Article:

Nash, H., Dixon, C., Clarke, P. et al. (2025) Dynamic assessment of word learning as a predictor of vocabulary, reading comprehension and risk status for the poor comprehender reading profile. Reading and Writing, 38 (9). pp. 2633-2659. ISSN: 0922-4777

https://doi.org/10.1007/s11145-024-10603-x

Reading and Writing
https://doi.org/10.1007/s11145-024-10603-x

Dynamic assessment of word learning as a predictor of vocabulary, reading comprehension and risk status for the poor comprehender reading profile

Hannah Nash[1] · Chris Dixon[1] · Paula Clarke[2] · Emily Oxley[1] · Anna Steenberg Gellert[3] · Anna Weighall[4]

Accepted: 17 October 2024

Abstract

The Poor Comprehender (PC) reading profile is characterised by difficulty comprehending text despite age-appropriate decoding skills. Risk for this profile is typically identified through static screening instruments measuring pre-existing knowledge, which may produce biased estimates for culturally and linguistically diverse children. In contrast, Dynamic Assessment (DA) measures potential to learn new knowledge and has been shown to reduce bias in screening. To date, however, DA has not been used to identify PC reading profile risk status. Adopting a longitudinal design, we used an adapted DA of word learning to measure growth in vocabulary and reading comprehension among a diverse sample of 322 primary school children (aged 9 years at the first time point) in England over a period of 16-19 months, and to classify later PC reading profile risk status. Two separate factor scores representing phonological and semantic aspects of word learning predicted unique variance in the growth of vocabulary knowledge and reading comprehension after accounting for traditional static predictors. The DA of word learning achieved excellent classification accuracy for identifying poor comprehenders, identifying all of the poor comprehenders with EAL when added to the static assessments. Results suggest that DA of word learning may be a promising tool in screening for the PC reading profile, particularly for EAL pupils, ensuring that the risk of reading comprehension difficulties does not go unidentified in such learners.

Keywords Dynamic assessment · Vocabulary · Reading comprehension impairment

Intervention is most effective when reading difficulties are identified as early as possible (Lovett et al., 2017). Much research has examined the ability of traditional

Extended author information available on the last page of the article

Published online: 23 November 2024

screening assessments to predict future reading achievement and classify children at-risk for reading difficulties. Measures of phonological awareness, letter knowledge and rapid automatised naming are established predictors of early word recognition across a range of orthographies (Caravolas et al., 2012; Puolakanaho et al., 2007), and oral language skills are predictors of later reading comprehension (Adlof et al., 2010; Catts et al., 2016). By focusing on pre-existing knowledge and skills, traditional 'static' screening assessments are sensitive to a child's learning experiences, an issue that is problematic for those from culturally and linguistically diverse backgrounds increasingly found in many school populations (OECD, 2010; Peña & Halle, 2011). In contrast, Dynamic Assessment (DA) aims to measure a child's potential to learn from feedback and may provide a more accurate indication of an underlying learning difficulty as opposed to lack of learning opportunity (Grigorenko & Sternberg, 1998). If effective, DA may obviate the 'wait to fail' model necessitated by a Response to Intervention (RTI) framework or the 'wait and see' approach in the case of English language learners still acquiring basic linguistic proficiency (Verpalen et al., 2018). DAs of decoding and phonological awareness offer good accuracy in the identification of children who experience difficulties with word recognition (Dyslexic profile) and can offer advantages over static assessments (Bridges & Catts, 2011; Cho et al., 2020; Gellert & Elbro, 2017a, b; Petersen et al., 2018; Dixon et al., 2023a). However, it is unknown whether such classification advantages also apply in the identification of difficulties in other aspects of reading such as comprehension. In this study, we investigate the ability of a novel computer assisted DA of word learning to predict vocabulary and reading comprehension growth, and to identify risk of the Poor Comprehender (PC) reading profile in a diverse sample of primary school children in England.

According to the Simple View of Reading (SVR), decoding and linguistic comprehension represent two dissociable but related strands of reading skills and processes, which through a multiplicative relationship, are essential for successful reading comprehension (Gough & Tunmer, 1986). The discrepancy-based profile of accurate and fluent decoding coupled with comprehension difficulties, is aligned with this view (Nation, 2019) and has been described using a variety of terms, most notably, Specific Reading Comprehension Disorder (Landi & Ryherd, 2017), Reading Comprehension Impairment (Snowling, 2012), and Poor Comprehenders (Cain & Oakhill, 2006; Nation et al., 2004). There is no consensus regarding the precise criteria for the discrepant profile. Spencer and Wagner (2018, pp. 370-371) identified four main approaches: (a) a reading comprehension and decoding discrepancy, (b) approach a plus average range decoding (c) approach b plus discrepancy with chronological age expectations (d) approach c plus comprehension scores below a cut-off point. Additionally, research has not consistently used the same measures of component reading skills in operationalising these criteria, meaning that resulting profiles differ according to assessment demands (e.g., Keenan & Meenan, 2014).

Although the SVR has support from a wealth of studies that have shown decoding and linguistic comprehension to be significant predictors of reading comprehension development (Hulme et al., 2015; Language & Reading Research Consortium, 2015) the simplicity of the model may not fully convey the complex interrelationships between component skills over time. Protopapas et al. (2013) argued that the interrelationships between SVR components and vocabulary, in their longitudinal data from 8–11-year-olds in Greece, were best explained using a combination of the SVR and the Lexical Quality Hypothesis (LQH, Perfetti, 2007). According to the LQH, (Perfetti, 2007), reading relies on lexical representations of phonology (pronunciation), semantics (word meaning and grammar) and orthography (spelling). The quality of these representations is key, as are the connections between them. More recently, Nation (2019) has argued that the inclusion of an underlying language factor (which includes vocabulary and morphology) that contributes to both decoding and linguistic comprehension, would be a valuable addition to the SVR.

The PC reading profile is associated with weaknesses in oral language skills including vocabulary, morphosyntax, and listening comprehension (Landi & Ryherd, 2017). The existence of oral language difficulties prior to reading instruction and their persistence over time is suggestive of their causal role in the PC reading profile (Catts et al., 2006; Clarke et al., 2010; Elwér et al., 2013; Landi & Ryherd, 2017). This is supported by studies showing that explicit instruction in vocabulary and oral language skills has been shown to improve reading comprehension among poor comprehenders (Clarke et al., 2010; Lee & Tsai, 2017; Proctor et al., 2020). Approximately 5-10% of school-aged children present with the PC reading profile (e.g., Clarke et al., 2010; Kelso et al., 2020; Nation et al., 2010); though prevalence is also likely to increase with age (Hogan et al., 2014) and is higher among those learning English as an additional language, at 10-18%, due to the typically lower levels of English vocabulary knowledge among this population (Li et al., 2021). Despite its attendant oral language weaknesses, identification of the PC reading profile remains challenging. Compared to difficulties in reading accuracy and fluency, comprehension problems may go unnoticed in the classroom, and teachers have been shown to misclassify such students (Kelso et al., 2020; Nation et al., 2004).

The accuracy of a screening assessment is determined by examining the agreement in classification between a 'reference' or 'gold' standard such as a standardised reading assessment 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). Static screening assessments of reading and reading-related skills have been criticised for their high false positive rates as well as high false negative rates (Catts et al., 2001; Jenkins & O'Connor, 2002; Petersen et al., 2016), the latter of which is deemed to be more problematic since individuals who might benefit from intervention could easily be missed.

Children acquiring the majority language as an additional language and children from socio-economically disadvantaged backgrounds (collectively termed culturally and linguistically diverse; CLD) are likely to be at elevated risk of misclassification. Crucially, these children may score poorly on screening assessments due to lack of English language learning opportunity rather than an underlying learning difficulty (Tzuriel, 2000). In the UK both language learner status and socio-economic status (SES) predict educational achievement during primary school, and these risk factors may interact (Strand et al., 2015). In the UK, over 4 million children are living in poverty and some regions are seeing rapid increases in this number (Stone, 2023). Alongside this, there has also been a gradual increase in the number of children for whom English is an Additional Language (EAL); across primary and secondary schools just over 20% of children have EAL (Department for Education, 2023). These increases corroborate the pressing need for less biased forms of screening assessment.

DA is a measure of learning potential or latent capacity rather than developed ability (Grigorenko & Sternberg, 1998) and is therefore less sensitive to prior learning experiences. DA involves an element of teaching or mastery learning with explicit feedback, and data are interpreted in terms of within-subject changes (Haywood & Lidz, 2007). Although not novel, there has been increasing interest in DA of language and reading skills in recent years. DA of spoken word learning was precipitated by the need for speech and language therapists to establish whether young children's language difficulties represented a clinically significant impairment that was likely to persist. Research has shown that performance on DAs of word learning correlates with concurrent standardized measures of vocabulary knowledge and with change in vocabulary knowledge over time, in typically developing children (e.g., Camilleri & Botting, 2013; Gellert & Elbro, 2013) and those with low language skills (Camilleri & Law, 2014). However, DA of word learning has most often been used to distinguish between children whose language is delayed, but who are capable of learning, and those with an underlying language disorder. This has been demonstrated in monolingual (e.g., Camilleri & Law, 2007; Hasson et al., 2013) and bilingual children (e.g., Kapantzoglou et al., 2012; Matrat et al., 2023). Recently, Matrat et al. (2023) found particularly high diagnostic accuracy using a DA that captured learning the phonological form of new words in young children.

