DA was that it was computer assisted, with the materials shown on screen and played via laptop speakers and the administrator recording performance and monitoring attention. Few studies have used computer assisted DA of decoding tasks (Aravena et al., 2016; Horbach et al., 2015, 2018), yet computerisation could have advantages over traditional hard copy tasks (Dixon et al., 2023a, 2023b). The next phase of this research will involve teachers and teaching assistants administering a web-based application version of the DA of decoding. During this phase, we will collect data on the acceptability and reliability of the DA when implemented directly by educators within a school context, without the presence of a research team.

Our study, and indeed our DA, are not without limitations. Due to circumstances beyond our control, the first timepoint was later in the reception year than originally intended. By this point the children had received approximately 8 months of formal literacy instruction, meaning that they had been taught letters and how to decode words. Consequently, most children in the sample could read. Ideally the DA of decoding should be given near the beginning of this first year of schooling, once the children have settled into their new learning environment but before literacy instruction has begun. There were a small number of children, very new to English, who could not complete the assessments, including the DA of decoding. The aim of creating a narrative around the task was to make it more fun and engaging for children, but it introduced dialogue which could have increased task difficulty for children less proficient in English. Indeed, although modest, the correlation between vocabulary and performance on our DA of decoding was stronger than that reported in the original Danish longitudinal study, in which the DA did not include a story context (Gellert & Elbro, 2017b). Future versions of the DA of decoding will seek to maintain a narrative structure but will minimise verbal demands through use of animation and non-verbal cues.

Our DA of decoding has demonstrated considerable promise as an effective screening tool for identifying young children at risk of reading difficulties upon entry into formal education, particularly those from CALD backgrounds. When used independently, the DA achieved excellent to outstanding classification accuracy, and its predictive strength improved further when combined with a static measure of LSK. A two-step or gated screening procedure first employing LSK to identify children

who have not yet acquired these foundational skills, and subsequently using the DA of decoding to determine whether this is due to a learning difficulty could provide a highly accurate and practical screening tool method for schools. Fundamentally, the DA of decoding mirrors and condenses the early phonics instruction that children receive in reception and Year 1, and a child's performance on the DA of decoding appears to reliably predict responsiveness to classroom phonics teaching.

In addition to identifying whether or not a young child has an underlying learning difficulty, the DA of decoding task can provide an indication of whether the learner is experiencing challenges with learning symbol-sound correspondences or blending sequences of sounds, or both components, and this information could be used to plan targeted support. However, we would recommend that a more in-depth analysis of a learner's profile be carried out following the screening if they appear to be experiencing difficulties. For example, *Sound Linkage: An Integrated Programme for Overcoming Reading Difficulties* (Hatcher et al., 2014). Qualitatively DA of decoding provides the administrator with a valuable opportunity to observe the learner progress through a structured condensed version of phonics instruction, and this may reveal other useful insights into their learning preferences, needs and strengths. While it was beyond the scope of the present study to examine individual differences in this way; it is certainly something that would be worth investigating in future research.

The longer term goal of this research is to develop and validate an accessible, practical DA screener that educators can easily implement in school contexts at minimal cost and with limited training requirements. In collaboration with teachers and pupils, we have already co-designed a web-based application version, ensuring it is engaging, intuitive, and compatible with existing, stable, and affordable school IT systems. The next stage involves working directly with educators to create targeted training and support materials, determining how the DA can be integrated effectively within current school assessment and monitoring procedures, and evaluating whether schools find this approach valuable. Following this, we will assess the accuracy and usability of the web-based DA when implemented by teachers in everyday classroom settings. Collectively, these steps aim to establish a robust, equitable, and scalable screening procedure to better support young learners at risk of reading difficulties, especially those traditionally underserved by static assessments.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11145-025-10758-1>.

Acknowledgements This research was supported by the Nuffield Foundation (Grant No. EDO/43441). The authors would like to thank the schools and children who took part in this research project, along with the research assistants who worked with them. The authors would also like to thank the expert advisory board who provided valuable input and feedback throughout the project.

Data availability The study's preregistration, data and analysis code are available online via the following link: https://osf.io/agr6b/overview?view_only=3ca0bfa8e0b14b4c86878b85bd217190, https://osf.io/agr6b/?view_only=a1edaaf59dde4026ac98b898f28acc80).

