



# Dynamic assessment as a predictor of reading development: a systematic review

Chris Dixon<sup>1</sup>  · Emily Oxley<sup>1</sup> · Anna Steenberg Gellert<sup>2</sup> · Hannah Nash<sup>1</sup>

Accepted: 4 May 2022 / Published online: 4 July 2022  
© The Author(s) 2022

## Abstract

Assessments of reading and reading-related skills which measure acquired knowledge may pose problems for the prediction of future reading performance. Such static measures often result in floor effects in the early stages of reading instruction, and may be particularly inaccurate predictors for children from culturally and linguistically diverse (CLD) backgrounds. Dynamic assessment (DA), in contrast, focuses on learning potential by measuring response to teaching, and may therefore be a less biased form of assessment. We conducted a systematic review of the literature to assess the ability of dynamic measures of reading and related skills to predict variance in the growth of children’s reading skills over time. Seventeen peer-reviewed articles met inclusion criteria, representing 18 studies published between 1992 and 2020. After static predictors were accounted for, dynamic measures of phonological awareness and decoding explained a significant amount of variance in the growth of word reading accuracy (1–21%) and word reading fluency (typically 1–9%), while variance in reading comprehension outcomes was accounted for by dynamic measures of morphological awareness (4–33.4%) and one dynamic decoding assessment (1%). Finally, a single paired-associate nonword learning task predicted 6% unique variance in future nonword reading accuracy and fluency. Results support the ability of DA to tap into variance unexplained by traditional static measures, though no studies explicitly examined the validity of DA for children from CLD backgrounds. We call for future studies of DA of reading to adopt longer developmental windows and assess proximal as well as distal reading outcome measures.

**Keywords** Dynamic assessment · Systematic review · Prediction · Decoding · Comprehension

---

✉ Chris Dixon  
cddixon90@gmail.com

Extended author information available on the last page of the article

## Introduction

Learning to read is a foundational outcome of formal education. Good reading skills allow school pupils to access curriculum content and progress on to further education and training; as a result, early difficulties in reading skills have deleterious impacts on future life outcomes including educational attainment and earnings (McLaughlin et al., 2014). Depending on diagnostic criteria, small but significant proportions of school-aged children experience specific difficulties in accurate and/or fluent word recognition (5–17%; Grigorenko et al., 2020) or the ability to extract meaning and make inferences in written texts (5–11%; Kelso et al., 2020). Given the stability of poor reading skills over time, reading assessments administered in the earlier stages of education could feasibly be used to predict growth in reading and to identify children at risk of poor reading development (Catts et al., 2006). Indeed, early intervention for reading difficulties is highly desirable as it is most likely to be effective when provided earlier rather than later (e.g., before third grade; Lovett et al., 2017); and may be crucial in efforts to break a negative feedback loop in which poor reading skills lead to lower motivation, engagement, and less print exposure over time (van Bergen et al., 2018).

Standardised assessments of phonological awareness (PA), rapid automatized naming (RAN), and letter-sound knowledge are robust predictors of later word-level reading skills across a range of alphabetic orthographies including English, Spanish, Czech, Slovak, and Finnish (Caravolas et al., 2013; Puolakanaho et al., 2007). In contrast, measures of vocabulary, morphosyntax, and listening comprehension are significant predictors of future reading comprehension performance (Muter et al., 2004; Verhoeven & Van Leeuwe, 2008). By measuring pre-existing knowledge on the day of the test, such assessments are considered to be ‘static’ in nature, whereby the role of the examiner is neutral and detached, and corrective feedback on examinees’ responses is strictly prohibited. In certain circumstances, the use of static assessments to predict growth in reading skill over time is problematic. Firstly, when administered prior to or around the onset of formal reading instruction, static measures of letter knowledge and decoding often result in floor effects (Catts et al., 2009), and therefore provide insufficient variance for predictive statistical modelling. Secondly, by measuring pre-existing knowledge, static assessments are insensitive to variation in children’s learning experiences and opportunities, which is a particular issue for children from culturally and linguistically diverse (CLD) backgrounds who may not have access to the same learning resources or linguistic input (Peña & Halle, 2011). In other words, if the static measurement of pre-existing knowledge is not an accurate indication of current reading skills for such children, then it is also unlikely to provide accurate prediction of their reading skills in the future. One alternative to static testing which does not depend as heavily on prior learning opportunity is dynamic assessment.

## Dynamic assessment

Dynamic assessment (DA) is an approach to psychological testing that conceptualises cognitive ability as ‘developing’ as opposed to ‘developed’. Crucially, it makes a distinction between what an individual is capable of (‘latent capacity’)

and what an individual has achieved as a result of this capability plus environmental factors such as education and parental support ('developed ability'; Sternberg & Grigorenko, 2002). Theoretically, DA is closely linked to response to intervention (RTI) frameworks (Grigorenko, 2009), though typically takes place within a single or small number of sessions as opposed to progress monitoring over the course of many weeks. The goal of DA is to shift the focus from the product of learning to the process of learning by measuring individuals' response to teaching and feedback. Learning potential may be quantified as gains within a test-teach-retest procedure, or the amount of assistance (e.g., prompting) required to achieve learning goals (see Method for more information on these formats). In both cases, performance is interpreted in an 'ideographic' or within-individual fashion (Haywood & Lidz, 2007), contrasting sharply with the interpretation of performance on static assessments in which an individual's test score is compared to that of a norming population.

DA is rooted in the work of Lev Vygotsky and in particular the concept of a *zone of proximal development* (Dumas et al., 2020). Today there exist multiple distinct DA frameworks including structural cognitive modifiability, learning potential testing, testing-the-limits, *Lerntest*, and graduated prompts (see Sternberg & Grigorenko, 2002 for a review). According to Haywood's (1997) nomenclature, DA involves either (i) restructuring the test situation, (ii) learning within the test, or (iii) metacognitive intervention. Sternberg and Grigorenko (2002) propose a fourth category, namely 'training a single cognitive function', for example DA of constructs such as working memory, phonological awareness, and narrative retelling. The present review aligns with this latter category, focusing on DA of reading and reading-related skills.

Criticism has been levied at DA for its 'concept fuzziness', questionable psychometric properties, and time-consuming nature (Grigorenko & Sternberg, 1998), and its take-up among educational psychologists is low (Hill, 2015). However, DA does offer promise in situations where static tests result in floor effects or underestimate the abilities of children from CLD backgrounds (Haywood & Lidz, 2007). Additionally, there is growing evidence for the ability of DA to enhance accuracy and reduce bias in the prediction of future performance across a range of skills. In a review of 24 studies examining a broad range of achievement measures, Caffrey et al. (2008) found that scores derived from dynamic and static assessments correlated similarly with future achievement, but when entered into multiple regression equations, measures of learning potential accounted for additional variance (over and above that explained by static assessment) in non-verbal reasoning, mathematics, phonemic awareness, reading, and writing outcomes. The value-added nature of DA is also supported by a number of more recent studies. For instance, dynamic measures of word learning have been shown to predict unique variance in vocabulary growth among both mono- and bi-/multilingual children in Denmark and the UK (Gellert & Elbro, 2013; Oxley, 2019), and a dynamically administered oral narrative retell instrument has been shown to improve classification accuracy for language disorder risk status among Spanish-English bilingual children (Petersen et al., 2020).