While some research has examined the contribution of DA to the prediction of reading comprehension (e.g., Wolter & Pike, 2015), to date DA has not been used to classify risk for the PC reading profile. Gruhn et al. (2020) have speculated that DAs could help to provide a more nuanced understanding of the learning profiles of individuals experiencing reading comprehension difficulties as they may tap additional abilities not captured by static measures. It is possible that accurate classification of reading comprehension difficulties could be obtained from a DA of reading comprehension itself (e.g., training in comprehension strategies). However, this may be prohibited by the length of time required to administer a DA of reading comprehension (e.g., up to 60 min in Elleman et al., 2011), as well as the relatively high level of linguistic proficiency needed to access passages of text. Instead, given the critical role of vocabulary knowledge in reading comprehension and reading comprehension difficulties, DA of word learning may hold promise in identifying risk for the PC reading profile. The ability to learn meanings of derived words (e.g., equalize, oceanaut) through graduated prompts has been found to correlate with static measures of vocabulary and reading comprehension (Wolte and Pike, 2015) and performance on DA of word learning appears to be less strongly associated with language learning background (Camilleri & Law, 2007) and SES (Burton & Watkins, 2007) than standardised static measures of vocabulary knowledge. Finally, children fitting the PC profile perform similarly to good comprehenders in acquiring phonological information about novel words (labels) but experience difficulty retaining semantic information in short-term follow-up (Nation et al., 2007; Ricketts et al., 2008).

The present study is a conceptual replication of a longitudinal dynamic word learning study conducted in Denmark with 90 children aged 9 years 7 months (Gellert & Elbro, 2013). Based upon the learning paradigm employed by Nation et al. (2007), participants were exposed to six novel labels depicting animals, accompanied by four attributes, e.g., *Targeli*, *a fat*, *white*, *spotted cow*. After initial exposure to each novel word and its attributes, phonological learning was measured through a learning-to-criterion training phase with corrective feedback in which participants were asked to provide the novel name for each animal as its picture was displayed. Three post-tests followed: definition knowledge, (examiner provided novel name and child recalled relevant attributes), immediate recall (child provided novel name after hearing definition) and immediate recognition, (child presented with picture of novel animal and three distractors and asked to select the one matching the word spoken by examiner). A composite variable of vocabulary training and immediate recall (representing phonological aspects of word learning) explained 1-6% unique variance in receptive and expressive vocabulary knowledge measured nine months later after controlling for vocabulary knowledge at the first timepoint (the autoregressor) and general cognitive ability. This finding has been replicated using a computerised version of the DA in a sample of children learning EAL in England (Oxley, 2019).

Present study

DA holds potential as a less biased form of screening for reading accuracy difficulties but has not been used to classify risk for the PC profile. To examine the predictive and prognostic value of a DA of word learning for this purpose we posed three research questions and formed corresponding hypotheses:

1. Does learning performance on a computer assisted DA of word learning correlate less strongly than static assessments with SES and English language proficiency? Given that performance on a DA of word learning has been shown to be less strongly related to language learning background (Camilleri & Law, 2007) and SES (Burton & Watkins, 2007), our first hypothesis was that the DA of word learning would not correlate strongly with SES or English language proficiency at the first time point.
2. Does learning performance on a computer assisted DA of word learning predict growth in vocabulary and reading comprehension ability over time? We hypothesised that the DA of word learning would predict growth in vocabulary over time as found previously (Gellert & Elbro, 2013) and that it would predict growth in reading comprehension (as suggested by the findings of Wolter & Pike, 2015). In addition to planned analyses for the whole sample, we conducted exploratory analyses for monolingual (ML) and EAL children separately, to test whether any predictive relationship differed between the two groups.
3. Can a computer assisted DA of word learning accurately screen for later reading comprehension difficulties, specifically the PC profile? Based on the presence of

vocabulary weaknesses (Landi & Ryherd, 2017) and word learning difficulties (e.g., Ricketts et al., 2008) in children with the PC profile, our third hypothesis was that our DA of word learning would differentiate good/poor comprehenders, more specifically that it would achieve clinically acceptable levels of accuracy in identifying children at risk for of the PC profile at the second time point. We tested this for the whole sample and for the ML and EAL groups separately, as there is some suggestion that DA of decoding improves classification accuracy more in bilingual learners (Petersen et al., 2016, 2018). Although there are no absolute thresholds, sensitivity (proportion of affected children correctly identified) and specificity (proportion of non-affected children correctly identified) upwards of 80% has been suggested as acceptable, while an AUC (which uses sensitivity and specificity to work out the likelihood of correct identification) 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).

In addition to the three main questions, we included two exploratory questions relating to screening accuracy; 4) How does the DA of word learning compare in terms of screening accuracy to static assessments? 5) Does the DA of word learning improve screening accuracy when added to static assessments?

Method

Participants

Based on guidance in Bujang and Adnan (2016), we aimed to recruit a minimum sample of 220 children to afford adequate statistical power to detect a 10% prevalence rate of the PC profile with at least 80% sensitivity. To recruit a representative sample, we stratified local state-maintained (government-funded) primary schools into high/low SES and high/low EAL enrolment according to Department for Education statistics. Participating schools were asked to share data on demographic characteristics as well as EAL status and Special Educational Needs (SEN) status (yes/no) of their pupils. SES was measured using the Income Deprivation Affecting Children Index (IDACI) ranks derived from participants' home post codes; the rank indicates the proportion of children in that area living in low-income families with a higher proportion reflecting lower SES. The sample at the first time point (t1) consisted of 414 children ($n = 226$ male) with a mean age of 110.45 months (SD = 3.47) or 9 years 2 months, recruited from seven schools in Leeds, West Yorkshire. The sample included 58 children with SEN and 145 children learning EAL. Based on a home language questionnaire, EAL learners spoke a total of 33 different languages, primarily Bengali (32%), Romanian (11%), and Urdu (11%). Confounding between EAL status and IDACI ranks was evident, with significantly lower ranks among EAL learners (see Table 1).

Children were scheduled to be reassessed on measures of vocabulary and reading comprehension at a second time point (t2) 12 months later. However, due to COVID-19-related school disruption, this interval was elongated to 16-19 months. Reasons

Table 1 Descriptive statistics at t1 and t2

Measure	ML	EAL	D	*p*
SES	17126.09 (10443.55)	7846.9 (6649.87)	1.058	0.000
NVIQ	27.26 (4.84)	24.2 (5.37)	0.598	0.000
DA: initial exposure	18.96 (3.97)	17.59 (4.63)	0.316	0.014
DA: training	40.26 (10.64)	39.02 (12.11)	0.108	0.606
DA: definition	12.3 (3.63)	10.94 (3.81)	0.365	0.003
DA: recall	4.83 (1.33)	4.59 (1.47)	0.175	0.377
DA: recognition	5.7 (0.74)	5.69 (0.81)	0.017	0.874
DA: phonological factor	0.06 (0.94)	-0.1 (1.09)	0.155	0.426
DA: semantic factor	0.13 (0.97)	-0.24 (1.01)	0.368	0.003
Reading accuracy	55.33 (8.57)	50.88 (10.9)	0.454	0.000
BPVS t1	120.25 (15.65)	100.08 (23.12)	1.020	0.000
BPVS t2[a]	129.89 (15.36)	113.69 (21.03)	0.878	0.000
RC t1	58.32 (7.96)	50.01 (10.9)	0.869	0.000
RC t2[a]	61.46 (8.48)	54.85 (9.44)	0.736	0.000

Note: Statistics shown in mean and (SD) for ML and EAL subgroups; D = Cohen's D with Hedge's correction; p = p-value of independent groups t-test with Holm correctio; SES = IDACI rank; NVIQ = CPM; DA = dynamic word learning assessment; RC = YARC-P ability score; $n = 414$ unless stated; [a]$n = 322$

for attrition at t2 were children moving school ($n = 25$; 27.2%) and being absent ($n = 11$; 12%). A further 56 children (60.9%) in one school could not be re-assessed due to a nationwide lockdown in January 2021 and associated school closures. The final sample consisted of 322 children at t2 ($n = 174$ male; $n = 124$ EAL; $n = 44$ SEN; mean age in months = 128.10 [SD = 3.52] or 10 years 8 months). We compared the demographic profiles of the children who were missing at t2 to those of the children who remained in the study (missingness data can be found in Supplementary Material). The two groups did not differ significantly in terms of gender split, proportion of children with SEN or in chronological age. There were, however, significant differences in terms of the proportion of children with EAL in each group and average SES; there were more monolingual children and children from higher SES backgrounds in the missing group. This reflects the fact that most of the missing children came from a school that had been selected for the study because it was high SES/low EAL. This served to increase the proportion of children with EAL in our final sample from 35 to 38.5% and reduce the average IDACI rank.