Declarations

Conflict of interest The authors report there are no competing interests to declare.

Ethical approval This research was conducted in accordance with the guidelines of the British Psychological Society (BPS, 2021). Ethical approval was obtained from School of Psychology Research Ethics Committee PSC-753.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2016). Predicting responsiveness to intervention in dyslexia using dynamic assessment. *Learning and Individual Differences*, 49, 209–215. <https://doi.org/10.1016/j.lindif.2016.06.024>
- Bartón, K. (2019). *MuMIn: Multi-model inference*. <https://cran.r-project.org/web/packages/MuMIn/index.html>
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bridges, M. S., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities*, 44(4), 330–338. <https://doi.org/10.1177/0022219411407863>
- Bristol Dyslexia Centre (2011). *Dyslexia Quest Screener*. Nessy. Retrieved from <https://www.nessy.com/en-gb/product/dyslexia-quest-school>
- British Psychological Society (2021). Valid Consent. <https://doi.org/10.53841/bpsrep.2021.infl80.4>
- Bujang, M. A., & Adnan, T. H. (2016). Requirements for minimum sample size for sensitivity and specificity analysis. *Journal of Clinical and Diagnostic Research*, 10(10), 1–6. <https://doi.org/10.7860/JCDR/2016/18129.8744>
- Caravolas, M., Lervåg, A., Defior, S., Malkova, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, 24, 1398–1407. <https://doi.org/10.1177/0956797612473122>
- Carroll, J. M., Holden, C., Kirby, P., Thompson, P. A., Snowling, M. J., Panel, D. D., & Rack, J. (2024). Toward a consensus on dyslexia: Findings from a Delphi study. *Journal of Child Psychology and Psychiatry*. <https://doi.org/10.1111/jcpp.14123>
- Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, 42(2), 163–176. <https://doi.org/10.1177/00222194083262>
- Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2014). Examining the predictive validity of a dynamic assessment of decoding to forecast response to tier 2 intervention. *Journal of Learning Disabilities*, 47(5), 409–423. <https://doi.org/10.1177/0022219412466703>
- Cho, E., Compton, D. L., Gilbert, J. K., Steacy, L. M., Collins, A. A., & Lindstrom, E. R. (2017). Development of first-graders' word reading skills: For whom can dynamic assessment tell Us more? *Journal of Learning Disabilities*, 50, 95–112. <https://doi.org/10.1177/0022219415599343>
- Cho, E., Compton, D. L., & Josol, C. K. (2020). Dynamic assessment as a screening tool for early identification of reading disabilities: A latent change score approach. *Reading & Writing*, 33(3), 719–739. <https://doi.org/10.1007/s11145-019-09984-1>
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102(2), 327–340. <https://doi.org/10.1037/a0018448>

