DART
Dynamic Assessment of Reading Test

Dynamic Assessment of Reading Test (DART) Project Summary Report

Can dynamic assessment predict reading development and be used to identify reading difficulties?

Nash, H., Clarke, P., Dixon, C., Oxley. E., Grainger, K., Weighall, A., & Gellert, A.S.

Nuffield Foundation

UNIVERSITY OF LEEDS

Contents

DOI: https://doi.org/10.48785/100/128

Executive Summary

Context

A quarter of children leave primary school without having achieved the expected standard in reading (Department for Education, 2022). It is crucial that we identify children at risk of reading difficulties early in their school experience, so that appropriate support can be put in place. Current screening practices involve assessing reading ability itself or related skills such as letter knowledge or vocabulary. However, these measures are static, assessing a child's existing knowledge, which is a product of their ability to learn and their experiences.

Opportunities to learn vary greatly between children; children from disadvantaged backgrounds or those for whom English is an additional language may have had less opportunity to learn the foundation skills of reading in English. Dynamic assessment (Grigorenko & Sternberg, 1998) offers a potentially fairer screening method, measuring a child's capacity to learn while completing a task. In systematic reviews of the existing literature we found that dynamic assessments of reading-related skills explained differences in children's reading growth even after accounting for variance associated with static measures (Dixon et al., 2022b) and achieved good identification accuracy for later reading difficulties, when used alone or in combination with static measures (Dixon et al., 2022a). However, there was a lack of evidence from the UK context, for skills relating to reading comprehension and for children from diverse backgrounds.

Aim

Our systematic reviews of the literature identified the potential for using dynamic assessment in the screening process for reading difficulties. The overarching aim of the DART project was to create computerised dynamic assessments of learning, which if shown to be effective screeners, could be developed in future work as low-cost screeners for use in schools. The research was conducted in three work packages.

1. Dynamic assessment of decoding with children in reception, focusing on early reading ability.
2. Dynamic assessment of sight word learning with children in year three, focusing on the development of more skilled reading.
3. Dynamic assessment of vocabulary learning with children in year four, focusing on reading comprehension.

Research Questions

1. Do dynamic assessments correlate less strongly with socio-economic status and English language proficiency than static measures?
2. Does learning in each of the dynamic assessments predict growth in reading ability over time?
3. Can dynamic assessments accurately screen for later reading difficulties?
a. How do they compare to static measures?
b. Do they improve screening when added to static measures?

Method

Each work package used the same longitudinal design, with two assessment time points. Children were first assessed using a battery of static tests (to measure reading ability and traditional predictors of reading) and one dynamic assessment. At the second time only the reading ability tests needed for diagnostic classification were completed.

Key Findings

Performance on the dynamic assessments was only weakly to moderately correlated with socio-economic status and English language proficiency. The static measures were also only weakly to moderately correlated with socio-economic status but more strongly related to English language proficiency. This suggests that our dynamic assessments offer a less biased approach to screening for the increasing number of children with English as an Additional Language (EAL) entering primary school.

All of our dynamic assessments predicted unique growth in reading ability after controlling for demographic factors and static tests:

1. The dynamic assessment of decoding predicted growth in early word reading.
2. The dynamic assessment of sight word learning predicted growth in reading accuracy and fluency.
3. The dynamic assessment of vocabulary learning predicted growth in reading comprehension.

All of our dynamic assessments achieved excellent or outstanding levels of accuracy as screeners for later reading difficulties and two showed potential to add value to a battery of static tests for children with EAL (decoding and vocabulary learning). Data from a sub-sample of non-readers suggest that administering the decoding task earlier in the school year, as originally planned, could improve screening accuracy for all reception children.

Future Directions

The DART project has provided evidence of 'proof of concept'.

We now need to work in partnership with educators to establish how dynamic assessments of decoding and vocabulary learning may fit within existing practices in schools and determine for what age group and when in the school year our dynamic assessments should be administered to maximise their value.

An additional cycle of participatory co-design work with children and educators is needed to refine the presentation and delivery of the dynamic assessments on an accessible, stable and low-cost platform, that would be suitable for use in schools.

How does reading develop?

The foundations

Learning to read in English is complex and usually begins with mastering the alphabet and learning how to decode.

Decoding is the ability to recognise patterns in written text (graphemes) and convert these into sounds (phonemes). A child's letter-sound knowledge, phoneme awareness and rapid automatised naming ability are important for decoding success (Caravolas et al. 2013). Letter-sound knowledge means recognition of letters and the phonemes they represent. Phoneme awareness refers to the identification of the smallest units of sound in speech. Rapid automatised naming is the ability to name sequences of letters, numbers, objects or colours, fluently and at speed. Using this knowledge and these skills, it is possible to transform some written English into its spoken form.

English is not, however, a transparent orthography, meaning that learning to read in English is complicated by words that do not follow the typical rules of pronunciation. These are often referred to as exception words and include examples such as 'island' and 'colonel'. Research suggests these are read with support from vocabulary knowledge (e.g., Nash, Davies & Ricketts, 2021). So if a child has previously encountered the word 'island' in spoken language, and has stored it in memory, they can use this knowledge to help decode it.

Becoming a skilled reader

To become a skilled reader a child needs to progress from decoding to recognising words and this process is called orthographic learning (Castles & Nation, 2006). Orthographic learning is underpinned by the development of a store of written words, called a lexicon, which is driven by decoding and supported by vocabulary knowledge (Share, 1995; Wang, Castles, Nickels & Nation 2011).

Once a written word has been recognised, spoken language skills are used to understand the text. Evidence from longitudinal studies, which track childrens development over time, (Muter, Hulme, Snowling & Stevenson, 2004) and intervention research, which tests the effectiveness of teaching methods, (Clarke, Snowling, Truelove & Hulme, 2010) shows that vocabulary knowledge is particularly important for reading comprehension.

Why do we need to assess reading?

For educational and social inclusion

26% of children in 2022 left primary school without having achieved the government expected standard in reading (Department for Education, 2022). This is concerning as research has shown that low literacy is associated with a range of poor life outcomes (Morrisroe, 2014). The number of children with English as an Additional Language (EAL) (21.2% in 2022) and children receiving Free School Meals (FSM) (22.5% in 2022; Schools, pupils and their characteristics: June 2022 DfE) continues to rise. Children from these backgrounds are at a potential disadvantage entering formal education because their home learning environments may not provide sufficient opportunities to develop the English language skills which underpin reading development.