Measures and procedures

At t1 children were assessed on the DA of word learning as well as static measures of nonverbal reasoning, vocabulary, passage reading accuracy and comprehension. The test battery was split into two sessions on separate days each lasting approximately 30 min. Task order was constant across all sessions. Ethical approval was obtained from the School of Psychology Research Ethics Committee at the University of Leeds. Parental consent was sought on an opt-out basis while ongoing child assent was sought by researchers during data collection. The test battery at t2 consisted of static vocabulary, reading accuracy and comprehension measures. All assessments were carried out individually in quiet areas of the school by members of the research team and a small number of research assistants (undergraduate and postgraduate students) who had received 2.5 h of training.

Static assessments

Nonverbal reasoning. Raven's coloured progressive matrices (CPM; Raven et al., 2008) are a series of diagrammatic puzzles, in which examinees are asked to select one of six solutions to 'fit' the puzzle. The CPM is standardised on a UK sample of children aged 4 to 11 years and reports a split-half reliability of $r = 97$.

Vocabulary. The British Picture Vocabulary Scale-III (BPVS-3; Dunn et al., 2009) is a multiple-choice measure of receptive vocabulary knowledge. Examinees are asked to match pictures to words spoken by the examiner and testing discontinues once eight or more errors are made in a set of 12 items. The BPVS-3 is standardised on a UK sample of individuals aged 3 to 16 years. 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 $r = 94$.

Reading accuracy and comprehension. The York Assessment of Reading for Comprehension-Primary (YARC-P; Snowling et al., 2011) assesses passage reading accuracy and comprehension. The YARC-P consists of six fiction and six non-fiction passages, and reading scores are calculated from the two highest passages attempted. The starter passage was determined by each participant's score on the Single Word Reading Test. Comprehension questions assess both literal and inferential understanding. The YARC-P is standardised on a large sample of primary school children in the UK and reports internal reliability of $r = 71\text{-}0.84$. In addition to age-standardised scores, the YARC-P derives ability scores (ranging from 1 to 91) which take passage difficulty into account (as the two highest passages can be different for different children), these can then be converted into standardised scores based on age. In the present study, standardised scores for passage accuracy and comprehension were used to determine 'poor comprehender' status (discussed below), while reading accuracy and comprehension ability scores were used in hierarchical regression models and reading comprehension ability score as an autoregressor in logistic regression models, this was in place of children's raw scores which would differ depending on the two highest passages completed.

Dynamic assessment of word learning

DA may take several forms, but typically involves restructuring the test situation, learning within the test, or metacognitive intervention (Haywood & Lidz, 2007). We operationalised DA as learning within the test, whereby examinees would be given multiple opportunities to learn phonological and semantic information and the opportunity to act upon feedback. We adapted the DA of Gellert and Elbro (2013), piloting changes with a sample of similarly aged children prior to data collection. Firstly, given evidence for the superior ability of bilingual children to map novel labels on to known objects (Kaushanskaya et al., 2014), we sought to remove this potential source of bias by using unknown referents (aliens). Visual stimuli depicting alien referents were sourced and adapted from Gupta et al. (2004). Secondly, based on a replication by Oxley (2019), we adapted labels from the original Danish study, for instance changing mafyk to masik due to confusion between /f/ and /th/, and simplifying the three-syllable label *targeli* to the two-syllable *tarom*. Thirdly, to prevent children from merely describing visual characteristics, a third 'unseen' attribute (e.g., brave, organised) was also included in definitions (see Supplementary Material for all names and attributes used).

Through colourful illustrations and a narrative script 'Galaxy Explorers' recorded by a female native English speaker, participants were asked to help 'Commander Stan McKenzie' document the names and attributes of six novel aliens encountered on a space mission. The task was administered using DMDX software (Forster & Forster, 2003) with stimuli presented in two blocks of three aliens. The test administrator recorded participants' responses with keystrokes. To reduce fatigue, children were given a break after completing the first block (the entire assessment took on average 25 minutes to complete). The structure of the assessment is presented in Fig. 1 and each phase is described below. Reliability of each phase is reported using the omega coefficient (ω) and its 95% confidence interval provided by the MBESS package (Kelley, 2022) in R (R Core Team, 2021). Omega has been shown to be practical alternative to alpha, addressing the deficiencies of the former (Dunn et al., 2014).

Initial exposure. Children were presented with a picture of each alien and asked to repeat the novel name and accompanying definition, consisting of three adjectives, for instance '*Salu*, a green, spotted, bossy alien'. There was no time limit, and no corrective feedback was given. Reliability of initial exposure was $? = 0.80$, [0.76, 0.84].

Vocabulary training. Immediately after initial exposure, aliens were presented in a randomised order and children were asked to provide the corresponding novel name. If the child gave the correct response they then heard 'that's right, it's *Tarom*, a pink, tentacled, brave alien'. For incorrect responses, the child heard 'that's not quite right, it's *Tarom*, a pink, tentacled, brave alien'. Training discontinued when participants correctly named all three aliens in one block on two consecutive trials, or when they reached 10 trials. One point was awarded for each item named correctly, with a maximum of 30 points per block. Reliability of vocabulary training was $? = 0.94$, [0.93, 0.94].

Following the training trials there were three post-tests, described below and administered in that order, during which no feedback was provided.

Definition knowledge. Children were assessed on their knowledge of the alien attributes. For instance, children were asked 'How would you describe *Tarom*?' No visual stimuli were provided. Each correctly named attribute received one point, up to a maximum of three points per alien and nine points per block. Reliability of the definition knowledge post-test was $? = 0.78$, [0.75, 0.81].

Immediate recall. Children heard definitions of the aliens and were asked to name the alien, for instance 'What was the name of the red, bearded, lazy alien?'. No visual stimuli were provided. The score was one point per correct response and the maximum score was therefore three per block. Reliability of immediate recall was $? = 0.60$, [0.53, 0.67].

Immediate recognition. Children were presented with pictures of the three target aliens and three distractors, and then asked to point to a particular target. The pictures

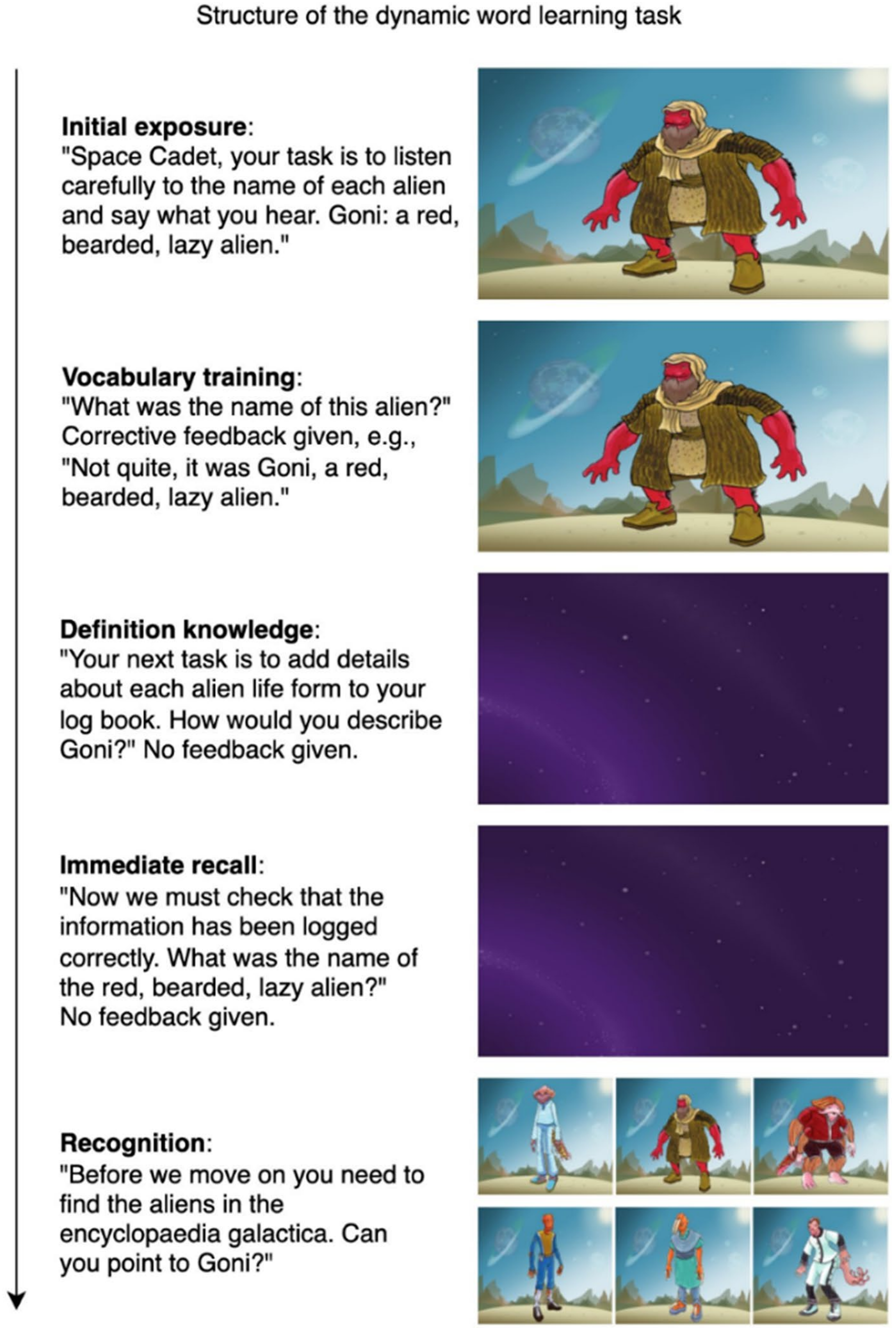

Fig. 1 Structure of the dynamic assessment of word learning assessment 'Galaxy Explorers'