## The present study

In recent years there has been renewed interest in the use of DA of reading skills: converging in their research questions, instrument design, and analytical procedures, existing studies make important contributions to the evidence base for DA and may assist researchers and practitioners alike in their choice of assessment procedure. Although previous reviews of DA exist (e.g., Caffrey et al., 2008), the present review adopts a narrower focus by only including studies which examine the contribution of DA to reading skills over and above that of traditional static tests. One previous systematic review (Dixon et al., in press), did examine the ability of dynamic measures of reading and related skills to enhance classification accuracy of screening for reading disorder (i.e., to assign at-risk versus not-at-risk status). In a complementary fashion, the present review instead focuses on the ability of dynamic measures to explain variance in the *growth* of children's reading skills over time or performance at a future point in time, over and above the explanatory power of static predictors. The prediction of growth, as opposed to binary classification, represents a distinct but important aspect of screening and speaks to the utility of DA as a tool for educators and speech and language therapists in considering children's likely developmental trajectories. This is particularly true of children from CLD backgrounds whose progress may be difficult to predict, and we therefore explore the value of DA over and above that of static tests to predict reading skills for children from CLD backgrounds.

## Method

### Literature search

We developed search terms through an iterative process using different combinations of key words from relevant articles already known to the authors. We searched the electronic databases ERIC, LLBA, Medline, PsycInfo, and Web of Science (20/01/2022) using the following search terms with no restrictions on language, publication year, or search fields: (child\* OR under-18 OR pupil\* OR student\*) AND ("dynamic assessment\*" OR "dynamic test\*" OR "dynamic task\*" OR "mediated learning" OR "mediated assessment\*" OR "interactive assessment\*" OR testing-the-limits OR "learning potential" OR "learning task") AND (read\* OR "phonological awareness" OR decod\* OR "word recognition" OR accur\* OR fluen\* OR "reading comp\*"). Additional hand searches were carried out on reference lists of articles at the full-text screening stage ( $n=44$ ) and one study (Gruhn et al., 2020) was identified as potentially eligible for the review by the third author.

## Inclusion criteria

We stipulated the following criteria for inclusion in the review: (a) study uses a DA of reading or a reading-related skill to predict (i) variance in the growth of reading skill between two or more points in time or (ii) at one future time point only; (b) study examines variance accounted for in reading outcomes by DA over and above static measures; (c) study reports empirical data and appropriate statistical information for determining relationships between DA and reading including path coefficients and change in  $R^2$  (we excluded studies focusing only on the statistical reliability of DA or using DA to classify participants into groups, e.g., with logistic regression); (d) participants aged up to 18 years; (e) peer reviewed and published in English (although there were no restrictions on the language in which DA was administered). Note that DA was operationalised as testing procedures within which explicit teaching and/or feedback was provided, and which participants were given the opportunity to act upon (e.g., repeated attempts at the same stimuli/question or application of learning in a teaching phase to novel items or without assistance).

## Screening and Coding

We employed double screening of all article titles and abstracts returned from database searches, with the first author screening all records ( $n=5,600$ ) and the second author and a research assistant screening half of the records respectively ( $n=2,800$  each). Percentage agreement and interrater reliability (Cohen's kappa) were computed using the irr package (Gamer et al., 2019) in R (R Core Team, 2021). We achieved agreement of 99.7% to 99.8% between both pairs of raters, resulting in an overall agreement rate of 99.8% at the title/abstract screening stage. Full-text screening was carried out by the first two authors on a total of 73 articles arising from database searches, backward citation searches and one record identified by the third author, achieving 95.9% agreement and a Cohen's kappa statistic of 0.892,  $p < 0.01$  (representing 'near perfect' agreement; Landis & Koch, 1977). All disagreements were resolved through discussion.

Data were extracted by the first two authors from the 17 studies meeting inclusion criteria using an adapted version of the Cochrane Collaboration data extraction form for non-RCT studies, including information on country, sample size, gender, age, and second language status. DA procedures were coded for the construct in which examinees received training or instruction (e.g., decoding), and whether the task was administered by computer. The format of each DA was coded as 'pretest-teach-posttest' (PTP) if it employed a teaching and posttest phase and provided feedback during the teaching phase. DA format was coded as 'graduated prompts' (GP) if it employed a graduated set of hints for each incorrect response and incorporated the number of prompts required into the operationalisation of learning potential.

## Quality assessment

We assessed the quality of each study included in the review using the Quality Assessment Tool for Studies of Diverse Designs (QATSDD) instrument (Sirriyeh et al., 2012). The QATSDD provides 16 study quality indicators such as ‘explicit theoretical framework’ and ‘description of procedure for data collection’ which are scored from 0 (no mention at all) to 3 (described in full). Note that quality scores in the present study are based only on the 14 indicators in the QATSDD relevant for quantitative research designs (a maximum possible score out of 42). All studies were independently rated by the first two authors, achieving a Kappa statistic of 0.703,  $p < 0.01$ , representing ‘substantial’ agreement (Landis & Koch, 1977). Using the method of Murphy and Unthiah (2015), in the case of discrepancies of 1 point we selected the lower of the two scores, while discrepancies of 2 or more points were discussed and a final score was agreed upon through discussion. Mean study quality, expressed as a percentage out of 42, was judged to be 64% (range 42.9–76.2%).