Gaps in attainment between children from low Socio-Economic Status (SES) groups and their peers are evident in primary school. For example, in England, only 43% of 11 year olds classed as disadvantaged achieved expected standards in reading, writing and maths in 2022, compared to 65% of children who were not classed as disadvantaged. The 'disadvantage gap' (the difference in mean percentile rank of those classed as disadvantaged and their peers) is growing. It was calculated as 2.91 in 2019 and 3.21 in 2022 (National curriculum assessments: Key Stage 2, October 2022 update, DfE). These gaps persist as children progress through the school system. In 2022 the percentage of secondary school pupils classed as disadvantaged attaining 5 or above in English and Maths was 29.5% compared with 56.8% for non-disadvantaged peers. This corresponds to a disadvantage gap of 3.84, which has risen from 3.7 in 2019 (Key Stage 4 Performance, 2022, DfE).

With regards the attainment of EAL pupils, historically gaps have been reported between their performance and that of their peers, with fewer EAL pupils achieving the expected levels for reading at Key Stage 2 or in English GCSE (Strand, Malmberg & Hall, 2015). However, more recent data suggests that gaps are closing and in 2022 there was only a 2% gap in reading performance at key stage 2 between EAL pupils and their peers. The direction of the gap appears to have switched at GCSE level with EAL pupils outperfoming their peers by 4% in English and Maths. EAL pupils are a diverse population and variation in attainment has been linked to English language proficiency (Demie, 2018).

To identify reading difficulties

Reading difficulties are commonly observed in school aged children and can be categorised into three main profiles, dyslexic, poor comprehender and poor reader.

The dyslexic profile, characterised by inaccurate or slow, effortful word reading (DSM 5, American Psychiatric Association, 2013), affects 10-17% of children (Pennington et al. 2012; Rutter et al., 2004). It's thought to arise from a phonological impairment, which delays letter sound learning (Melby-Lervag, Lyster & Hulme, 2012), leading to difficulties decoding words and in orthographic learning (e.g., Bailey, Manis, Pedersen & Seidenberg, 2004).

Conversely, learners who show the poor comprehender profile are able to read words aloud efficiently, but have difficulties understanding what they have read, a profile that often goes unnoticed in the classroom but affects around 10% of children (Keenan et al. 2014) and a higher proportion of children with EAL (up to 18% Li et al., 2021). Poor comprehenders have been shown to have a range of oral language difficulties and vocabulary knowledge has been highlighted as a particular area of weakness (Clarke, Snowling, Truelove & Hulme, 2010).

Some children experience difficulties with both word reading and comprehension and have a generally poor reader profile.

The 'gold standard' assessments for identifying these profiles are typically administered by specialists such as educational psychologists, special educational needs co-ordinators and dyslexia specialist tutors. Whilst comprehensive, they are time consuming and can be costly and difficult to access. Screening tools enable schools to gauge which children may require further assessment and support. They can be administered more easily and quickly to more children, earlier in development. Currently screeners measure a child's existing reading-related skills and knowledge, however, we propose that this is problematic and that a more dynamic approach (Grigorenko & Strenberg, 1998) is required.

How should we assess reading?

Static vs. dynamic assessment

Static

- Assesses existing knowledge which is the product of a child's capacity to learn plus their learning experiences
- Captures developed ability i.e. what a child has learned up until the day of the test

Example:

Letter Knowledge: How many letters can a child provide the accompanying speech sounds for?

Dynamic

- Assesses a child's capacity to learn and their learning potential
- Captures latent capacity i.e. what a child can achieve with assistance during the test

Example:

Learning novel letters: How well can a child learn to pair novel letter-like shapes with speech sounds?

The case for dynamic assessment

Early screening to identify children at risk of reading difficulties currently involves assessing reading-related skills such as phoneme awareness, letter-sound knowledge, rapid automatised naming and vocabulary. However, traditional screening measures are static and assess a child's existing knowledge, which is a product of their capacity to learn and their experience. So as well as identifying children with learning difficulties they will also pick up children who have had less opportunity to learn English, such as children of low socioeconomic status (SES) and those for whom English is an additional language (EAL).

That traditional screeners are not sensitive to variations in children's home learning experiences is problematic for children from culturally and linguistically diverse backgrounds (Peña & Halle, 2011).

Static tests used at, or shortly after, the onset of formal reading instruction can also be problematic due to floor effects (Catts et al., 2009). A floor effect occurs when many children score zero, or close to zero, on a test and this can make it difficult to accurately gauge their future reading ability.

Dyslexic and poor comprehender profiles are characterised by difficulties with language learning that impact upon reading. Assessing language learning ability can provide useful information to inform targeted support at the source of learners difficulties.

There is emerging evidence that dynamic assessment can result in higher classification accuracy of reading difficulties, than traditional static tests alone, including for bilingual children (Petersen & Gillam, 2015).

Dynamic assessments could therefore provide a less culturally biased and more accurate screening tool for the identification of children at risk of developing reading difficulties.

How is assessment accuracy calculated?

The process

To investigate whether or not an assessment could be used as a screening tool to identify learners at risk of developing reading difficulties, there are a series of steps to follow:

1. The type of reading difficulty that the screening tool is intended to identify is specified.
2. The diagnostic criteria that will be used to decide whether a learner has a reading difficulty is determined.
3. The current existing 'gold standard' assessments used to diagnose the reading difficulty are identified.
4. A large sample of learners are assessed using the potential screening tool and then the 'gold standard' assessments.
5. Learners are categorised using the diagnostic criteria based on their 'gold standard' assessment performance.
6. Learners are categorised according to their performance on the potential screening tool.
7. The categorisations are compared to determine the extent to which they are in agreement.

Assessing accuracy

Three main concepts are referred to when assessing the accuracy of a screening tool:

⊕ Sensitivity– proportion of True Positives (TP)

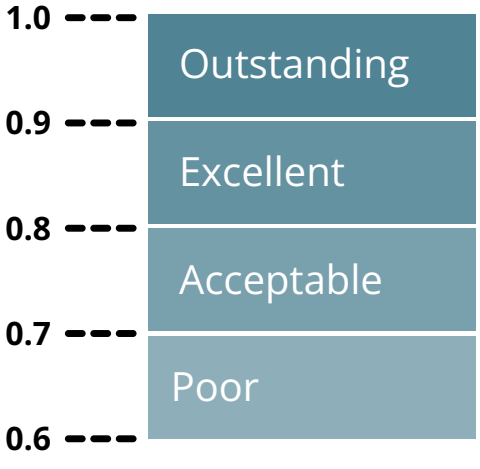
- This is the proportion of those classified using the 'gold standard' assessments as having a reading difficulty, that are also identified by the screening tool as being at risk of a reading difficulty.