were then randomly ordered, and the procedure was repeated for the remaining novel words. No feedback was provided, and the score was the number of targets correctly chosen, with a maximum of three points per block. Reliability of immediate recognition was $? = 0.70$, [0.62, 0.78].

Analysis

To address hypothesis one, correlation analysis was carried out to assess relationships between dynamic assessment variables, static assessments, and covariates of interest (SES and NVIQ). To address hypothesis two, growth in vocabulary (BPVS raw scores) and reading comprehension (YARC-P ability scores) between the two time points was estimated using linear mixed effects (LME) models in R (data, analysis code, and the study's preregistration are available online via the following link: https://osf.io/5zkem/?view_only=dce7f66d870b410d910697738cc1ad94). In most cases, likelihood ratio tests suggested significantly improved fit for unconditional linear models containing random intercept terms for both subject (n = 322) and school (n = 7); however, the random intercept for school was dropped in two language subgroup models (BPVS growth for ML learners and YARC growth for EAL learners) in order to deal with model non-convergence. Random slopes for time were not appropriate due to only two repeated measurements. LME models were fitted using the lme4 package (Bates et al., 2015) and marginal pseudo R-squared was estimated using the MuMIn package (Bartón, 2019). All LME models were fitted with maximum likelihood estimation to allow comparison of nested models, and all met assumptions of nonconstant error variance and approximately normal distribution of residuals. Continuous predictors were scaled and centred (except for BPVS-3 and YARC-P which were the only variables to be administered at two time points), while EAL status and PC reading profile risk status were dummy coded with reference categories of monolingual (ML) and non-PC reading profile risk, respectively. We employed a 'step-up' model building process for linear and logistic regression models. Covariates were added initially, followed by either dynamic or static assessments such that these were allowed to compete on an even basis after the contribution of covariates. In a final exploratory step, we then fitted 'final' models with dynamic variables entered after covariates and static predictors and repeated this process for ML and EAL subgroups separately. Statistical inference of predictors at each step was estimated using the results of likelihood ratio tests.

To address hypothesis three, we estimated the ability of static and dynamic variables to classify children at risk for the PC reading profile with logistic regression models using the glm() function in R. We defined PC reading profile status at t2 as a YARC-P comprehension standard score of < -1 SD as well as a discrepancy of at least 1 SD between YARC-P reading accuracy and comprehension standard scores. Classification accuracy was assessed using AUC, sensitivity, specificity, false positive (FPR) and false negative rates (FNR), computed from the reportROC package (Du & Hao, 2020). We compared AUC values at the final step of static and dynamic models using DeLong's method with the roc.test() function of the pROC package (Robin et al., 2011). The total sample size in all statistical models presented here is $n = 320$ due to one child missing a CPM score and one child with no valid IDACI rank.

Gellert and Elbro (2013) predicted growth in vocabulary knowledge with phonological (vocabulary training and recall) and semantic (definition) aspects of the DA of word learning. In a similar effort to simplify our analysis, we conducted principal components analysis (PCA) on the phases of the DA: vocabulary training, definition, and recall. Initial exposure was excluded due to its resemblance to a nonword repetition task, and recognition was excluded due to a considerable ceiling effect (as was the case in Gellert & Elbro, 2013). We extracted two components, with training and recall loading highly on a first 'phonological' factor, and definition loading highly on a second 'semantic' factor. The three remaining variables indicated an acceptable level of sampling adequacy, with all Keyser-Meyer-Olkin values ≥ 0.61. A two-factor solution was preferred, with eigenvalues of 1.61 and 1.00 for the first and second factor, respectively, explaining 87% of the variance. As factors correlated $r =$ .45, oblique factor rotation was applied to aid interpretation. Vocabulary training and recall loaded onto a first factor (0.92 and 0.87 respectively) and definitions loaded onto a second factor (1.00). Thus, the first factor represents phonological learning, the tasks loading onto it prompt for the alien names, while the second factor represents semantic learning, the definition task prompts for alien attributes. These factor scores were entered separately into statistical models as DA factors of word learning.

Results

Descriptive statistics are presented in Table 1 and intercorrelations in Table 2. Between the two time points children made significant growth in receptive vocabulary (raw scores; $t(321) = -21.65$, $p < .001$) and reading comprehension (ability scores; $t(321) = -9.87$, $p < .001$). This applied equally to EAL and ML subgroups. The largest standardised mean differences between the two language groups were found in static assessments of vocabulary and reading comprehension at t1 ($d = 1.02$ and 0.87, respectively). In contrast, group differences on DA factors were much reduced in magnitude, ranging from $d = 0.02$ for recognition to $d = 0.37$ for definition score (Table 1).

Correlations between the DA factors, English language proficiency, and SES

In terms of the first hypothesis, the raw DA scores correlated only weakly with SES (*rs* = 0.03–0.12) and weakly to moderately with English receptive vocabulary at t1 (*rs* = .09–.45). Similarly, DA factor scores correlated only weakly with SES (*rs* = .09 & .13), and although correlations with vocabulary were somewhat stronger, these were only of moderate strength (*rs* = .36 & .44). Vocabulary and reading comprehension at the first time point had stronger correlations with SES (*rs*. .34 and .32 respectively) and there were strong relationships between vocabulary and reading comprehension at both t1 ($r = .78$) and t2 ($r = .71$). Also of note was the high degree of stability between the two time points in vocabulary ($r = .89$) and reading comprehension ($r = .73$).

Table 2 Descriptive statistics and correlations for static and dynamic variables at t1 and t2

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. SES														
2. NVIQ	0.30**													
3. DA: initial exposure	0.04	0.29**												
4. DA: training	0.12*	0.24**	0.29**											
5. DA: definition	0.09	0.32**	0.39**	0.39**										
6. DA: recall	0.12*	0.19**	0.26**	0.61**	0.42**									
7. DA: recognition	0.03	0.07	0.07	0.23**	0.27**	0.37**								
8. DA: phonological factor	0.13**	0.24**	0.30**	0.90**	0.45**	0.89**	0.33**							
9. DA: semantic factor	0.09	0.32**	0.38**	0.37**	1.00**	0.44**	0.27**	0.45**						
10. Reading accuracy	0.21**	0.33**	0.33**	0.57**	0.25**	0.37**	0.11*	0.52**	0.24**					
11. BPVS t1	0.34**	0.52**	0.45**	0.39**	0.45**	0.26**	0.09	0.36**	0.44**	0.49**				
12. BPVS t2	0.37**	0.58**	0.63**	0.39**	0.46**	0.28**	0.10	0.37**	0.45**	0.47**	0.89**			
13. RC t1	0.32**	0.47**	0.38**	0.41**	0.42**	0.26**	0.11*	0.37**	0.42**	0.55**	0.78**	0.75**		
14. RC t2	0.30**	0.47**	0.53**	0.33**	0.38**	0.24**	0.11*	0.31**	0.37**	0.44**	0.72**	0.71**	0.73**	

Note: SES = IDACI rank; Matrices = CPM; DA = dynamic word learning assessment; RC = YARC-P ability score; $n = 414$ unless stated; [a]$n = 413$, [b]$n = 322$; ** $p < .01$, * $p < .05$

The predictive validity of DA factors in vocabulary and reading comprehension growth

LME models estimating growth in vocabulary are presented in Tables 3 and 4. The results of likelihood ratio tests indicated that EAL predicted vocabulary growth, however SES was not a significant predictor. Both the phonological and semantic DA factors were significant predictors of vocabulary growth when entered directly after the covariates (Table 3), accounting for 4.4% and 2.4% of variance respectively. Both DA factors remained significant predictors in an exploratory final model (Table 4) when entered at steps 5 and 6 after covariates and a vocabulary autoregressor. This was also the case for both language groups separately (Table 4): however, this effect was slightly more pronounced in the EAL group (7.3-8.5% of variance) compared to the ML group (2.2-3.0% of variance).