## Results

From an initial pool of 6,784 records resulting from all literature searches, 17 articles met criteria for inclusion in the review, representing 18 individual studies published between 1992 and 2020 (see Fig. 1 for PRISMA flowchart and Table 1 for study characteristics). Participants were followed up at various points between pre-school and sixth grade. Studies reported a median sample size of  $n = 120$  (range: 38–1,988) and were conducted in the USA ( $n = 7$ ), Netherlands ( $n = 2$ ), Chile ( $n = 2$ ), Germany ( $n = 2$ ), Denmark ( $n = 2$ ), Canada ( $n = 1$ ), and Iran ( $n = 1$ ). One study recruited participants across four different countries (Australia, USA, Norway, and

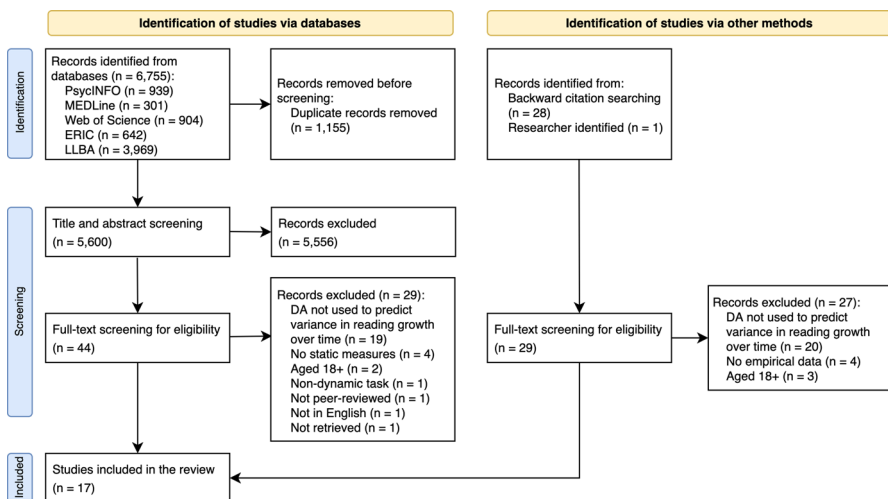


Fig. 1 PRISMA flowchart of literature search

**Table 1** Study characteristics (ordered by construct targeted)

Reference	N (Age/Grade range)	Country	% Male	% Bilingual	Construct	Format	Computerised	Dynamic test instrument/ procedure
Bridges and Catts (2011) Study 1	90 (Sep K to Apr K)	USA	NR	NR	PA	GP	Yes	The Dynamic Screening of Phonological Awareness (DSPA): examinee is asked to produce words without particular syllables and phonemes
Bridges and Catts (2011) Study 2	96 (Sep K to Apr K)	USA	NR	NR	PA	GP	No	DSPA as in Bridges and Catts (2011) Study 1
Coventry et al. (2011)	1,988 (Preschool to K)	Australia, USA, Norway, Sweden	51*	NR	PA	PTP	No	Phoneme identification task: examinee is taught 4 phonemes using up to 4 increasingly explicit levels of instruction and then asked to indicate which word begins or ends with a given phoneme
Gellert and Elbro (2017a, 2017b)	160 (Nov K, end K, Nov G1, end G1)	Denmark	48	37	PA	GP	No	Phoneme identification task: examinee is asked to point to picture beginning with a given phoneme
Spector (1992)	38 (Nov K to May K)	USA	NR	0	PA	GP	No	Dynamic phoneme segmentation task: examinee attempts to segment the phonemes of 12 words with CV, VC, and CVC structures

Table 1 (continued)

Reference	N (Age/Grade range)	Country	% Male	% Bilingual	Construct	Format	Computerised	Dynamic test instrument/ procedure
Krenca et al. (2020)	62 (Fall G1 to Spring G1)	Canada	44	31	PA	PTP	Yes	Computerised lexical specificity training (Ziggy's Word Game): examinee is presented with 2 unfamiliar minimal-pair targets, e.g. <i>foal</i> and <i>sole</i> , 1 unfamiliar control e.g. <i>knoll</i> , and 1 familiar control e.g. <i>bowl</i> and asked to "show me the [target]"
Aravena et al. (2016)	55 (7;4–11;1 to 10 months later)	Netherlands	55	0	Decoding	PTP	Yes	Training: examinee is taught to match speech sounds with unfamiliar Hebrew graphemes in a computer game. Across 2 posttests, examinees match spoken speech sounds with corresponding graphemes and read as many high-frequency Dutch words written in the unfamiliar orthography as possible within 3 min



Table 1 (continued)

Reference	N (Age/Grade range)	Country	% Male	% Bilingual	Construct	Format	Computerised	Dynamic test instrument/ procedure
Cho et al. (2014)	134 (Nov G1 to 14 weeks later)	USA	51	1	Decoding	GP	No	Training: examinee is taught to read nonwords using three decoding skills: CVC ( <i>vop</i> ), CVCe ( <i>vope</i> ), and CVC(C)ing ( <i>vopping</i> ). Mastery must be met on a set of untaught nonwords (5/6 correct) before attempting the next stage. Failure to reach mastery results in reteaching the skill using a more explicit (graduated) level of instruction
Cho et al. (2017)	105 (Fall G1 to May G1)	USA	55	0	Decoding	GP	No	Training: paired-associate sound-symbol learning of 6 unfamiliar Mandarin characters (9 trials). Posttest 1 blending: examinee is asked to blend symbol-sound pairs into CVC (real) words (4 trials). Posttest 2 rule-based learning: examinee is prompted to infer a silent-e rule and decode CVC(e) words (5 trials). Each phase consisted of multiple learning trials with a mastery test
Cho et al. (2020)	104 (Fall G1 to May G1)	USA	55	0	Decoding	GP	No	As in Cho et al. (2017)

Table 1 (continued)

Reference	N (Age/Grade range)	Country	% Male	% Bilingual	Construct	Format	Computerised	Dynamic test instrument/ procedure
Fuchs et al. (2011)	318 (Fall G1 to Spring G2)	USA	49	NR	Decoding GP	GP	No	DA of nonword decoding as in Cho et al. (2014)
Horbach et al. (2015)	243 (K to G1)	Germany	56	0	Decoding PTP	PTP	Yes	Symbol-sound paradigm (SSP): examinee is taught to associate strings of dots and dashes with CV syllables /ta/ and /ma/. A final test phase assesses the ability to name syllable strings consisting of combinations of three or four symbols
Horbach et al. (2018)	56 (Preschool to G3)	Germany	39	52	Decoding PTP	PTP	Yes	SSP as in Horbach et al. (2018)
Navarro and Mourgués-Codern (2018)	813 (G3–G6 to 11 months later)	Chile	62	NR	MA	GP	Yes	Dynamic syntactic awareness task (DSA) of the EDPL-BAI battery: examinee is presented with 44 plausible and implausible sentences ranging in syntactic structure and answer multiple-choice questions about their meaning

Table 1 (continued)

Reference	N (Age/Grade range)	Country	% Male	% Bilingual	Construct	Format	Computerised	Dynamic test instrument/ procedure
Navarro et al. (2018)	324 (G3-G6 to 5 months later)	Chile	54	NR	MA	GP	Yes	Subtests of the EDPL-BAI as in Navarro and Mourgués-Codern (2018): morpho-syntactic awareness ('MS'; cloze task), rearranging disordered sentences ('OF'), and syntactic awareness ('CS'; multiple-choice questions about short sentences)
Hamavandi et al. (2017)	50 (age 14-18; 2 months later)	Iran	0	NR	MA	GP	No	Adaptation of DATMA (Larsen & Nippold, 2007) for intermediate EFL learners: examinee is asked to provide definitions for 10 low-frequency morphologically complex words with high frequency bases shown on cards