⊖ Specificity– proportion of True Negatives (TN)

- This is the proportion of those classified using the 'gold standard' assessments as not having a reading difficulty, that are also identified by the screening tool as not being at risk of a reading difficulty.

? AUC (Area under the curve)

- This is calculated using the sensitivity and specificity values to summarise the likelihood of correct diagnosis (values all range from 0=not at all accurate to 1=complete accuracy).
- The AUC values can be used to judge the acceptability of the accuracy of the screening tool. A value of 0.7 or higher is considered acceptable (Hosmer et al. 2013).



What does previous research tell us?

Systematic Reviews

A systematic review involves an exhaustive search of research studies to provide an up to date synthesis of relevant evidence. Two systematic reviews were conducted as part of the DART project. Both reviews were restricted to including only peer-reviewed publications, published in English.

Review 1 (Dixon et al., 2022a)

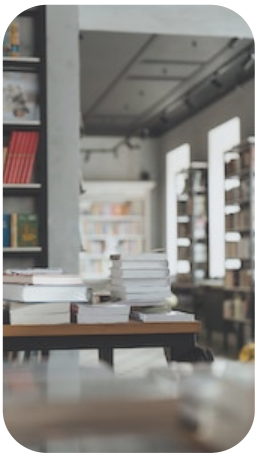
How well can dynamic assessments of reading and reading-related constructs **accurately identify** children who have, or who at risk of having, reading difficulties?

15 studies were included:

- 8 on Decoding
- 5 on Phonological awareness
- 2 on Working memory

Dynamic assessments can achieve good classification accuracy of reading difficulties both when used alone or when used in addition to traditional static tests.

AUC values were reported to be above 0.8.



Review 2 (Dixon et al., 2022b)

How well can dynamic assessments of reading and reading-related constructs accurately **predict growth** in reading?

18 studies were included:

- 7 on Decoding
- 6 on Phonological awareness
- 3 on Morphological (word structure) awareness
- 1 on Nonword learning
- 1 on Reading comprehension

Dynamic assessments explain growth in reading skills. Scores obtained on dynamic assessments of phonological awareness predicted between 4% and 21% of unique variance associated with growth in reading accuracy. Scores obtained on dynamic assessments of decoding predicted between 1% and 17% of unique variance associated with growth in reading accuracy. These values evidence the potential of dynamic assessment, but to fully test the added value a comprehensive analysis including a range of static tests is required.

Insights from the systematic reviews

Dynamic assessment shows promise for the early identification of children at risk of developing reading difficulties and taps into the growth of reading skills.

Dynamic assessment can be time-consuming and complex in terms of administration and scoring. Computerisation of dynamic assessments has the potential to address these issues.

The extent to which dynamic assessment could add value to the identification of reading difficulties in Culturally and Linguistic Diverse (CLD) populations requires further exploration.

Further research is needed to investigate the potential use of dynamic assessments of vocabulary and reading comprehension and to examine whether dynamic assessment can predict growth in reading in the transition between early and later reading development.

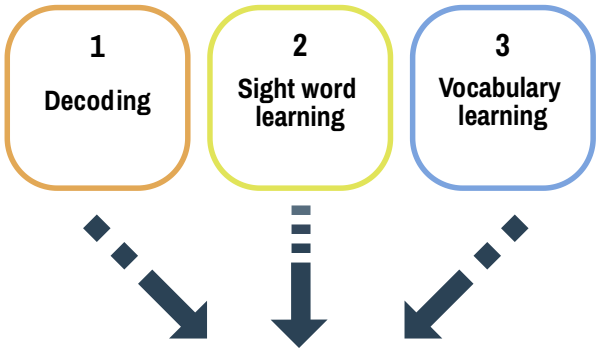
Project aims

Overview

The overarching aim of the DART project was to develop and trial dynamic assessments of language learning in order to examine the extent to which they have potential for use as screening tools to identify reading difficulties. The intention throughout was to create dynamic assessments that could potentially be used in the future by educators in schools, at low cost and with minimal training.

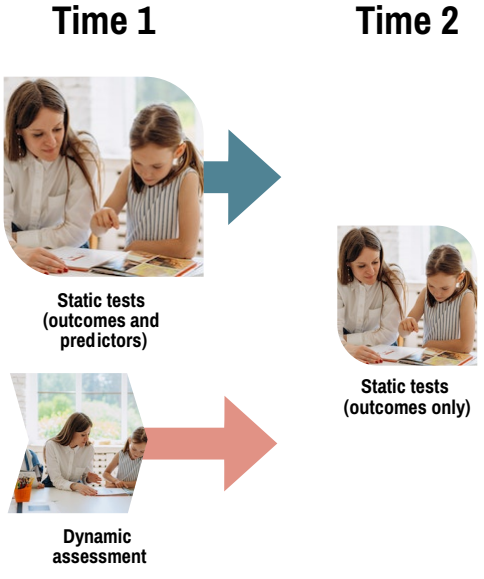
The DART project activities were carried out within three work packages. The research questions addressed by each work package were the same, what differed was the type of dynamic assessment being trialled.

Work packages



Design

Each work package used the same longitudinal design. With two assessment time points at least ten months apart. Children were first assessed using a battery of static tests (to measure reading ability and predictors of reading) and one dynamic assessment. At the second time point a smaller battery including only the reading ability tests needed for diagnostic classification was completed. A different cohort of children took part in each work package.



Research questions

1. **Do dynamic assessments correlate less strongly than static tests with SES and English language proficiency?**

2. **Does learning in each dynamic assessment predict growth in reading ability over time?**

3. **Can dynamic assessments accurately screen for later reading difficulties?**

How do dynamic assessments compare in terms of screening accuracy to static assessments?

Do the dynamic assessments improve screening accuracy when added to static assessments?

Hypotheses

The dynamic assessment will not correlate strongly with SES or English language proficiency at assessment time point one.

The dynamic assessment will predict growth in reading ability between assessment time points one and two.

The dynamic assessment will achieve acceptable levels of screening accuracy (AUC 70 or above).