In terms of predicting growth in reading comprehension (Tables 5 and 6), again EAL significantly predicted growth and SES did not. Both DA factor scores were significant predictors of reading comprehension growth when entered directly after covariates, accounting for 3.8% and 2.6% of variance, respectively (Table 5). In an exploratory final model, however, only the DA semantic factor score remained significant at step 8 when entered after covariates, static predictors, and DA phonological factor score, accounting for 0.5% of variance (Table 6). Results varied by language subgroup: for ML learners, neither DA factor was a significant predictor, but for EAL learners the DA semantic factor accounted for 1.3% of unique variance.

Classifying risk status for the PC reading profile

We identified 20 children at risk of the PC reading profile, representing an overall prevalence rate of 6.25%. EAL learners were significantly more likely to be at risk ($n = 14$; 11.38%) compared to their ML peers ($n = 6$; 3.05%) ($\chi^2(1) = 7.57$, $p = .006$). At t2, children in the PC group did not differ significantly from their typically developing (TD) peers in YARC-P reading accuracy (standard score means: PC: 102.2 [range: 89-115]; TD: 99.4; $t(29.38) = -1.76$, $p = .088$, $d = -0.29$) but performed significantly lower in YARC-P comprehension (PC: 81.0 [range: 69-84]; TD: 98.6; $t(47.27) = 16.16$, $p < .001$, $d = 2.04$). Logistic regression models predicting PC risk status are presented in Tables 7 and 8 and ROC curves are presented in Fig. 2 below.

A dynamic model (both DA factor scores entered after all covariates) accounted for 19.7% of the variance in predicting PC risk status (step 5, Table 7). The model achieved excellent classification accuracy (AUC = 0.818) though with poorer specificity (0.643 with 107 false positives representing a 35.7% False Positive Ratio, FPR). In contrast, a model with only covariates and static predictors accounted for 33.2% of the variance in predicting later PC risk status (step 6, Table 7). This model represented better classification accuracy overall with an AUC of 0.878, although this came with a higher FPR of 37.7% (113 false positive cases). A comparison of AUC for the static and dynamic step-up models in Table 7 revealed no statistically significant difference in classification accuracy ($\chi^2(6) = 17.87$, $z = -1.43$, $p = 151$). In a final exploratory model, both DA factor scores were entered last after covariates and static predictors, where neither was statistically significant (Table 8). Comparing classifica-

Table 3 Linear mixed effects step-up models predicting growth in vocabulary knowledge for full sample and language subgroups

	Model	Step	Predictor	χ^2	R^2	ΔR^2
Full sample	Covariates	1	SES	1.81	0.010	0.010
(n=320)		2	EAL	16.12**	0.117	0.107
		3	NVIQ	83.73**	0.351	0.235
	Static	4	BPVS t1	286.94**	0.430	0.079
	Dynamic	4	DA Phon	29.46**	0.395	0.044
		5	DA Sem	21.05**	0.418	0.024
ML group	Covariates	1	SES	12.75**	0.053	0.053
(n=197)		2	NVIQ	74.17**	0.301	0.248
	Static	3	BPVS t1	203.78**	0.403	0.102
	Dynamic	3	DA Phon	11.17**	0.331	0.030
		4	DA Sem	8.7**	0.353	0.022
EAL group	Covariates	1	SES	1.29	0.012	0.012
(n=123)		2	NVIQ	20.08**	0.152	0.140
	Static	3	BPVS t1	100.14**	0.239	0.088
	Dynamic	3	DA Phon	16.73**	0.236	0.085
		4	DA Sem	10.67**	0.309	0.073

Note: * $p < .05$; ** $p < .01$; R^2 = marginal pseudo R-squared; ΔR^2 = change in pseudo R-squared; beta (SE) from final step; SES = IDACI rank

Table 4 Linear mixed effects final models predicting growth in vocabulary knowledge for full sample and language subgroups

	Step	Predictor	χ^2	R^2	ΔR^2	Beta (SE)
Full sample	1	SES	1.81	0.010	0.010	1.76 (0.97)
(n=320)	2	EAL	16.12**	0.117	0.107	-8.78 (2.03)
	3	NVIQ	83.73**	0.351	0.235	6.89 (0.85)
	4	BPVS t1	286.94**	0.430	0.079	11.51 (0.53)
	5	DA Phon	29.46**	0.473	0.044	2.74 (0.85)
	6	DA Sem	21.05**	0.498	0.024	4.09 (0.87)
ML group	1	SES	12.75**	0.053	0.053	1.88 (0.82)
(n=197)	2	NVIQ	74.17**	0.301	0.248	7.41 (0.97)
	3	BPVS t1	203.78**	0.403	0.102	10.55 (0.56)
	4	DA Phon	11.17**	0.433	0.030	2.10 (0.94)
	5	DA Sem	8.7**	0.455	0.022	2.80 (0.94)
EAL group	1	SES	1.29	0.012	0.012	4.53 (2.74)
(n=123)	2	NVIQ	20.08**	0.152	0.140	6.84 (1.52)
	3	BPVS t1	100.14**	0.239	0.088	13.04 (1.05)
	4	DA Phon	16.73**	0.325	0.085	3.20 (1.62)
	5	DA Sem	10.67**	0.397	0.073	5.96 (1.76)

Note: * $p < .05$; ** $p < .01$; R^2 = marginal pseudo R-squared; ΔR^2 = change in pseudo R-squared; beta (SE) from final step; SES = IDACI rank

Table 5 Linear mixed effects step-up models predicting growth in reading comprehension for full sample and language subgroups

	Model	Step	Predictor	χ^2	R^2	ΔR^2
Full sample	Covariates	1	SES	2.11	0.011	0.011
(*n*=320)		2	EAL	8.46**	0.066	0.055
		3	NVIQ	58.48**	0.229	0.163
	Static	4	BPVS t1	204.45**	0.522	0.293
		5	YARC accuracy t1	24.11**	0.539	0.018
		6	YARC comp t1	84.02**	0.580	0.040
	Dynamic	4	DA Phon	32.32**	0.267	0.038
		5	DA Sem	15.91**	0.293	0.026
ML group	Covariates	1	SES	3.17	0.024	0.024
(*n*=197)		2	NVIQ	51.9**	0.218	0.194
	Static	3	BPVS t1	117.46**	0.478	0.260
		4	YARC accuracy t1	6.69**	0.487	0.009
		5	YARC comp t1	48.33**	0.530	0.043
	Dynamic	3	DA Phon	7.53**	0.235	0.017
		4	DA Sem	4.62*	0.243	0.008
EAL group	Covariates	1	SES	1.69	0.011	0.011
(*n*=123)		2	NVIQ	16.16**	0.112	0.100
	Static	3	BPVS t1	90.04**	0.482	0.370
		4	YARC accuracy t1	24.31**	0.543	0.061
		5	YARC comp t1	35.79**	0.588	0.045
	Dynamic	3	DA Phon	27.51**	0.255	0.143
		4	DA Sem	10.59**	0.302	0.047

Note: * $p < .05$; ** $p < .01$; R^2 = marginal pseudo R-squared; ΔR^2 = change in pseudo R-squared; beta (SE) from final step; SES = IDACI rank

tion metrics from Steps 6 and 8 in this model, the addition of DA factor scores did increase overall accuracy (AUC = 0.899), with the effect of reducing the FPR from 37.7% to 17.0%, but also failing to identify 3 true positive cases (sensitivity = 0.850).

Different results were seen for the language subgroups in the classification analysis. For ML learners (Table 7), a static model yielded significantly higher classification accuracy (AUC = 0.928) than a dynamic model (AUC = 0.834; $\chi^2(5) = 13.04$, $z = -2.17$, $p = 029$). Neither DA factor score was a significant predictor of PC risk status when included after covariates and static predictors. For EAL learners, a static model did not yield significantly higher classification accuracy (AUC = 0.854) than a dynamic model (AUC = 0.760; $\chi^2(5) = 10.74$, z = -1.15, $p = 249$; Table 7). When added after covariates and static predictors, the DA phonological factor was a significant predictor of PC reading profile risk status, serving to increase sensitivity from 71.4 to 100%, but to also increase FPR from 5.5 to 24.8% (Table 8).

Discussion

The present study investigated the predictive and prognostic value of a DA of word learning as a screening tool for identifying children at risk of developing reading comprehension difficulties, specifically those who show the PC reading profile.