Table 1 (continued)

Reference	N (Age/Grade range)	Country	% Male	% Bilingual	Construct	Format	Computerised	Dynamic test instrument/ procedure
Gruhn et al. (2020)	275 (G3, G4, G5 to 6 months later)	Netherlands	NR	NR	RC	GP	Yes	Self-paced computerised test of reading comprehension: examinee is presented with 25 texts containing 7 questions assessing orthography, vocabulary, sentence-integration, and global comprehension. One hint is provided for each incorrect answer. Response accuracy is measured on first and second attempts (0–2)
Poulsen and Elbro (2018)	137 (G0, G1, G5)	Denmark	53	4	WL	PAL	No	Visual-verbal paired-associate word learning task. Three non-words corresponding to cartoon animals are repeated in short story contexts. In a subsequent learning phase, examinees are asked to recall the names across a maximum of 42 trials with feedback. Score is the percent of names recalled correctly

PA = phonological awareness; MA = morphological awareness; RC = reading comprehension; WL = word learning; PAL = paired-associate learning; NR = not reported; GP = graduated prompts; PTP = pretest-teach-posttest; PAL = paired-associate learning; \* averaged across mono- and dizygotic samples

Sweden; Coventry et al., 2011). Studies employed dynamic assessments of the following constructs: decoding ( $n=7$ ), phonological awareness ( $n=6$ ), morphological awareness ( $n=3$ ), nonword learning ( $n=1$ ), and reading comprehension ( $n=1$ ). The most commonly adopted assessment procedure was a graduated prompts format ( $n=12$ ), with the remainder of studies employing a pretest-teach-posttest format ( $n=5$ ). One study employed a paired-associate learning (PAL) paradigm (Poulsen & Elbro, 2018) in which participants were trained to learn the nonword labels of cartoon animals across a series of learning trials with corrective feedback. Assessments were computerised in only a minority of cases ( $n=7$ ). Results of the systematic review are discussed below for each of the five constructs in turn.

### Phonological awareness (PA)

Six studies examined the contribution of dynamic measures of PA to reading and reading-related skills, following participants between preschool and first grade (Bridges & Catts, 2011; Coventry et al., 2011; Gellert & Elbro, 2017a; Krenca et al., 2020; Spector, 1992). Studies recruited children on an unselected basis, with the exceptions of Bridges and Catts (2011) who purposely recruited approximately 50% of children in their second sample to be at risk of reading difficulties according to DIBELS scores, and Gellert and Elbro (2017a) who oversampled children deemed to be at risk of reading difficulties according to PA and letter knowledge. Studies generally used a combination of static and dynamic measures to predict variance in reading accuracy and/or fluency. Dynamic assessments of PA included phoneme deletion (Bridges & Catts, 2011), segmentation (Spector, 1992), identification (Coventry et al., 2011; Gellert & Elbro, 2017a), and lexical specificity training (onsets, codas, vowels; Krenca et al., 2020). A graduated prompts procedure was employed in all of the studies.

A number of studies in the sample reported floor effects for static assessments of PA and reading measures, while scores derived from dynamic tests showed either very little skew or negative skew, indicating generally high performance (Bridges & Catts, 2011; Gellert & Elbro, 2017a; Spector, 1992). After controlling for static predictors including letter knowledge, PA, and in one case a word reading autoregressor, dynamic PA scores accounted for 4–21% of unique variance in word reading accuracy outcome measures (Bridges & Catts, 2011; Gellert & Elbro, 2017a; Spector, 1992). However, the contribution of DA was no longer statistically significant when used to predict reading scores 17 months into the future (Gellert & Elbro, 2017a) or when a sound fluency measure was included as a static predictor (Bridges & Catts, 2011 Study 2). Interesting findings emerged from Gellert and Elbro (2017a) in which children were assessed at four time points between kindergarten and end of first grade. Static measures of PA and letter knowledge and a dynamic PA score measured in kindergarten were entered in hierarchical regression models as predictors of performance on a word reading accuracy composite at different time points in the study. Dynamic PA score predicted significant and unique variance in word reading at the end of kindergarten (7%) and November of first grade (3%), but failed to reach significance at the end of first grade. Indeed, a similar pattern was found in

logistic models predicting reading disorder risk status at each time point, and thus results suggest that the predictive power of DA may be developmentally constrained (see also Dixon et al., in press, on this point).

Dynamic PA scores were also found to predict 1–9% of variance in future non-word reading fluency measures over and above static predictors (Bridges & Catts, 2011; Coventry et al., 2011). Mediation analysis was used by Krenca et al. (2020) in their sample of emergent French–English bilingual children in Canada to assess direct and indirect contributions of dynamic phonological training score to current and later performance on a composite variable of word reading accuracy and fluency. Results provided no evidence for the direct contribution of the dynamic training score on future word reading skill in English, though there was a significant indirect effect through concurrent PA performance (elision). In other words, children who were able to learn high-quality phonological representations during the dynamic task also obtained high scores on a static PA measure, which itself was a predictor of later word reading skill.

## Decoding

Seven studies used dynamic decoding assessments to predict variance in the growth of reading skills, following participants between kindergarten and first grade (Horbach et al., 2015), the beginning to end of first grade (Cho et al., 2017, 2020), over a 14-week period in first grade (Cho et al., 2014), fall to spring of first grade (Fuchs et al., 2011), and in one case from preschool to third grade (Horbach et al., 2018). Additionally, one study used DA to predict reading gains ten months after an intervention amongst a sample of 7- to 11-year-old children with a dyslexia diagnosis (Aravena et al., 2016), and participants in Cho et al. (2014) were those who were shown to be unresponsive to Tier-1 classroom instruction. Similar to PA studies discussed above, dynamic decoding studies mostly used combinations of static and dynamic variables to predict variance in reading fluency, accuracy, or composites thereof, though two studies also examined reading comprehension as an outcome (Fuchs et al., 2011; Horbach et al., 2018).

Participants were trained to decode nonwords either in a novel orthography (in Mandarin: Cho et al., 2017, 2020; in Hebrew: Aravena et al., 2016; in dots and dashes: Horbach et al., 2015, 2018) or in the same orthography of instruction (English; Cho et al., 2014; Fuchs et al., 2011). Two studies used the same dynamic test of decoding in which children are taught to apply different strategies for reading nonwords with CVC (*vop*), CVCe (*vope*) and CVC(C)ing (*vopping*) structures (Cho et al., 2014; Fuchs et al., 2011), and a further two studies adapted this procedure by substituting English letters for Mandarin characters (Cho et al., 2017, 2020).