No hypotheses as these were exploratory questions.

Project team

Dr Hannah Nash

Principal Investigator
University of Leeds

Dr Chris Dixon

Research Fellow
University of Leeds

Dr Paula Clarke

Co-Investigator
University of Leeds

Dr Emily Oxley

Research Assistant
University of Leeds

Dr Anna Weighall

Co-Investigator
University of Sheffield

Katy Grainger

Research Assistant
University of Leeds

Dr Anna Steenberg Gellert

Co-Investigator
University of Copenhagen

Advisory Board

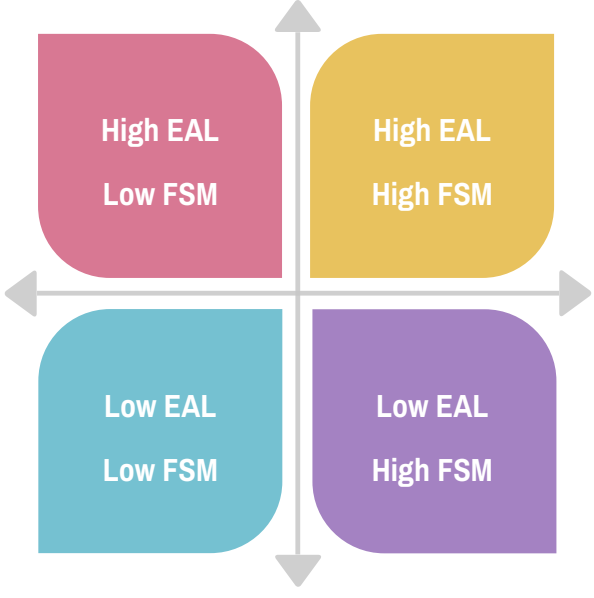
Prof. Cecile De Cat, University of Leeds
Dr Yvonne Griffiths, University of Leeds
Prof. Beaton, Mhairi, Leeds Beckett University

Kevin Smith, PATOSS
Pat Payne, Yorkshire Rose Dyslexia
Dr Hazel Trotter, Leeds City Council

Participating schools

Sampling

We categorised schools in Leeds along two continua – proportion of EAL pupils (low to high) and proportion of children receiving FSM (low to high) – and aimed to recruit two schools from each of the 4 resulting quadrants. The intention was to form a representative sample by recruiting similar numbers of pupils in each quadrant.



We successfully recruited 7 schools to the project. We were only able to recruit one school that had a high proportion of EAL pupils coupled with a low proportion of pupils in receipt of FSM, however this was a very large school and therefore contributed a large number of pupils to the sample.

Schools were located across Leeds with the maximum distance from the University being 8 miles. Four were community schools, one was a foundation school, one an academy converter school and one a voluntary aided school. 6 of the schools had a 'Good' Ofsted rating and one an 'Outstanding' rating.

6 schools took part in all three project work packages. 1 school took part in work packages 2 and 3 only.

Schools

1
Pupils on roll: ~750
EAL%: High
FSM%: Low

2
Pupils on roll: ~500
EAL%: High
FSM%: Average

3
Pupils on roll: ~400
EAL%: Average
FSM%: Average

4
Pupils on roll: ~600
EAL%: Low
FSM%: Low

5
Pupils on roll: ~250
EAL%: Low
FSM%: Low

6
Pupils on roll: ~400
EAL%: Low
FSM%: Average

7
Pupils on roll: ~400
EAL%: Low
FSM%: Average

Project essentials

Ethics and data protection

Approval for the research to go ahead was granted by the University of Leeds School of Psychology ethics committee. The reference numbers for each application are as follows:

- Work Package 1 - PSC – 753 17/09/2019
- Work Package 2 - PSC – 767 14/10/2019
- Work Package 3 - PSC – 671 12/04/2019

Schools signed a memorandum of understanding and a data sharing agreement. The data sharing agreement covered them sharing demographic information with us and us sharing standardised test data with them.

We created parent and child information sheets and parents also received the data collection notice. Parents / legal guardians were given the opportunity to request their child opt-out of the study, in line with BPS guidelines for research activities that are similar to normal curriculum activities and pose no harm.

Ongoing assent was obtained from the children and they could choose to withdraw from the research activities without any consequence. The children received a small gift (coloured pencils; rubbers) as a thank you for taking part.

The research assistants were trained and supported by the wider team to ensure consistency of approach. All had enhanced DBS clearance and prior experience of working with primary school age children.

Data were initially transferred using encrypted USBs and later using secure OneDrive links. Data were stored securely on the university server, and all research data were anonymised.

Equipment

Lenovo X280 Thinkpad Laptops with 8th generation intel i5 cores and 12.5 inch screens were used to run the dynamic assessment tasks.

PsychoPy (Peirce, 2007) was used to run the decoding and sight word learning tasks and DMDX (Forster & Forster, 2003) was used to run the vocabulary learning task.

OSF Pre-registration

- Work Package 1 https://doi.org/10.17605/OSF.IO/CJ2GM
- Work Package 2 https://doi.org/10.17605/OSF.IO/PCWRT
- Work Package 3 https://doi.org/10.17605/OSF.IO/ND7M3

Static tests of reading outcomes

Overview

The static tests of reading outcomes chosen for use in the DART project are all standardised assessments that are widely used by practitioners and researchers. They were all administered according to manual instructions by trained research assistants. The tests are described in detail here and the reliability of each test is reported. Reliability is typically calculated using a statistical test called Cronbach's Alpha, as a general guide values of 0.7 or higher indicate good reliability.

Reading accuracy

Word reading ability was measured using two different assessments. In the Early Word Recognition (EWR) subtest of the YARC-ER, children attempt to read aloud a series of 15 regular words and 15 exception words of increasing difficulty (e.g., cat, in, ... biscuit, giant). The test discontinues after 10 consecutive errors.

The Diagnostic Test of Word Reading Processes (DTWRP; Forum for Research in Literacy and Language, 2012) is a more challenging measure of reading accuracy and separately assesses recognition of regular words, exception words, and nonwords, with 30 items each. Testing on each of the three lists discontinues after five consecutive errors.

The DTWRP is standardised on a nationally representative sample of 1,125 children in England aged 5 years 4 months to 12 years 3 months, 12.4% of whom are EAL learners. Internal reliability (alpha) for the complete test of 90 items is .99.