Table 6 Linear mixed effects final models predicting growth in reading comprehension for full sample and language subgroups

	Step	Predictor	χ^2	R^2	ΔR^2	Beta (SE)
Full sample	1	SES	2.11	0.011	0.011	0.21 (0.38)
(*n*=320)	2	EAL	8.46**	0.066	0.055	-0.05 (0.8)
	3	NVIQ	58.48**	0.229	0.163	0.67 (0.33)
	4	BPVS t1	204.45**	0.522	0.293	5.00 (0.40)
	5	YARC accuracy t1	24.11**	0.539	0.018	0.17 (0.04)
	6	YARC comp t1	84.02**	0.580	0.040	3.76 (0.38)
	7	DA Phon	1.70	0.579	0.000	0.05 (0.34)
	8	DA Sem	6.41*	0.584	0.005	0.83 (0.32)
ML group	1	SES	3.17	0.024	0.024	0.39 (0.40)
(*n*=197)	2	NVIQ	51.9**	0.218	0.194	1.12 (0.42)
	3	BPVS t1	117.46**	0.478	0.260	5.81 (0.55)
	4	YARC accuracy t1	6.69**	0.487	0.009	0.13 (0.05)
	5	YARC comp t1	48.33**	0.530	0.043	3.50 (0.47)
	6	DA Phon	0.07	0.530	0.000	-0.21 (0.42)
	7	DA Sem	0.43	0.530	0.000	0.26 (0.39)
EAL group	1	SES	1.69	0.011	0.011	-0.26 (0.85)
(*n*=123)	2	NVIQ	16.16**	0.112	0.100	0.28 (0.51)
	3	BPVS t1	90.04**	0.482	0.370	4.35 (0.58)
	4	YARC accuracy t1	24.31**	0.543	0.061	0.25 (0.06)
	5	YARC comp t1	35.79**	0.588	0.045	4.19 (0.65)
	6	DA Phon	3.32	0.595	0.007	0.33 (0.57)
	7	DA Sem	6.07*	0.608	0.013	1.42 (0.57)

Note: * $p < .05$; ** $p < .01$; R^2 = marginal pseudo R-squared; ΔR^2 = change in pseudo R-squared; beta (SE) from final step; SES = IDACI rank

Adopting a longitudinal design, we measured the ability of a novel computer assisted DA of word learning to predict growth in vocabulary and reading comprehension, and to classify children at risk of the PC reading profile.

We posed three research questions and corresponding hypotheses. The first hypothesis predicted that a DA of word learning would correlate less strongly with SES or English language proficiency at the first time point compared to static assessments. This was supported, as scores on the DA task were weakly correlated with SES and moderately correlated with English language proficiency, whereas reading comprehension and vocabulary knowledge measured using static assessments were more strongly correlated with SES, and reading comprehension was more strongly correlated with English language proficiency.

The second hypothesis was that the DA of word learning would predict growth in vocabulary over time as found previously and that it would predict growth in reading comprehension (as suggested by the findings of Wolter & Pike, 2015). In addition to planned analyses for the whole sample, we conducted exploratory analyses for ML and EAL groups separately, to test whether any predictive relationship differed between them. Both the phonological and semantic DA factors were significant predictors of vocabulary growth when entered directly after the covariates. This held true for both language groups although the effect was slightly more pronounced in the EAL group. The semantic DA factor predicted significant unique variance in reading

Table 7 Logistic regression step-up models and classification accuracy for poor comprehender risk status (whole sample and language subgroups)

		Step	Predictor	χ^2	R^2	ΔR^2	AUC	Sen.	Spec.	TN	TP	FN	FP	FPR	FNR
Full sample	Covariates	1	SES	1.00	0.008	0.008	0.530	0.700	0.510	150	14	6	150	0.500	0.300
(*n*=320)		2	EAL	7.97**	0.074	0.066	0.677	0.700	0.660	208	13	7	92	0.307	0.350
TP=20		3	NVIQ	3.96*	0.106	0.032	0.742	0.700	0.720	216	14	6	84	0.280	0.300
TN=300	Static	4	BPVS t1	11.01**	0.193	0.087	0.836	0.750	0.837	251	15	5	49	0.163	0.250
		5	YARC acc ab t1	14.36**	0.302	0.109	0.862	1.000	0.607	180	20	0	120	0.400	0.000
		6	YARC comp ab t1	4.03*	0.332	0.030	0.878	1.000	0.623	187	20	0	113	0.377	0.000
	Dynamic	4	DA phon	2.46	0.126	0.020	0.756	0.700	0.730	219	14	6	81	0.270	0.300
		5	DA sem	9.07**	0.197	0.071	0.818	0.900	0.643	193	17	3	107	0.357	0.150
ML group	Covariates	1	SES	0.34	0.007	0.007	0.545	0.833	0.524	95	5	1	96	0.503	0.167
(*n*=197)		2	Matrices	4.24*	0.096	0.089	0.753	1.000	0.518	97	6	0	94	0.492	0.000
TP=6	Static	3	BPVS t1	4.53*	0.189	0.093	0.844	1.000	0.675	128	6	0	63	0.330	0.000
TN=191		4	YARC acc ab t1	7.21**	0.333	0.144	0.902	0.833	0.864	164	5	1	27	0.141	0.167
		5	YARC comp ab t1	6.50*	0.458	0.125	0.928	1.000	0.770	146	6	0	45	0.236	0.000
	Dynamic	3	DA phon	0.05	0.097	0.001	0.756	1.000	0.513	98	5	1	93	0.487	0.167
		4	DA sem	5.14*	0.203	0.105	0.834	0.833	0.806	154	5	1	37	0.194	0.167
EAL group	Covariates	1	SES	2.04	0.032	0.032	0.569	0.786	0.367	40	11	3	69	0.633	0.214
(*n*=123)		2	Matrices	1.37	0.054	0.021	0.614	0.857	0.413	45	12	2	64	0.587	0.143
TP=14	Static	3	BPVS t1	7.71**	0.170	0.116	0.783	1.000	0.587	64	14	0	45	0.413	0.000
TN=109		4	YARC acc ab t1	10.09**	0.312	0.142	0.841	0.714	0.881	96	10	4	13	0.119	0.286
		5	YARC comp ab t1	0.74	0.322	0.010	0.854	0.714	0.954	103	10	4	6	0.055	0.286
	Dynamic	3	DA phon	3.44	0.107	0.053	0.704	0.857	0.514	56	12	2	53	0.486	0.143
		4	DA sem	4.36*	0.172	0.065	0.760	0.857	0.679	74	12	2	35	0.321	0.143

Note: YARC acc ab t1 = YARC accuracy ability score; YARC comp ab = YARC comprehension ability score; 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

Table 8 Logistic regression final models and classification accuracy for poor comprehender risk status (whole sample and language subgroups)

	Step	Predictor	χ^2	R^2	ΔR^2	AUC	Sen.	Spec.	TN	TP	FN	FP	FPR	FNR
Full sample	1	SES	1.00	0.008	0.008	0.530	0.700	0.510	150	14	6	150	0.500	0.300
(*n*=320)	2	EAL	7.97**	0.074	0.066	0.677	0.700	0.660	208	13	7	92	0.307	0.350
TP=20	3	NVIQ	3.96*	0.106	0.032	0.742	0.700	0.720	216	14	6	84	0.280	0.300
TN=300	4	BPVS t1	11.01**	0.193	0.087	0.836	0.750	0.837	251	15	5	49	0.163	0.250
	5	YARC acc ab t1	14.36**	0.302	0.109	0.862	1.000	0.607	180	20	0	120	0.400	0.000
	6	YARC comp ab t1	4.03*	0.332	0.030	0.878	1.000	0.623	187	20	0	113	0.377	0.000
	7	DA phon	3.37	0.356	0.025	0.893	0.950	0.720	216	19	1	84	0.280	0.050
	8	DA sem	2.22	0.372	0.016	0.899	0.850	0.830	249	17	3	51	0.170	0.150
ML group	1	SES	0.34	0.007	0.007	0.545	0.833	0.524	95	5	1	96	0.503	0.167
(*n*=197)	2	NVIQ	4.24*	0.096	0.089	0.753	1.000	0.518	97	6	0	94	0.492	0.000
TP=6	3	BPVS t1	4.53*	0.189	0.093	0.844	1.000	0.675	128	6	0	63	0.330	0.000
TN=191	4	YARC acc ab t1	7.21**	0.333	0.144	0.902	0.833	0.864	164	5	1	27	0.141	0.167
	5	YARC comp ab t1	6.50*	0.458	0.125	0.928	1.000	0.770	146	6	0	45	0.236	0.000
	6	DA phon	0.01	0.458	0.000	0.928	1.000	0.770	146	6	0	45	0.236	0.000
	7	DA sem	1.18	0.481	0.022	0.928	1.000	0.712	133	6	0	58	0.304	0.000
EAL group	1	SES	2.04	0.032	0.032	0.569	0.786	0.367	40	11	3	69	0.633	0.214
(*n*=123)	2	NVIQ	1.37	0.054	0.021	0.614	0.857	0.413	45	12	2	64	0.587	0.143
TP=14	3	BPVS t1	7.71**	0.170	0.116	0.783	1.000	0.587	64	14	0	45	0.413	0.000
TN=109	4	YARC acc ab t1	10.09**	0.312	0.142	0.841	0.714	0.881	96	10	4	13	0.119	0.286
	5	YARC comp ab t1	0.74	0.322	0.010	0.854	0.714	0.954	103	10	4	6	0.055	0.286
	6	DA phon	4.86*	0.386	0.064	0.873	1.000	0.697	76	14	0	33	0.303	0.000
	7	DA sem	1.32	0.403	0.017	0.881	1.000	0.752	82	14	0	27	0.248	0.000

Note: YARC acc ab t1 = YARC accuracy ability score; YARC comp ab = YARC comprehension ability score; 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 comprehension, even after static predictors had been entered. This was the case for the whole sample and the EAL group, but not for the ML group. These data offer new evidence that extends that of Wolter and Pike (2015) by controlling for key variables such as reading accuracy. Our models show the separable contributions of decoding (reading accuracy) and linguistic comprehension (vocabulary knowledge) which aligns with the SVR (Gough & Tunmer, 1986). Furthermore, for the EAL group, we have shown that the capacity to learn new vocabulary makes a unique contribution, over and above existing vocabulary knowledge, to reading comprehension growth. This indicates that static measures of vocabulary knowledge only provide a partial account of the vocabulary skills important for reading comprehension in this population. The SVR does not explicitly account for new word learning ability, and it remains unclear whether this would be best captured within the linguistic comprehension component or as Nation (2019) suggests as an underlying language component that contributes to both decoding and linguistic comprehension.