Skewness statistics are not reported by the majority of dynamic decoding studies in the sample, though particularly low performance is reported for a static measure of PA in Cho et al. (2017) and for letter knowledge in Horbach et al. (2018). After accounting for static predictors such as PA, RAN, vocabulary, and autoregressors in regression models, dynamic measures accounted for between 1 and 17% of unique variance in word reading accuracy outcome measures (Aravena et al., 2016; Cho

et al., 2014, 2017, 2020; Fuchs et al., 2011; Horbach et al., 2015). In terms of reading fluency outcomes, dynamic scores accounted for between 4 and 8% of unique variance when entered after static predictors of RAN and PA, though coefficients did not reach the threshold of statistical significance in two studies (Cho et al., 2017; Fuchs et al., 2011), and in one study this effect was significant only among a subgroup of 'preliterate' children (Horbach et al., 2015). Horbach et al. (2018) assessed the predictive validity of their dynamic sound-symbol learning paradigm (SSP) score against static measures of IQ, letter knowledge, and age in predicting word and nonword reading fluency measured three years later: in both models, SSP was the only significant predictor, with models accounting for 63%–72% of variance in total.

Finally, two studies assessed the contribution of DA to reading comprehension. Using the same regression model structure described above, Horbach et al. (2018) found their SSP score to be the only significant predictor in a model predicting future reading comprehension, accounting for 82% variance in total. The relatively large coefficients of determination in models reported by Horbach et al. (2018) may be due to the limited number of covariates included in regression models (though all models did contain a static measure of reading in the form of a letter naming task). Rather more robust results are offered by Fuchs et al. (2011) in which dynamic decoding score contributed a small but significant 1% of unique variance to reading comprehension even after controlling for letter naming, RAN, PA, vocabulary, listening comprehension, reading fluency, IQ, and measures of attention.

### **Morphological awareness (MA)**

Three studies evaluated dynamic measures of MA. In Navarro and Mourgues-Codern (2018) and Navarro et al. (2018), large samples of children were initially recruited in third to sixth grade in Chile, and followed up at 11 and 5 months, respectively. In Hamavandi et al. (2017), 14 to 18 year-old participants in Iran were followed up approximately two and a half months after pretest. Therefore, participants in dynamic MA studies were substantially older than in studies examining PA or decoding. In two studies, DA of MA was measured through judgement of (im) plausible sentences and cloze exercises, assessing learning potential through graduated prompts (the dynamic EDPL-BAI battery; Navarro & Mourgues-Codern, 2018; Navarro et al., 2018). Hamavandi et al. (2017) measured DA of MA with an adapted version of the Dynamic Assessment Task of Morphological Awareness (DATMA; Larsen & Nippold, 2007). In the DATMA, participants are asked to provide verbal definitions for low-frequency morphologically derived words (e.g., where *beast* is a root form and *beastly* is a derived form) and to justify their answers with morphological knowledge defined as “awareness of the constituent morphemes of a word, knowledge of their meanings, and the ability to integrate that information” (Larsen & Nippold, 2007, p.204). A series of graduated prompts is employed in the event of an incorrect answer, drawing attention to meaning and constituent parts. The task was adapted specifically for adolescent intermediate English as a foreign language students by relaxing scoring criteria for definitions (i.e., not requiring formal literate language found in dictionary definitions).

All three studies examined reading comprehension as an outcome. Dynamic MA scores accounted for 33.4% variance (entered after static MA) in Hamavandi et al. (2017). This apparently large proportion of explained variance is tempered by the relatively lower coefficients of determination reported by Navarro and Mourgues-Codern (2018): controlling for an autoregressor, dynamic MA scores here accounted for only 4–5% variance in reading comprehension, though this was no longer statistically significant after a measure of nonverbal reasoning was added to the regression model. In a similar analysis, Navarro et al. (2018) used structural equation modelling to assess the contribution of DA on later reading comprehension. The authors found that a latent variable derived from all three subtests of the EDPL-BAI battery was a significant and unique predictor of later reading comprehension even after controlling for an autoregressor and nonverbal reasoning.

### Paired-associate learning

One study used a paired-associate learning (PAL) paradigm to predict growth in reading skills. Poulsen and Elbro (2018) administered their task to a sample of 137 children in Denmark in Grade 0 (age 6;10 before the onset of formal reading instruction), with follow-up assessments in Grade 1 and Grade 5 to measure growth in word-reading. This task uses a visual-verbal PAL paradigm in which examinees are taught nonword labels of three cartoon animals (*sput*, *laf*, *ky*). Stimuli are initially introduced across repeated short story contexts before the onset of a learning phase consisting of a maximum of 42 trials. Corrective feedback is provided throughout the learning phase and the final score is the percentage of trials in which each nonword animal label is correctly named. At the first time point, participants were also administered measures of RAN (digits and objects), letter knowledge, and PA, with measures of decoding accuracy and fluency administered in Grades 1 and 5. In a hierarchical regression model accounting for static measures in Grade 0, PAL score was a significant and unique predictor of real word decoding in Grade 1 (2% variance). Additionally, in another set of models predicting Grade 5 reading outcomes, PAL score was a significant and unique predictor of nonword reading accuracy and fluency (accounting for 6% unique variance in both cases) after controlling for reading precursor measures in Grade 0 and decoding in Grade 1.

### Reading comprehension (RC)

One study evaluated the predictive validity of a DA of reading comprehension, following a sample of children in third, fourth, and fifth grade over a 9-month period (Gruhn et al., 2020). Children in this study participated in a self-paced computerised DA of reading comprehension. Across a total of 25 texts, examinees answered multiple-choice and short-answer questions tapping knowledge of orthography, vocabulary, and sentence-integration before attempting a single global inference question. A single prompt is provided for each incorrect answer (apart from inference questions); for instance, pictures are presented on vocabulary definition items, and relevant parts of sentences are highlighted on sentence-integration items. As



a result, participants have only one opportunity to utilise feedback on incorrectly answered questions. Note that although this procedure differs from that of other studies included in the review, participants are nevertheless given opportunity to act upon feedback. Scores on the assessment therefore consist of subtotals grouped by orthographic, vocabulary, and integration questions for correct responses after a first attempt and after a second attempt, with this second score indexing potential to learn from feedback. In a linear mixed effects model accounting for an autoregressor, it was only first- and not second-attempt subscores that were significantly predictive of reading comprehension growth.

### **The value-added nature of DA in CLD populations**

As indicated in Table 1, ten studies reported statistics on the language status of participants, with six reporting fully or approximately fully monolingual samples (Ara-vena et al., 2016; Cho et al., 2014, 2017, 2020; Horbach et al., 2015; Poulsen & Elbro, 2018; Spector, 1992). Other studies recruited both mono- and bi-/multilingual children, ranging from 28 to 52% (Gellert & Elbro, 2017a; Horbach et al., 2018; Krenca et al., 2020). No study in our sample assessed the differential predictive validity of DA for bi-/multilingual children as an explicit research question.