Reading fluency

The Test of Word Reading Efficiency-II (TOWRE-2; Torgesen et al., 2011) consists of two subtests (Sight-Word Efficiency [SWE]; real words), Phonemic Decoding Efficiency [PDE]; nonwords) in which examinees are asked to read aloud as many words as possible from a list within 45 seconds. The score is the total number of words read correctly within the time limit.

The TOWRE-2 is standardised on a US population of individuals aged 6 to 24 years and reports test-retest reliability of between r = .89-.93.

Reading comprehension

The York Assessment of Reading for Comprehension-Primary (YARC-P; Snowling et al., 2011) assessed 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 is determined by each participant's score on a Single Word Reading Test, an accompanying measure of real word reading accuracy. 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-.84.

In addition to age-standardised scores, the YARC-P derives ability scores (ranging from 1 to 91) which take passage difficulty into account. In the present study, standardised scores for accuracy and comprehension were used to determine 'poor comprehender' status, while ability scores were used as a measure of comprehension in the statistical analysis of growth.

Static tests of reading predictors

Overview

The static tests of reading predictors used in the DART project were chosen based on previous studies. Measures of receptive vocabulary and nonverbal ability were included as predictors in all three work-packages. They were included to account for the role of verbal and nonverbal ability in the development of reading. Measures of pre or early reading skills were only included in work package 1; phoneme awareness, letter sound knowledge and rapid automatized naming, because together these skills predict differences in early reading skills in alphabetic languages (e.g., Caravolas et al 2013). These tests were all administered according to manual instructions by trained research assistants.

Receptive vocabulary

The British Picture Vocabulary Scale-III (BPVS-3; Dunn et al., 2009) is a standardised, norm-referenced assessment of receptive vocabulary knowledge. In a multiple-choice format, children are asked to match colour illustrations to stimulus words spoken by the examiner. Testing discontinues once a child makes eight or more errors 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 statistical reliability, however the PPVT-4, the assessment on which the BPVS-III is based, reports a high split-half reliability coefficient of .94 (Dunn & Dunn, 2007).

Nonverbal ability

Raven's coloured progressive matrices (CPM; Pearson, 2008) was administered as a measure of nonverbal reasoning. Children are presented with a series of diagrammatic puzzles and 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 .97.

Phonological awareness

Two subtests of the PhAB-2 (Gibbs & Bodman, 2014) were used as measures of phonological awareness, Alliteration and Blending.

In Alliteration, the child is presented with a series of illustrated coloured slides depicting three objects. After the name of each object is spoken by the examiner (e.g., well, peg, pot), the child is asked to indicate which two objects share the same initial phoneme (e.g. peg and pot). We administered parts 1 and 2 (up to age 6), though part 2 is only attempted if a child scores three or more correct on part 1. Internal reliability (alpha) for parts 1 and 2 is .86.

In Blending, the child hears phonemes spoken by the examiner in sequence (e.g., /f/ /i/ /sh/ [fish]) and is asked "what word do we get when we put these sounds together?" The test consists of two parts, though part 2 is only attempted if at least five items are answered correctly or there are no more than four consecutive incorrect responses on part 1. Internal consistency (alpha) for Blending is .96.

Letter knowledge

The letter-sound knowledge (LSK) subtest of the York Analysis of Reading for Comprehension-Early Reading (YARC-ER; Hulme et al., 2011) is a measure of alphabetic knowledge. Children are shown a series of printed letters and digraphs and asked "what sound does this letter make?" We administered the Extended test, consisting of 26 letters and 6 digraphs. Children attempt all 32 items. The LSK Extended test has an internal reliability of .98. The YARC-ER is standardised on a nationally representative sample of 662 children in the United Kingdom, 16.3% of whom are EAL learners.

Rapid Autonomised Naming

We administered the Picture Naming subtest of the Phonological Assessment Battery-II (PhAB-2; Gibbs & Bodman, 2014) as a measure of rapid automatised naming of objects. Children are presented with a matrix of randomly repeating rows of five objects (ball, hat, door, table, box) and asked to name all items as quickly as possible. To reduce the length and attentional demands of the battery, we administered Part 1 only. Time taken to name all items was recorded in seconds. Internal reliability (alpha) is .89. The PhAB-2 is standardised on a sample of 773 children aged 4 years 7 months to 11 years 11 months in England, Scotland, and Wales (8% with EAL).

Dynamic assessment of decoding

Aims

To create a computerised version of a Danish dynamic decoding assessment (Gellert & Elbro, 2017).

To trial the new dynamic assessment of decoding to determine its utility in predicting reading outcomes and identifying children at risk of reading difficulties, specifically the dyslexic profile.

What did we expect to find?

1. The dynamic assessment of decoding will not correlate strongly with SES or English language proficency at assessment time point one.

2. The dynamic assessment of decoding will predict growth in word reading accuracy between time points one and two.

3. The dynamic assessment of decoding will achieve acceptable levels of screening accuracy (AUC 70 or above).

Study design and participants

Participants were first assessed when they were in Reception. The second assessment time point was approximately 10 months later when they were in Year one. The diagnostic criteria used to categorise the dyslexic profile of reading difficulties was a composite reading score on the YARC-ER EWR and DTWRP single word reading tests in the lowest 15% of the sample at Time 2. There are no agreed assessment based criteria for diagnosing the dyslexic profile, however we have used two well established standardised assessments of reading and our criteria are aligned with the current Diagnostic and Statistical Manual (DSM 5) description of Dyslexia.

Time 1

May-June 2021

n=317

Male: 167 (53%)
Female: 150 (47%)

SEN: 47 (15%)
EAL: 75 (24%)

Average age: 5 years 3 months

10 months

Time 2

March-April 2022

n=286

Male: 146 (51%)
Female: 140 (49%)

SEN: 41 (14%)
EAL: 65 (23%)

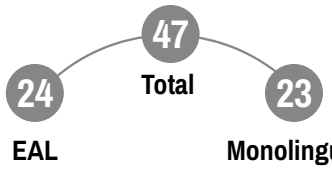
Average age: 6 years 0 months

Reading ability outcomes
Reading accuracy (YARC EWR & DTWRP)

Static predictors of reading
Phonological awareness; Letter knowledge; Rapid automatised naming; Vocabulary; Nonverbal ability.