Our third hypothesis was that the DA of word learning would differentiate good/poor comprehenders and more specifically that it would achieve acceptable levels of accuracy in identifying children at risk for of the PC profile at the second time point. We examined this for the whole sample and for the ML and EAL groups separately, as there is some suggestion that DA of decoding improves classification accuracy more in bilingual learners (Petersen et al., 2016, 2018). Although there are no absolute thresholds, sensitivity and specificity upwards of 80% has been suggested as acceptable, while an AUC of 0.8 to 0.9 represents excellent discrimination (Hosmer et al., 2013). We identified 20 children with the PC profile at the second time point, representing an overall prevalence rate of 6.25%; this was higher in the EAL group (14 children, 11.38%) compared to the ML group (6 children, 3.05%). This replicates the prevalence estimates obtained by Li et al. (2021). In terms of differentiating good/poor comprehenders in the whole sample, the DA task achieved an excellent level of classification accuracy comparable to that of the static assessments of vocabulary, text reading accuracy and earlier reading comprehension. This novel finding supports our third hypothesis.

In addition to the three main questions, we included two exploratory questions relating to screening accuracy. First, we examined how the DA of word learning compared to the static assessments in terms of screening accuracy. For the whole sample the two types of assessment were comparable, as the AUC values were not statistically different. This pattern holds for the EAL subgroup, however in the ML group there was a statistically significant difference, with the static assessment outperforming the DA. We then explored whether DA of word learning could improve classification accuracy when added to static assessments. For the whole sample the addition of the DA factor scores did increase accuracy (by reducing the number of false positives) but this was not statistically significant. However, for EAL learners the phonological DA factor did significantly improve classification accuracy, helping to identify all 14 EAL learners fitting the PC profile (whereas 4 children were missed when only the static assessments were included in the model). This is an important finding in the context of the increasing numbers of pupils with EAL in schools and the raised prevalence of the PC profile in this group (Li et al., 2021), which we confirmed in our sample.

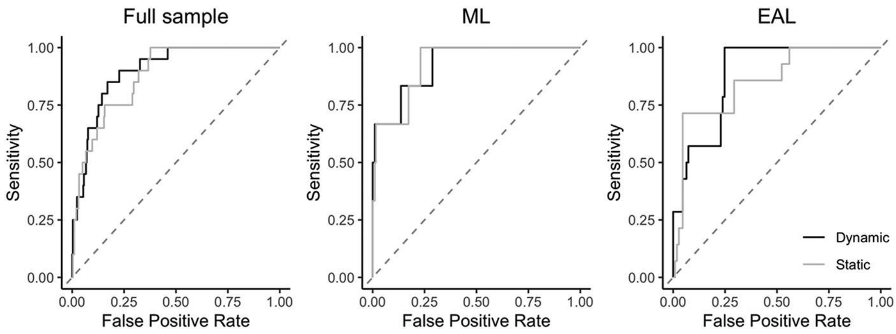

Fig. 2 ROC curves predicting risk for the Poor Comprehender reading profile in the full sample and subsamples

Whilst not a study aim, the separable DA factors provide estimates of the relative contributions of phonological and semantic aspects of vocabulary learning. However, the task was not originally designed for this purpose and there are differences between how these different aspects are taught (with phonological learning more active than semantic) and measured. Furthermore, in our models we only include a single static assessment of vocabulary which measures receptive understanding. Therefore, the different predictive relationships in the models are difficult to interpret. In future research it will be advantageous to scrutinize the training and recall phases of the task, potentially revising them to ensure that the DA is fully capturing both phonological and semantic aspects of vocabulary.

When interpreting the findings of this study it is necessary to acknowledge the potential influence of contextual factors, most notably, the COVID-19 pandemic which caused widespread disruption to schools in England during our intended second time point (April to July 2020). This did have the unforeseen advantage of allowing us to measure growth in vocabulary and reading skills over a longer period (16-19 months), representing a longer interval than in many previous longitudinal studies of DA (e.g., Bridges & Catts, 2011; Gellert & Elbro, 2013; Krenca et al., 2020). However, because of school disruption, participants missed an average of 13 weeks of regular classroom instruction, and it is not possible to assess the extent to which this may have influenced growth rates. Although we were able to reassess most children, we did lose one school at follow-up shortly before the onset of a second national lockdown in early 2021, reducing the proportion of ML children in the final sample.

To the best of our knowledge this is the first demonstration of the utility of a DA for identifying learners at risk of developing reading comprehension difficulties. Whilst the added benefit of the DA appears modest this is because of our conservative analytical approach and small sample. When comparing the DA to static assessments we have included a combination of demographic factors and a range of static assessments (nonverbal ability, vocabulary, word reading accuracy). This means that we are comparing its accuracy as a screener to a comprehensive battery of measures, which represents an idealised research context rather than the reality in schools. If we scale our findings to the wider population, using England as an example, then the added benefit of DA becomes clear. According to Government data, there were 146,898

EAL children in the upper years of primary school in England in 2022/23 who were eligible to sit the national Statutory Assessments Tests at the end of primary school (Explore Education Statistics, 2023). According to our estimated prevalence of 11%, 16, 158 of these will show the PC profile. Only 11, 472 (71%) of these children would be identified by static assessments alone whereas all the at-risk children would be identified if our DA was added to the screening battery. Although we have used England as an example, our findings can be applied to other countries where there are children whose home language is different to the language of instruction in school. These children's difficulties are at risk of going unnoticed and unsupported in the absence of suitable screening tools. It remains unclear precisely how the DA would work with static assessments in practice, a gated approach in which the DA is given to all children prior to further assessment of an at-risk subsample may be appropriate. With some adaptations we believe our DA could be suitable for children as young as six years old, which would enable practitioners to be able to identify children at risk of developing reading comprehension difficulties early in their development.

The novel computer assisted DA presented in this study offers benefits in sensitivity for detecting risk status for the PC profile, particularly for EAL learners. However, such benefits need to be considered against the potentially time-consuming nature of dynamic assessment (Dixon et al., 2023b). Our DA of word learning took approximately twenty-five minutes per child, although it is worth noting that static assessments can also be time consuming to administer and often require specialist training. In future development of our DA, we intend to revise the method to reduce the administration time as well as engaging in participatory co-design with children and educators to ensure that the DA is appealing and user-friendly, and to determine the support and training needed for educators to use the assessment independently. There is also the potential to extend the task down to younger pupils, to help identify poor comprehenders earlier and thus provide timely support.

In conclusion, our findings make three important contributions to the literature. First, a dynamic measure of word learning is less strongly related to SES and English language proficiency than traditional static assessments. Second, that performance on the DA of word learning uniquely predicts growth in vocabulary knowledge and reading comprehension. The latter finding was particularly apparent in EAL children, suggesting that both existing vocabulary knowledge and new word learning capacity need to be taken into account when considering the development of their reading comprehension. Third and finally, our DA of word learning achieved excellent classification accuracy when identifying poor comprehenders. Furthermore, the DA task helped to identify all of the poor comprehenders with EAL, when added to the static assessments. This underlines the value of the DA task for ensuring that the risk of reading comprehension difficulties does not go unidentified in EAL pupils. Given the potential utility and impact of our DA of word learning, future work with children and teachers is needed to optimise the task and develop it as a web-based app with accompanying training and support, which teaching staff can use easily in schools as part of their monitoring processes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11145-024-10603-x.

Acknowledgements 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.

Funding This research was supported by the Nuffield Foundation (Grant No. EDO/43441).

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

References

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of Second vs. Eighth Grade Reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332–345. https://doi.org/10.1177/0022219410369067

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

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

Burton, V. J., & Watkins, R. V. (2007). Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders*, *40*(5), 335–356. https://doi.org/10.1016/j.jcomdis.2006.06.015

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*, 683–696. https://doi.org/10.1348/000709905X67610

Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: the dynamic assessment of word learning (DAWL). *International Journal of Language & Communication Disorders*, *48*(5), 565–581.