Studies ranged similarly in their reporting of socio-economic status. Of the five studies reporting such data, the proportion of children eligible for free lunch ranged from 15 to 66%. Horbach et al. (2015) report lower levels of parental education among a ‘preliterate’ group of children, and participants in Horbach et al. (2018) were recruited from “regions of relatively low socioeconomic status” (p.4). Again, no studies explicitly compared the predictive validity of DA according to different metrics of SES.

## **Discussion**

In contrast to traditional static tests, dynamic assessments focus on individuals’ learning potential by measuring their ability to benefit from feedback. Performance on static measures may be strongly influenced by variation in learning opportunities, parental support, home language, and socioeconomic status (Sternberg & Grigorenko, 2002); therefore by measuring the processes of learning rather than its products, DA is said to offer a more sensitive measure of latent capacity particularly for children from CLD backgrounds (Tzuriel, 2000). We conducted a systematic review to synthesise research on DA of reading and reading-related skills: specifically, we asked to what extent DA is able to predict variance in the growth of children’s reading skills or to predict reading performance at a future point in time, and to what extent DA may tap into variance unexplained by static assessment. A total of 17 articles met inclusion criteria for the review, representing 18 peer-reviewed studies published between 1992 and 2020. Studies were carried out in a range of countries and followed participants between the ages of four and eighteen, the majority administering dynamic tests of code-based skills such as phonological awareness

and decoding, with the remainder focusing on DA of morphological awareness, paired-associate learning, and reading comprehension.

The results of regression modelling support the role of dynamic measures as statistically significant predictors of growth in reading skills. In some cases, this proportion of variance was rather large (e.g., 72–82% in the studies of Horbach et al., 2015, 2018), though in studies controlling for various static predictors of reading and reading-related skills, the median amount of variance explained by DA specifically was approximately 5%. The contribution of dynamic measures varied somewhat across different reading outcomes. After accounting for static predictors, DA of PA explained 4–21% of variance in reading accuracy and 1–9% in reading fluency outcomes. Similar results were found for DA of decoding, typically predicting between 1% and 17% of variance in reading accuracy and 4–8% in fluency outcomes. Reading fluency outcomes may be considered slightly more distal to the skills targeted by dynamic PA and decoding tasks, thus accounting for the relatively lower predictive power of DA here. Two dynamic decoding studies examined a yet more distal outcome in the form of reading comprehension (Fuchs et al., 2011; Horbach et al., 2018). The results of Fuchs et al. (2011), particularly, provide robust evidence for the ability of a dynamically administered measure of decoding ability to predict unique variance in a reading comprehension measured over a year later; though only a very small proportion (1%), this nonetheless suggests that dynamic decoding score (as indexed here through graduated prompts), was reliably tapping into variance unexplained by static measures.

Reading comprehension outcomes were also examined in studies of dynamic morphological awareness (explaining 4–33.4% of variance after static measures). The relatively older participants in these studies engaged in dynamic tasks including grammatical judgement, cloze exercises, and justifying responses through evidence of morphological awareness. Dynamic MA scores, all indexed through a graduated-prompts procedure, explained 4–33.4% of variance in reading comprehension after accounting for static MA task performance, though in one study this effect was no longer statistically significant when nonverbal reasoning was included in the statistical model (Navarro & Mourgues-Codern, 2018). Only one study sought to predict reading comprehension performance with a DA of the same construct. In their computerised adaptive reading comprehension test, Gruhn et al. (2020) found that second-attempt responses to incorrectly answered questions did not predict future reading comprehension performance. This null finding may have been due to the limited number of attempts at incorrectly answered questions (1 hint) and the relatively short interval between time points of three months. Despite this, there is support elsewhere for the predictive validity of DA in concurrent reading comprehension performance. In Elleman et al. (2011), children in second grade were taught ‘reading detective’ strategies and assessed on their application across a range of passages. When entered after static measures of word reading accuracy and vocabulary, a dynamic score derived from a graduated prompting procedure predicted a small but unique amount of variance (4%) in concurrent reading comprehension performance, again providing evidence for the ability of DA to tap into variance unexplained by traditional static predictors. More longitudinal work is needed to establish the predictive power of DA of reading comprehension.

While a number of studies in the review did recruit participants from CLD backgrounds (e.g., bi-/multilingual or those from socio-economically disadvantaged homes), none sought to explicitly compare the predictive validity of DA according to these variables. Although the present review is not able to address this issue, some tentative evidence for the differential sensitivity of DA for children from CLD backgrounds was found in the systematic review of Dixon et al. (in press). In particular, in a longitudinal study spanning kindergarten to fifth grade, Petersen et al., (2016, 2018) found that static tests afforded particularly poor classification sensitivity for a Hispanic subsample and that the addition of a dynamic decoding variable resulted in relatively larger improvements in sensitivity than for a Caucasian comparison group. It remains to be seen whether such differential sensitivity of DA for CLD populations applies in the longitudinal modelling of reading development.

One particular concern with static assessments of reading is the risk of a highly skewed distribution (e.g., floor effect), particularly for younger children shortly after the onset of formal literacy instruction (Catts et al., 2009). This results in lack of variation and therefore poses challenges for predicting variance in growth. Statistically, linear regression models make no assumptions concerning the distribution of independent variables (though a highly skewed predictor may increase the risk of outliers and result in larger residual variance; Field et al., 2012); instead, it is floor effects in *dependent* variables (i.e., reading outcome measures themselves) that pose a more serious threat. While some studies in the review did report trends towards floor effects in independent variables of PA (Bridges & Catts, 2011; Cho et al., 2017, 2020; Spector, 1992) and letter knowledge (Horbach et al., 2018), only two studies reported such trends in dependent variables of word reading outcome measures (Gellert & Elbro, 2017a; Spector, 1992). Therefore, the issue of floor effects did not appear to be widespread in the sample of studies synthesised here, though such a conclusion is tentative given inconsistent reporting of skewness statistics across studies.

As alluded to above, DA of reading and related skills has also been used for the purposes of classifying children at risk of developing reading difficulties (Aravena et al., 2018; Cho et al., 2020; Bridges & Catts, 2011; Compton et al., 2010; Gellert & Elbro, 2017a, 2017b, 2018; O'Connor & Jenkins, 1999; Petersen et al., 2018; see Dixon et al., in press for a recent review). Here again, dynamically administered measures of learning potential have been found to explain unique variance over and above that of static measures in future risk status (Dixon et al., in press). However, classification analysis using logistic regression modelling necessarily imposes an arbitrary cut-off (e.g.,  $-1$  or  $-1.5$  SD below the sample or norming population mean). In contrast, the results of the present review support the predictive validity of DA to measure growth in reading skills in a continuous fashion without the need for such arbitrary discrimination. Consequently, DA may offer more sensitive predictions regarding rate of change in reading skills over time as well as the likelihood of reading difficulties.