Dynamic assessment of decoding

Number of children at risk of dyslexic reading profile



Attrition

10%

Moved school: 11
Absent: 17
Other: 1

Reading ability outcomes
Reading accuracy (YARC EWR & DTWRP)

Dynamic assessment of decoding

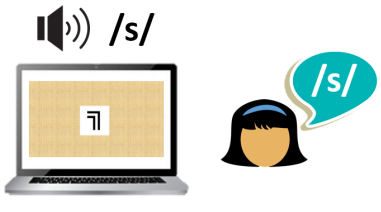
The story context

The goal of this assessment is to learn sound-symbol pairings. It is presented as an adventure story 'Chompy Croc and the Sacred Stone'. Chompy the crocodile is looking for hidden treasure. He needs help to understand a secret code and reveal its location. Illustrations indicate the child's progress through the story and are designed to be motivating and rewarding. The researcher controls progress through the task presented on a computer, recording whether a child's response is correct. Everything else is controlled by the computer program. The time the task takes from start to finish is on average 20 minutes, with a range + or - 10 minutes, depending on the child.

The learning sequence

1. Initial exposure

The children see each symbol and hear its corresponding sound. They immediately repeat each sound and receive spoken feedback via the computer to let them know whether or not they have produced it correctly. The children complete up to 30 trials depending upon how accurately they are able repeat the sounds.



2. Blending training

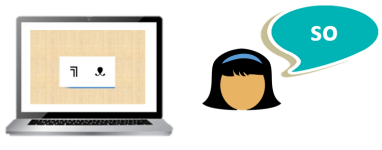
The children watch an animation of two symbol-sound pairs. Each sequence has three steps: 1) The symbols start far apart at the edges of the screen 2) The symbols move closer together 3) The symbols join together in the centre of the screen. A blue border appears each time a sound is presented to draw attention to the symbol it represents. A blue animated line appears under the joined symbols when the blended sounds are presented. Spoken feedback is provided throughout the training.

There are four phases to the blending training:

- Phase 1 - Full demonstration - The child watches and listens to the animation, no response is needed.
- Phase 2 - Repetition - There is a pause after each step in the sequence in which the child is asked to repeat the sounds and the final blend.
- Phase 3 - Independent blending - Same as Phase 2 except the child does not hear the final blend, they have to produce it independently.
- Phase 4 - Fully independent - No sounds are presented the child needs to re-produce them from memory and the final blend independently.

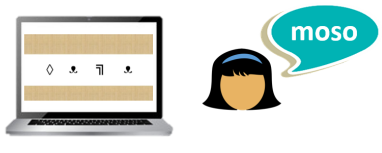
3. Blending test

First the children see the symbols and are asked to recall their corresponding sounds. If they are unable to do this then the researcher provides the sounds for them. Then they are asked to blend the two sounds depicted in the symbols. This process is supported by the same animation used in the blending training. The childrens responses are scored as 'correct' (1 point) or 'incorrect' (0 points). Spoken feedback is provided throughout. The children complete a maximum of 20 trials presented in five blocks of four items each. The test discontinues if a child scores 4/4 on two consecutive blocks.



4. Reading post test

This part is only completed by children who scored 4/4 on two consecutive blocks, or 16/20 or more overall in the blending test. It begins with a practice item which the child watches and repeats. Then they complete 12 test items independently. Each item is a sequence of symbols that need to be blended to form a nonsense word. The sequences become increasingly longer and more complex. Responses are scored as 'correct' (1 point) or 'incorrect' (0 points). No feedback is given.



Dynamic assessment of decoding

Did the dynamic assessment of decoding correlate with SES and/or English language proficiency?

- Performance across the blending training trials was not statistically significantly correlated with SES (*r*=.05), and was weakly correlated with English language proficiency (*r*=.27).
- Blending post test scores were weakly correlated with SES (*r*=.37) and moderately correlated with English language proficiency (*r*=.55).
- The static predictors were all weakly correlated with SES (Phonological awareness *r*=.39; Rapid automatised naming *r*=.21; Letter knowledge *r*= .41).
- Two of the static predictors were strongly correlated with English language proficiency (Phonological awareness *r*=.60; Letter knowledge *r*= .62). Rapid automatised naming was not significantly correlated (*r*= -.04).
- The standardised measures of reading ability were weakly correlated with SES (EWR *r*=.39; DTWRP *r*=.36) and moderately correlated with English language proficiency (EWR *r*=.53; DTWRP *r*=.47).

Hypothesis 1 supported

Did the dynamic assessment of decoding predict growth in reading and reading related skills?

- DA of decoding post test scores predicted an additional 6% of variance in word reading growth in the whole sample after the predictive value of the static tests had been accounted for.
- When the data from the 221 monolingual children who took part in this work package were analysed as a subgroup the DA of decoding post test scores continued to predict an additional 6% of unique variance in word reading growth.
- When the data from the 65 children with EAL were analysed as a subgroup, the DA of decoding post test scores did predict additional unique variance but to a lesser extent (approximately 3%).

Hypothesis 2 supported

Can dynamic assessments of decoding accurately screen for reading difficulties?

- 47 children were identified as at risk of developing the dyslexic reading profile, 23 monolingual children and 24 children with EAL.
- From Figure 1 it can be seen that the dynamic assessment of decoding achieved a high level of classification accuracy for the whole sample (as indicated by the AUC, .946), but the static tests achieved a slightly higher level when they were added together (.971). When added to the static tests, the dynamic assessment of decoding did not improve classification accuracy, this is likely to be because the AUC values are already both very high.
- From Figure 2 it can be seen that for children with EAL the dynamic assessment of decoding shows a perfect level of sensitivity (1.00). When added to the static tests, the dynamic assessment of decoding significantly improved the classification accuracy, increasing the AUC from .985 to .989.

Hypothesis 3 supported

Predicting growth in reading

When referring to growth we mean a change between time points 1 and 2. We would usually expect this to be an increase however for some learners there be no noticeable changes. The amount of change will vary between children and our analyses examines the factors that influence this variation and to what extent they have an effect.

The analyses predict reading scores obtained at time 2, first taking into account time 1 performance and then by examining the amount of variation explained by each factor. We consider in turn, demographic factors, then static tests and finally the dynamic assessments.

We are particularly interested in the amount variability predicted by the dynamic assessments after all other factors have been accounted for and we refer to this as **unique variance**.