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

Camilleri, B., & Law, J. (2014). Dynamic assessment of word learning skills of pre-school children with primary language impairment. *International journal of speech-language pathology*, *16*(5), 507–516.

Caravolas, M., Lervåg, A., Defior, S., Malkova, G., & Hulme, C. (2012). 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

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the Risk of Future Reading Difficulties in Kindergarten Children: A Research-Based Model and Its Clinical Implementation. *Language Speech and Hearing Services in Schools*, *32*(1), 38–50. https://doi.org/10.1044/0161-1461(2001/004)
Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*(9), 1407–1425. https://doi.org/10.1007/s11145-015-9576-x
Catts, H. W., Nielsen, D. C., Bridges, M. S., & Liu, Y. S. (2016). Early Identification of Reading Comprehension Difficulties. *Journal of Learning Disabilities*, *49*(5), 451–465. https://doi.org/10.1177/0022219414556121
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
Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties: A Randomized Controlled Trial. *Psychological Science*, *21*(8), 1106–1116. https://doi.org/10.1177/0956797610375449
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/
Department for Education (2023). *Statistics: Schools, pupils and their characteristics*. June 2023.
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*. https://doi.org/10.1177/00222194221117510
Du, Z., & Hao, Y. (2020). ReportROC: An Easy Way to Report ROC Analysis. https://CRAN.R-project.org/package=reportROC
Dunn, L., & Dunn, L.,NFER (2009). *British Picture Vocabulary Scale-Third Edition*. GL Assessment.
Elleman, A. M., Compton, D. L., Fuchs, D., Fuchs, L. S., & Bouton, B. (2011). Exploring dynamic assessment as a means of identifying children at risk of developing comprehension difficulties. *Journal of Learning Disabilities*, *44*(4), 348–357. https://doi.org/10.1177/0022219411407865
Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal Stability and Predictors of Poor Oral Comprehenders and Poor Decoders. *Journal of Experimental Child Psychology*, *115*(3), 497–516. https://doi.org/10.1016/j.jecp.2012.12.001
Explore Education Statistics Service Gov. UK (2023). Key Stage 2 Attainment, Academic year 2022/23. Office for National Statistics. https://explore-education-statistics.service.gov.uk/data-tables/fast-track/bc9ec011-afd6-433a-496f-08dbb0623a5d
Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods Instruments & Computers*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503
Gellert, A. S., & Elbro, C. (2013). Do experimental measures of word learning predict vocabulary development over time? A study of children from grade 3 to 4. *Learning and Individual Differences*, *26*, 1–8. https://doi.org/10.1016/j.lindif.2013.04.006
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
Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10.
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
Gruhn, S., Segers, E., Keuning, J., & Verhoeven, L. (2020). Profiling children's reading comprehension: A dynamic approach. *Learning and Individual Differences*, 82. https://doi.org/10.1016/j.lindif.2020.101923

Gupta, P., Lipinski, J., Abbs, B., Lin, P. H., Aktunc, E., Ludden, D., Martin, N., & Newman, R. (2004). Space aliens and nonwords: Stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods Instruments & Computers*, *36*(4), 599–603. https://doi.org/10.3758/BF03206540

Hasson, N., Camilleri, B., Jones, C., Smith, J., & Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child language teaching and therapy*, *29*(1), 57–75.

Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge University Press.

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, *16*(3), 199–207. https://doi.org/10.3109/17549507.2014.904441

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression* (3rd ed.). John Wiley.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychological Science*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. *Identification of learning disabilities: Research to practice* (pp. 99–149). Lawrence Erlbaum Associates.

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

Kaushanskaya, M., Gross, M., & Buac, M. (2014). Effects of Classroom Bilingualism on Task Shifting, Verbal Memory, and Word Learning in Children. *Developmental Science*, *17*(4), 564–583. https://doi.org/10.1111/desc.12142

Keenan, J. M., & Meenan, C. E. (2014). Test Differences in Diagnosing Reading Comprehension Deficits. *Journal of Learning Disabilities*, *47*(2), 125–135. https://doi.org/10.1177/0022219412439326

Kelso, K., Whitworth, A., Parsons, R., & Leitao, S. (2020). Hidden Reading Difficulties: Identifying Children Who Are Poor Comprehenders. *Learning Disability Quarterly*. https://doi.org/10.1177/0731948720961766

Kelley, K. (2022). MBESS: The MBESS r Package. https://CRAN.R-project.org/package=MBESS

Krenca, K., Gottardo, A., Geva, E., & Chen, X. (2020). English phonological specificity predicts early French reading difficulty in emerging bilingual children. *Annals of Dyslexia*, *70*(1), 27–42. https://doi.org/10.1007/s11881-019-00188-4

Landi, N., & Ryherd, K. (2017). Understanding specific reading comprehension deficit: A review. *Language and Linguistics Compass*, *11*(2), 1–24. https://doi.org/10.1111/lnc3.12234

Language and Reading Research Consortium (2015). Learning to Read: Should We Keep Things Simple? *Reading Research Quarterly*, 50 (2), 151– 169. https://doi.org/10.1002/rrq.99

Lee, S. H., & Tsai, S. F. (2017). Experimental intervention research on students with specific poor comprehension: A systematic review of treatment outcomes. *Reading And Writing*, *30*(4), 917–943. https://doi.org/10.1007/s11145-016-9697-x

Li, M., Geva, E., D'Angelo, N., Koh, P. W., Chen, X., & Gottardo, A. (2021). Exploring sources of poor reading comprehension in English language learners. *Annals of Dyslexia*, *71*(2), 299–321. https://doi.org/10.1007/s11881-021-00214-4

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

Matrat, M., Delage, H., & Kehoe, M. (2023). A new dynamic word learning task to diagnose language disorder in French-speaking monolingual and bilingual children. *Frontiers in Rehabilitation Sciences*, *3*, 1095023.

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, *24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272

OECD (2010). Closing the Gap for Immigrant Students: Policies, Practice and Performance. https://www.oecd-ilibrary.org/education/closing-the-gap-for-immigrant-students_9789264075788-en

Oxley, E. (2019). Word learning and vocabulary development in children with English as an additional language. [Unpublished doctoral thesis]. University of Leeds.

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
Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.
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., 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*, 875–888. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021
Proctor, C. P., Silverman, R. D., Harring, J. R., Jones, R. L., & Hartranft, A. M. (2020). Teaching Bilingual Learners: Effects of a Language-Based Reading Intervention on Academic Language and Reading Comprehension in Grades 4 and 5. *Reading Research Quarterly*, *55*(1), 95–122. https://doi.org/10.1002/rrq.258
Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology*, *116*(4), 914–929. https://doi.org/10.1016/j.jecp.2013.08.004
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.
Ricketts, J., Bishop, D. V. M., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading*, *31*(1), 117–135. https://doi.org/10.1111/j.1467-9817.2007.00365.x
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
Snowling, M. J., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language & Communication Disorders*, *47*(1), 27–34.
Snowling, M. J., Stothard, S., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K., & Hulme, C. (2011). *York Assessment of Reading for Comprehension*. GL Assessment.
Social Mobility Commission (2020). Monitoring social mobility 2013–2020: Is the government delivering on our recommendations? https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/891155/Monitoring_report_2013-2020_-Web_version.pdf
Spencer, M., & Wagner, R. K. (2018). The Comprehension Problems of Children With Poor Reading Comprehension Despite Adequate Decoding: A Meta-Analysis. *Review of Educational Research*, *88*(3), 366–400. https://doi.org/10.3102/0034654317749187
Stone, J. (2023). *Local indicators of child poverty after housing costs, 2021/22; Summary of estimates of child poverty after housing costs in local authorities and parliamentary constituencies, 2014/15 – 2021/22*. Loughborough University.
Strand, S., Malmberg, L., & Hall, J. (2015). *English as an additional language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. University of Oxford.
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
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
Verpalen, A., Van de Vijver, F., & Backus, A. (2018). Bias in dyslexia screening in a Dutch multicultural population. *Annals of Dyslexia*, *68*(1), 43–68.
Wolter, J. A., & Pike, K. (2015). Dynamic assessment of morphological awareness and third-grade literacy success. *Language Speech and Hearing Services in Schools*, *46*(2), 112–126. https://doi.org/10.1007/s11881-018-0155-0
Law, J., McBean, K., & Rush, R. (2015). (2011). Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. International *Journal of Language & Communication Disorders*, 46(6), 657–664. https://doi.org/10.1111/j.1460-6984.2011.00036.x

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

Authors and Affiliations

Hannah Nash[1] · Chris Dixon[1] · Paula Clarke[2] · Emily Oxley[1] · Anna Steenberg Gellert[3] · Anna Weighall[4]

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

4 School of Education, University of Sheffield, Sheffield, UK