Seven studies reported data on the administration time of DA procedures, ranging from 8–10 minutes (Bridges & Catts, 2011) to 40–60 minutes in total (Coventry et al., 2011), though most procedures lasted between 15 to 20 minutes. It may be questioned to what extent DA is justified as part of a screening battery given its time

requirements and the small to modest amount of variance it explains in future reading performance (O'Connor & Jenkins, 1999). However, it may also be the case that such time costs are justified particularly for children from CLD backgrounds whose future reading performance may be more accurately predicted by measures of learning potential. The time-consuming nature of DA may be addressed to some extent by computerisation, allowing for automatic scoring and standardisation of feedback (only seven studies in the current review employed computerised measures). However, there is also some evidence to suggest that computer-mediated feedback may result in poorer performance relative to human-mediated feedback (Golke et al., 2015). Indeed, the combination of a computer-delivered task and the presence of a 'knowledgeable other' may result in best performance and it may be argued that attentive and individualised (and therefore time-consuming) feedback represents a crucial mechanism by which DA quantifies learning potential, especially within the theoretical framework of a zone of proximal development (Dumas et al., 2020).

### Limitations and future directions

We searched five electronic databases for records pertaining to DA and reading skills, and chose not to impose exclusion criteria based on year of publication. To some extent, the results of the present study may have been limited or biased by what was not included in the review. Firstly, due to resource limitations we were unable to implement a comprehensive grey literature search strategy and therefore the review may have omitted relevant work from non-peer-reviewed sources (e.g., dissertations, preprints, etc.). Secondly, given the review's focus on the ability of DA to predict growth in reading or reading at a future point in time, at the full-text screening stage we excluded a number of studies with cross-sectional designs (e.g., Elleman et al., 2011). Although these studies did not serve to answer our research question, they are likely to be informative concerning the nature of shared and unique variance between static and dynamic measures of reading and related skills. Thirdly, some relevant studies may have been omitted due to the search criteria we employed. A number of different terms are used to describe dynamic tasks in the literature, and although we tried to incorporate many of these in our search (including 'mediated learning', 'interactive assessment', and 'learning potential'; see Method), the inclusion of other terms such as 'paired-associate learning' (PAL) may have led to different results. PAL in particular has been shown to predict unique variance in reading skills cross-sectionally (Warmington & Hulme, 2012) as well as spelling skills longitudinally (Nielsen & Juul, 2016). As a result, future reviews of DA and reading may seek to incorporate PAL more fully into search terms in an effort to identify studies which contrast PAL with traditional static measures to predict growth in reading skills over time.

We chose only to include studies in the review which explicitly contrasted the contributions made by static and dynamic measures in predicting future reading outcomes. Although not included in the review for this reason, three studies nevertheless warrant mention for insights they may provide. Petersen and Gillam (2015) used a dynamic decoding modifiability score in kindergarten to predict reading outcomes

in first grade among a cohort of Hispanic students. As no static measures were included in regression models, this score accounted for a relatively large amount of variance in word reading fluency (19%) and word reading accuracy (24%). It is important to note that this DA (taking only five minutes to administer) has been shown to predict variance in reading outcomes over and above static measures elsewhere, for instance in the classification of risk status for reading disorder (Petersen et al., 2018). We also identified two studies using a dynamic working memory measure to predict later reading outcomes (Swanson, 2010, 2011). The Swanson Cognitive Processing Test (S-CPT) consists of 11 subtests measuring verbal and nonverbal working memory and is shown in these studies to predict significant variance in reading achievement contemporaneously (7–32%), as well as its growth over a period of three years (40–74%). While neither of these studies is informative for our research question, they do provide some evidence for the validity of dynamic assessments in their shared relationship with reading outcomes, and the work of Petersen and colleagues in particular indicates the feasibility of a very brief dynamic measure in contrast to the often lengthy procedures found elsewhere in the literature.

Despite claims that DA provides a less biased form of assessment for children from CLD backgrounds, we were unable to assess the differential sensitivity of DA of reading for this purpose, though another recent review did find tentative evidence of this in the context of classification of at-risk status (Dixon et al., in press). This is an empirical question, and further research may conduct subgroup analyses or include interaction terms in models for socio-economic status or second language learning status. Lastly, although all studies in the present review adopted a longitudinal design, they provide a relatively limited developmental window through which to assess the predictive validity of DA. Studies of dynamic PA and decoding typically focused on the period between preschool and the end of first grade, with the remaining studies recruiting children in the later primary/elementary school grades. Consequently, there appears to be little work looking at the predictive validity of DA in the transition between early and later reading development. In line with Gellert and Elbro (2017a)'s finding that the predictive power of DA appears to decrease when administered beyond the first part of first grade, future work may seek to follow participants over a longer developmental period and assess the applicability of this finding to other educational contexts and more distal reading outcomes.

## Conclusion

Traditional static reading assessments focus on pre-existing knowledge and skills, but variation in learning opportunities and experiences may bias the ability of static tests to predict children's developmental trajectories. DA instead focuses upon a child's potential to learn with feedback. The results of this systematic review suggest that DA does tap into variance in the growth of reading skills that is unexplained by static tests—particularly for word-level reading outcomes, but with some evidence also for early reading comprehension. Additional work is required to address differential sensitivity in the predictive validity of dynamic test procedures for children

from CLD backgrounds, ideally with a longer developmental window and examination of proximal as well as distal reading outcome measures.