- Coventry, W. L., Byrne, B., Olson, R. K., Corley, R., & Samuelsson, S. (2011). Dynamic and static assessment of phonological awareness in preschool: A behavior-genetic study. *Journal of Learning Disabilities, 44*(4), 322–329. <https://doi.org/10.1177/0022219411407862>
- Department for Education (2012). *Assessment framework for the development of the year 1 phonics screening check*. Retrieved from <https://www.gov.uk/government/publications/assessment-framework-for-the-development-of-the-year-1-phonics-screening-check>
- Department for Education (2023). *The Reading Framework*. Retrieved from https://assets.publishing.service.gov.uk/media/664f600c05e5fe28788fc437/The_reading_framework_.pdf
- Department for Education (2024). *Statistics: Schools, pupils and their characteristics*. Retrieved from <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/2023-24>
- Department for Education (2025). *Statistics: Schools, pupils and their characteristics*. Retrieved from <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/2024-25>
- Dixon, C., Oxley, E., Gellert, A. S., & Nash, H. (2023a). Dynamic assessment as A predictor of reading development: A systematic review. *Reading and Writing, 36*, 673–698. <https://doi.org/10.1007/s11145-022-10312-3>
- Dixon, C., Oxley, E., Nash, H., & Gellert, A. S. (2023b). Does dynamic assessment offer an alternative approach to identifying reading disorder? A systematic review. *Journal of Learning Disabilities, 56*(1), 1–10. <https://doi.org/10.1177/00222194221117510>
- Dunn, L., & Dunn, L., NFER (2009). *British picture vocabulary Scale-Third edition*. GL Assessment.
- Du, Z., & Hao, Y. (2020). *ReportROC: An Easy Way to Report ROC Analysis*. Retrieved from <https://cran.r-project.org/web/packages/reportROC/index.html>
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- Elbro, C., Daugaard, H. T., & Gellert, A. S. (2012). Dyslexia in a second language?—A dynamic test of reading acquisition May provide a fair answer. *Annals of Dyslexia, 62*, 172–185. <https://doi.org/10.1007/s11881-012-0071-7>
- Forum for Research in Literacy and Language, Institute of Education. (2012). *The diagnostic test of word reading processes*. GL Assessment.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Cafrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*(4), 339–347. <https://doi.org/10.1177/0022219411407864>
- Gellert, A. S., & Elbro, C. (2017a). Does a dynamic test of phonological awareness predict early reading difficulties? A longitudinal study from kindergarten through grade 1. *Journal of Learning Disabilities, 50*(3), 227–237. <https://doi.org/10.1177/0022219415609185>
- Gellert, A. S., & Elbro, C. (2017b). Try a little bit of teaching: A dynamic assessment of word decoding as a kindergarten predictor of word reading difficulties at the end of grade 1. *Scientific Studies of Reading, 21*(4), 277–291. <https://doi.org/10.1080/10888438.2017.1287187>
- Gellert, A. S., & Elbro, C. (2018). Predicting reading disabilities using dynamic assessment of decoding before and after the onset of reading instruction: A longitudinal study from kindergarten through grade 2. *Annals of Dyslexia, 68*, 126–144. <https://doi.org/10.1007/s11881-018-0159-9>
- Gibbs, S., & Bodman, S. (2014). *PhAB2 Primary: Phonological Assessment Battery 2nd Edition Primary*.
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*(2), 111–132. <https://doi.org/10.1177/0022219408326207>
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75–111. <https://doi.org/10.1037/0033-2909.124.1.75>
- Hall, C., Dahl-Leonard, K., Cho, E., Solari, E. J., Capin, P., Conner, C. L., & Kehoe, K. F. (2023). Forty years of reading intervention research for elementary students with or at risk for dyslexia: A systematic review and meta-analysis. *Reading Research Quarterly, 58*(2), 285–312. <https://doi.org/10.1002/rq.477>
- Hatcher, P. J., Duff, F. J., & Hulme, C. (2014). *Sound linkage: An integrated programme for overcoming reading difficulties*. Wiley.
- Hoover, W. A., & Tunmer, W. E. (2020). *The cognitive foundations of reading and its acquisition*. Springer International Publishing.
- Horbach, J., Scharke, W., Cröll, J., Heim, S., & Günther, T. (2015). Kindergarteners' performance in a sound-symbol paradigm predicts early reading. *Journal of Experimental Child Psychology, 139*, 256–264. <https://doi.org/10.1016/j.jecp.2015.06.007>