Dynamic assessment of decoding

	Sensitivity	Specificity	AUC
DA Training	.894	.715	.876
DA Post test	.872	.900	.946
Static EWR	.894	.895	.964
Static LK	.894	.929	.971
Static + DA Training	.894	.933	.971
Static + DA Post test	.957	.858	.972

Whole sample

Figure 1 Sensitivity, specificity and AUC values from statistically significant predictors, obtained using DA decoding measures, static tests and both assessment types combined, and data from the whole sample (n=286).

Monolingual

EAL

	Monolingual Sensitivity	Monolingual Specificity	Monolingual AUC	EAL Sensitivity	EAL Specificity	EAL AUC
DA Training	.913	.747	.870	*Not applicable as not significant*		
DA Post test	.957	.758	.926	1.00	.902	.976
Static EWR	1.00	.793	.956	.958	.902	.984
Static LK	.957	.848	.963	.958	.927	.985
Static + DA Training	.913	.904	.958	*Not applicable as not significant*		
Static + DA Post test	.913	.899	.959	1.00	.927	.989

Figure 2 Sensitivity, specificity and AUC values from statistically significant predictors, obtained using DA decoding measures, static tests and both assessment types combined, and data from the three subgroups (Monolingual n=217, EAL n=65).

Dynamic assessment of sight word learning

Aims

To create a computerised dynamic sight word learning assessment, based on an orthographic learning task we had used in an unpublished study.

To trial the new computerised dynamic assessment of sight word learning to determine its utility in predicting reading outcomes and identifying children at risk of reading difficulties, specifically difficulties learning sight words that are required for fluent reading consistent with a dyslexic reading profile.

What did we expect to find?

1. The dynamic assessment of sight word learning will not correlate strongly with SES or English language proficiency at assessment time point one.

2. The dynamic assessment of sight word learning will predict growth in word reading accuracy and fluency between time points one and two.

3. The dynamic assessment of sight word learning will achieve acceptable levels of screening accuracy (AUC 70 or above).

Study design and participants

Participants were first assessed when they were in Year 3. The second assessment time point was approximately 14 months later when they were in Year 4. The diagnostic criteria used to categorise the dyslexic profile of reading difficulties was a composite reading score on the word reading subtests of the DTWRP (accuracy) and the TOWRE (fluency) in the bottom 15% of the sample at Time 2. As previously mentioned, there are no agreed assessment based criteria for diagnosing the dyslexic profile, however we have used two well established standardised assessments of reading and our criteria are aligned with the current Diagnostic and Statistical Manual (DSM 5) description of Dyslexia.

Time 1

Sept 2019 - March 2020

n=387

Male: 202 (52%)
Female: 185 (48%)

SEN: 56 (15%)
EAL: 126 (33%)

Average age: 7 years 10 months

14 months

Time 2

March-May 2021

n=328

Male: 172 (52%)
Female: 146 (48%)

SEN: 49 (15%)
EAL: 99 (30%)

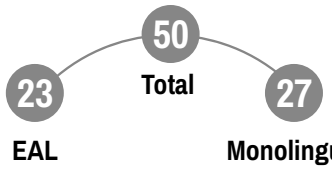
Average age: 9 years 1 month

Reading ability outcomes
Word reading accuracy (DTWRP) and fluency (TOWRE).

Static predictors of reading
Vocabulary; Nonverbal ability, Nonword reading accuracy and fluency.

Dynamic assessment of sight word learning

Number of children at risk of dyslexic reading profile



Attrition

15%

Moved school: 34
Absent: 24
Other: 1

Reading ability outcomes
Reading accuracy (YARC EWR & DTWRP)

Dynamic assessment of sight word learning

The story context

The goal of this assessment is to learn the correct spelling of novel words. It is presented as a story about 'Willow the Witch', who is casting spells using magical words. Illustrations indicate the child's progress through the story and are designed to be motivating and rewarding. The researcher controls progress through the task presented on a computer, recording whether a child's response is correct. Everything else is controlled by the computer program. The time the task takes from start to finish is on average 20 minutes, with a range + or - 10 minutes, depending on the child.

The learning sequence

1. Training trials

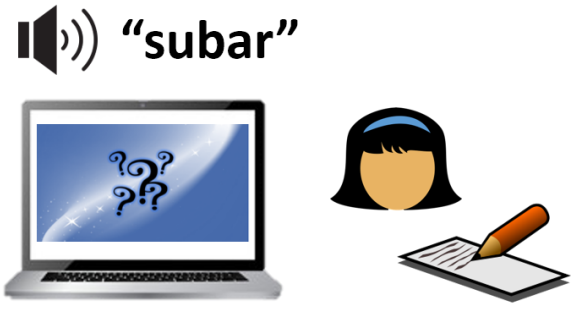
The children see each novel word written on the screen and are asked to read it aloud. There are eight novel words, based on real words, created by changing one sound.

Four words were based on real words with regular pronuciations (e.g., Streeb based on Street) and four were based on real words with irregular pronunciations (e.g., Subar based on Sugar).

The child's reading aloud attempt is scored as correct or not and they receive corresponding feedback via the computer, hearing the correct pronunciation. The children complete a maximum of 10 trials for each word, presented in 10 blocks. The test discontinues if a child reads all 8 words correctly in two consecutive blocks.

2. Spelling post test

The child hears each of the eight novel words spoken aloud and is asked to write down the correct spelling on a piece of paper. They score one point for each correctly spelled novel word.



3. Orthographic choice post test

For each of the eight novel words the child sees four word options presented on the screen. The target word (subar), a different but plausible spelling of the word (suber), and two visually similar made-up words (sular, suler). Their task is to point to the spelling of the word they learned earlier. They score one point for each correct response.

Dynamic assessment of sight word learning

Did the dynamic assessment of sight word learning correlate with SES and/or English language proficiency?

- Performance on the sight word training trials was not statistically significantly correlated with SES (*r*=.07), and was weakly correlated with English language proficiency (*r*=.32).
- Scores on the sight word learning post tests (orthographic choice and spelling combined) were weakly correlated with SES (*r*=.23) and moderately with English language proficiency (*r*=.50).
- Word reading (accuracy and fluency combined) was weakly correlated with SES (*r*=.23) and strongly correlated with English language proficiency (*r*=.60).
- Nonword reading (accuracy and fluency combined) was very weakly correlated with SES (*r*=.13) and moderately correlated with English language proficiency (*r*=.44).

Hypothesis 1 supported

Did the dynamic assessment of sight word learning predict growth in reading and reading related skills?