**Funding** Funding was provided by Nuffield Foundation (Grant No. EDO/43441).

**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

### Asterisks indicate studies included in the systematic review.

- \*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>
- Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2018). Predicting individual differences in reading and spelling skill with artificial script-based letter-speech sound training. *Journal of Learning Disabilities, 51*(6), 552–564. <https://doi.org/10.1177/0022219417715407>
- \*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>
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education, 41*(4), 254–270. <https://doi.org/10.1177/0022466907310366>
- 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>
- Catts, H., Adlof, S., & Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278–293. [https://doi.org/10.1044/1092-4388\(2006/023\)](https://doi.org/10.1044/1092-4388(2006/023))
- Catts, H., Petscher, Y., Schatschneider, C., Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163–176. <https://doi.org/10.1177/0022219408326219>
- \*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>
- Dixon, C., Oxley, E., Gellert, S. A., & Nash, H. (in press). Does dynamic assessment offer an alternative approach to identifying reading disorder? A systematic review. *Journal of Learning Disabilities*.
- Dumas, D., McNeish, D., & Greene, J. A. (2020). Dynamic measurement: A theoretical–psychometric paradigm for modern educational psychology. *Educational Psychologist*, 55(2), 88–105. <https://doi.org/10.1080/00461520.2020.1744150>
- 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>.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
- \*Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, 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>
- Gamer, M., Lemon, J., Fellows, I. & Singh, P. (2019). IRR: Various Coefficients of interrater reliability and agreement. <https://CRAN.R-project.org/package=irr>
- 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>
- Gellert, A., & 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(2), 126–144. <https://doi.org/10.1007/s11881-018-0159-9>
- Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learning and Instruction*, 39, 123–136. <https://doi.org/10.1016/j.learninstruc.2015.05.009>
- 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., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *The American Psychologist*, 75(1), 37–51. <https://doi.org/10.1037/amp0000452>
- 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*. <https://doi.org/10.1016/j.lindif.2020.101923>
- \*Hamavandi, M., Rezai, M. J., & Mazdayasna, G. (2017). Dynamic assessment of morphological awareness in the EFL context. *Cogent Education*, 4(1), 1324254. <https://doi.org/10.1080/2331186X.2017.1324254>
- Haywood, H. C. (1997). Interactive assessment. In R. Taylor (Ed.), *Assessment of individuals with mental retardation*. Singular Publishing Group.
- Haywood, H. C., & Lidz, C. S. (2007). *Dynamic assessment in practice: Clinical and educational applications*. Cambridge: Cambridge University Press.
- Hill, J. (2015). How useful is dynamic assessment as an approach to service delivery within educational psychology? *Educational Psychology in Practice*, 31(2), 127–136. <https://doi.org/10.1080/02667363.2014.994737>

- \*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., & Günther, 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>
- Kelso, K., Whitworth, A., Parsons, R., & Leitaio, S. (2020). Hidden reading difficulties: Identifying children who are poor comprehenders. *Learning Disability Quarterly, 43*(2), 147–157. <https://doi.org/10.1177/0731948720961766>
- \*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>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical. *Biometrics, 33*(1), 159–174. <https://doi.org/10.2307/2529310>
- Larsen, J. A., & Nippold, M. A. (2007). Morphological analysis in school-age children: Dynamic assessment of a word learning strategy. *Language, Speech, and Hearing Services in Schools, 38*(3), 201–212. <https://doi.org/10.1044/0161-1461%282007/021%29>
- 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>
- McLaughlin, M. J., Speirs, K. E., & Shenassa, E. D. (2014). Reading Disability and Adult Attained Education and Income: Evidence From a 30-Year Longitudinal Study of a Population-Based Sample. *Journal of Learning Disabilities, 47*(4), 374–386. <https://doi.org/10.1177/0022219412458323>
- Murphy, V., & Unthiah, A. (2015). *A systematic review of intervention research examining English language and literacy development in children with English as an Additional Language (EAL)*. University of Oxford.
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665–681. <https://doi.org/10.1037/0012-1649.40.5.665>
- \*Navarro, J.-J., & Mourgues-Codern, C. V. (2018). Dynamic assessment and computerized adaptive tests in reading processes. *Journal of Cognitive Education and Psychology, 17*, 70–96. <https://doi.org/10.1891/1945-8959.17.1.70>
- \*Navarro, J.-J., Mourgues-Codern, C., Guzman, E., Rodriguez-Ortiz, I. R., Conejo, R., Sanchez-Gutierrez, C., de la Fuente, J., Martella, D., & Saracostti, M. (2018). Integrating curriculum-based dynamic assessment in computerized adaptive testing: Development and predictive validity of the EDPL-BAI battery on reading competence. *Frontiers in Psychology, 9*, 1–19. <https://doi.org/10.3389/fpsyg.2018.01492>
- Nielsen, A.-M.V., & Juul, H. (2016). Predictors of early versus later spelling development in Danish. *Reading and Writing, 29*(2), 245–266. <https://doi.org/10.1007/s11145-015-9591-y>
- 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](https://doi.org/10.1207/s1532799xssr0302_4)
- 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>
- 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*, 875–888. [https://doi.org/10.1044/2018\\_LSHSS-DYSLC-18-0021](https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0021)
- Petersen, D. B., Penny, T., Spencer, T. D., & Foster, M. (2020). The classification accuracy of a dynamic assessment of inferential word learning for bilingual English/Spanish-speaking school-age children.



- Language, Speech, and Hearing Services in Schools*, 51(1), 144–164. [https://doi.org/10.1044/2019\\_LSHSS-18-0129](https://doi.org/10.1044/2019_LSHSS-18-0129)
- \*Poulsen, M., & Elbro, C. (2018). The short- and long-term predictions of reading accuracy and speed from paired-associate learning. *Journal of Experimental Child Psychology*, 174, 77–89. <https://doi.org/10.1016/j.jecp.2018.05.007>
- 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*, 48(9), 923–931. <https://doi.org/10.1111/j.1469-7610.2007.01763.x>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Sirriyeh, R., Lawton, R., Gardner, P., & Armitage, G. (2012). Reviewing studies with diverse designs: The development and evaluation of a new tool. *Journal of Evaluation in Clinical Practice*, 18(4), 746–752. <https://doi.org/10.1111/j.1365-2753.2011.01662.x>
- \*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>
- Sternberg, R., & Grigorenko, E. (2002). *Dynamic Testing: The nature and measurement of learning potential*. Cambridge University Press.
- Swanson, H. L. (2010). Does the dynamic testing of working memory predict growth in nonword fluency and vocabulary in children with reading disabilities? *Journal of Cognitive Education and Psychology*, 9(2), 139–165. <https://doi.org/10.1891/1945-8959.9.2.139>
- Swanson, H. L. (2011). Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities*, 44(4), 358–371. <https://doi.org/10.1177/0022219411407866>
- 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>
- van Bergen, E., Snowling, M. J., de Zeeuw, E. L., van Beijsterveldt, C. E. M., Dolan, C. V., & Boomsma, D. I. (2018). Why do children read more? The influence of reading ability on voluntary reading practices. *Journal of Child Psychology and Psychiatry*, 59(11), 1205–1214. <https://doi.org/10.1111/jcpp.12910>
- Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22, 407–423. <https://doi.org/10.1002/acp.1414>
- Warmington, M., & Hulme, C. (2012). Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading*, 16, 45–62.

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

## Authors and Affiliations

Chris Dixon<sup>1</sup>  · Emily Oxley<sup>1</sup> · Anna Steenberg Gellert<sup>2</sup> · Hannah Nash<sup>1</sup>

Emily Oxley  
e.oxley1@leeds.ac.uk

Anna Steenberg Gellert  
agellert@hum.ku.dk

Hannah Nash  
h.nash@leeds.ac.uk

<sup>1</sup> School of Psychology, University of Leeds, Woodhouse, Leeds LS2 9JZ, UK

- <sup>2</sup> Department of Nordic Studies and Linguistics, University of Copenhagen, 120 Njalsgade,  
2300 Copenhagen, Denmark