- Horbach, J., Weber, K., Opolony, F., Scharke, W., Radach, R., Heim, S., & Gunther, T. (2018). Performance in sound-symbol learning predicts reading performance 3 years later. *Frontiers in Psychology*, 9, 1–8. <https://doi.org/10.3389/fpsyg.2018.01716>
- Horne, J. (2020). *Rapid Digitised Dyslexia Screening for 4–15 years*. GL Assessment. Retrieved from https://support.gl-education.com/media/2910/rapid_usermanual2020_final_v3.pdf
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression* (3rd ed.). John Wiley.
- Hulme, C., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M. J. (2012). *York assessment of reading for comprehension (YARC): Early reading – Australian edition*. GL Assessment.
- Kelley, K. (2022). MBESS: The MBESS r package [Manual].
- Krenca, K., Segers, E., Chen, X., Shakory, S., Steele, J., & Verhoeven, L. (2020). Phonological specificity relates to phonological awareness and reading ability in English–French bilingual children. *Reading & Writing*, 33(2), 267–291. <https://doi.org/10.1007/s11145-019-09959-2>
- Lovett, M. W., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology*, 109(7), 889–914. <https://doi.org/10.1037/edu0000181>
- Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322. <https://doi.org/10.1037/a0026744>
- Mielke, J. (2018). Visualizing phonetic segment frequencies with density-equalizing maps. *Journal of the International Phonetic Association*, 48(2), 129–154. <https://doi.org/10.1017/S0025100317000123>
- Ministry of Housing, Communities & Local Government (2019). *English indices of deprivation*. Retrieved from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>
- Morrisroe, J. (2014). *Literacy Changes Lives 2014: A New Perspective on Health, Employment and Crime*. National Literacy Trust. Retrieved from <https://files.eric.ed.gov/fulltext/ED560667.pdf>
- O'Connor, R. E., & Jenkins, J. R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3(2), 159–197. https://doi.org/10.1207/s1532799xssr0302_4
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in Education, PISA*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Peña, E., & Halle, T. (2011). Assessing preschool dual Language learners: Traveling a multiforked road. *Child Development Perspectives*, 5(1), 28–32. <https://doi.org/10.1111/j.1750-8606.2010.00143.x>
- Peirce, J. W. (2007). PsychoPy—psychophysics software in python. *Journal of Neuroscience Methods*, 162(1–2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Petersen, D. B., Allen, M. M., & Spencer, T. D. (2016). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities*, 49(2), 200–215. <https://doi.org/10.1177/0022219414538518>
- Petersen, D. B., & Gillam, R. B. (2015). Predicting reading ability for bilingual Latino children using dynamic assessment. *Journal of Learning Disabilities*, 48(1), 3–21. <https://doi.org/10.1177/0022219413486930>
- Petersen, D. B., Gragg, S. L., & Spencer, T. D. (2018). Predicting reading problems 6 years into the future: Dynamic assessment reduces bias and increases classification accuracy. *Language, Speech, and Hearing Services in Schools*, 49(4), 875–888. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021
- Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A. M., Tolvanen, A., Torppa, M., & Lyytinen, H. (2007). Very early phonological and Language skills: Estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48(9), 923–931. <https://doi.org/10.1111/j.1469-7610.2007.01763.x>
- Raven, J. C., Court, J. H., & Raven, J. (2008). *Coloured progressive matrices and Crichton vocabulary scale*. Pearson Assessment.
- R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *Bmc Bioinformatics*, 12(1), 77. <https://doi.org/10.1186/1471-2105-12-77>
- Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J. F., Bricout, L., & Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, 52(10), 1503. <https://doi.org/10.1037/dev0000184>

- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua Non of reading acquisition. *Cognition*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Snowling, M. J., & Melby-Lervåg, M. (2016). Oral Language deficits in Familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, 142(5), 498–545. <https://doi.org/10.1037/bul0000037>
- Spector, J. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology*, 84(3), 353–363. <https://doi.org/10.1037/0022-0663.84.3.353>
- Streiner, D., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry*. <https://doi.org/10.1177/070674370705200210>
- Thompson, P. A., Hulme, C., Nash, H. M., Gooch, D., Hayiou-Thomas, E., & Snowling, M. J. (2015). Developmental dyslexia: Predicting individual risk. *Journal of Child Psychology and Psychiatry*, 56(9), 976–987. <https://doi.org/10.1111/jcpp.12412>
- Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, 12(4), 385–435. <https://doi.org/10.1023/A:1009032414088>
- Wood, E., Biggs, K., & Molnar, M. (2024a). Dynamic assessments of word reading skills in diverse school-age children: A meta-analysis. *Perspectives of the ASHA Special Interest Groups*, 9(3), 817–835. https://doi.org/10.1044/2024_PERSP-23-00262
- Wood, E., Biggs, K., & Molnar, M. (2024b). Characteristics of dynamic assessments of word reading skills and their implications for validity: A systematic review and meta-analysis. *SAGE Open*, 14(4). <https://doi.org/10.1177/21582440241300536>
- Yang, L., Li, C., Li, X., Zhai, M., An, Q., Zhang, Y., & Weng, X. (2022). Prevalence of developmental dyslexia in primary school children: A systematic review and meta-analysis. *Brain Sciences*, 12(2), 240. <https://doi.org/10.3390/brainsci12020240>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.