- The DA of sight word learning training and post test scores each predicted an additional 2% of variance in reading growth in the whole sample after the predictive value of the static tests had been accounted for.
- When the data from the 229 monolingual children who took part in this work package were analysed as a subgroup, the DA of sight word learning training and post test scores both continued to predict an additional 2% of unique variance in word reading growth.
- When the data from the 99 children with EAL were analysed as a subgroup, the DA of sight word learning training and post test scores both predicted additional unique variance but to a lesser extent (approximately 1%).

Hypothesis 2 supported

Can dynamic assessments of sight word learning accurately screen for reading difficulties?

- 50 children were identified as poor readers with a dyslexic profile, 27 were monolingual and 23 had EAL.
- From Figure 3 it can be seen that for the whole sample the dynamic assessment of sight word learning achieved acceptable levels of classification accuracy (as indicated by the AUC, .931), and the static tests of word reading as a composite achieved a slightly higher level (.958).
- When added to the static tests, the dynamic assessment of sight word learning did not improve classification accuracy so therefore these values are not presented in the figure. This is likely to be because the AUC values are already both very high.
- This pattern of findings was consistent across both monolingual and EAL subgroups as can be seen in Figure 4.

Hypothesis 3 supported

Dynamic assessment of sight word learning

	Sensitivity	Specificity	AUC
DA Training	.880	.806	.898
DA Post test	.940	.813	.931
Static Word Reading	.920	.942	.958

Whole sample

Figure 3 Sensitivity, specificity and AUC values from statistically significant predictors, obtained using DA sight word learning measures, static tests, and data from the whole sample (n=328).

Monolingual

	Sensitivity	Specificity	AUC
DA Training	.778	.881	.880
DA Post test	.891	.926	.927
Static Word Reading	.889	.970	.942

EAL

	Sensitivity	Specificity	AUC
DA Training	.870	.829	.924
DA Post test	.957	.750	.931
Static Word Reading	.957	.934	.976

Figure 4 Sensitivity, specificity and AUC values from statistically significant predictors, obtained using DA sight word learning measures, static tests, and data from the two subgroups (Monolingual n=229, EAL n=99).

Dynamic assessment of vocabulary learning

Aims

To create a computerised version of a Danish vocabulary learning task (Gellert & Elbro, 2013).

To trial the new dynamic assessment of vocabulary learning to determine its utility in predicting reading outcomes and identifying children at risk of reading difficulties, specifically the poor comprehender profile.

What did we expect to find?

1. The dynamic assessment of vocabulary learning will not correlate strongly with SES or English language proficiency at assessment time point one.

2. The dynamic assessment of vocabulary learning will predict growth in vocabulary and reading comprehension between time points one and two.

3. The dynamic assessment of vocabulary will achieve acceptable levels of screening accuracy (AUC 70 or above).

Study design and participants

Participants were first assessed when they were in Year 4. The second assessment time point was approximately 15 months later when they were in Year 6. The diagnostic criteria used to categorise the poor comprehender profile of reading difficulties at Time 2 was 1) Reading comprehension score 1 standard deviation or more below the average expected for chronological age 2) A discrepancy of 1 standard deviation or more between Reading Accuracy and Reading Comprehension, with performance in Reading Accuracy being stronger than Reading Comprehension. There is considerable debate regarding the specific criteria to use when diagnosing this profile. We have used a discrepancy based approach which is common in the literature (Spencer and Wagner, 2018) and added the more stringent requirement that reading comprehension is below the expected level for chronological age.

Time 1

May - July 2019

n=414

Male: 226 (55%)
Female: 188 (45%)

SEN: 58 (14%)
EAL: 145 (35%)

Average age: 9 years 2 months

15 months

Time 2

Nov - Dec 2020

n=320

Male: 173 (54%)
Female: 147 (46%)

SEN: 43 (14%)
EAL: 123 (39%)

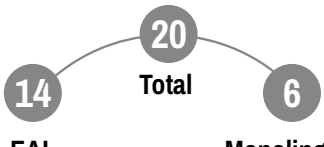
Average age: 10 years 8 months

Outcomes
Reading comprehension ability;
Receptive vocabulary.

Static predictors of reading comprehension
Nonverbal ability; Vocabulary;
Reading accuracy.

Dynamic assessment of vocabulary learning

Number of children at risk of poor comprehender reading profile



Attrition

22%

Moved school: 25
Absent: 11
Lockdown: 56
Other: 2

Outcomes
Reading comprehension ability;
Receptive vocabulary

Dynamic assessment of vocabulary learning

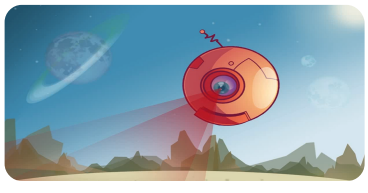
The story context

The goal of this assessment is to learn the names of six aliens and some semantic information about them.

The task is presented as a Galaxy Explorer Quest with Commander Stan Mackenzie, who is travelling to different planets meeting the aliens who live there. The researcher controls progress through the task presented on a computer, recording whether a child's response is correct. Everything else is controlled by the computer program. The time the task takes from start to finish is on average 25 minutes, with a range + or - 10 minutes, depending on the child.

Images and background illustrations are related to space travel and are designed to be motivating and rewarding.

The six aliens are divided into two blocks of three, for learning trials and post tests.



1. Initial exposure

Children hear the name of each alien and accompanying semantic information. For example, "Space Cadet, your task is to listen carefully to the name of each alien and say what you hear. Goni: a red, bearded, lazy alien."

They are asked to immediately repeat what they hear. Two of the three adjectives used to describe each alien can be seen in the picture (including one colour) and one is not visible (such as a personality attribute).

2. Vocabulary training

Next, images of the aliens are presented and the children are asked "What was the name of this alien?" One point is given for each correct name. Feedback is given in which the child hears the name of the alien and accompanying semantic information again via the computer. There is a maximum of 10 trials, but the task discontinues if the child is able to name all three aliens correctly in two consecutive blocks.

3. Definitions post test

Following vocabulary training, children are assessed on their knowledge of the semantic information about the aliens. They are asked "How would you describe Goni?" No visual information is provided. A maximum of three points per item is given for each of the correct attributes provided.

4. Recall post test

Next the child hears the semantic information about the alien and is asked to recall the name. For example, "What was the name of the red, bearded, lazy, alien?" No visual information is provided. They are awarded one point for each name they recall correctly.

5. Recognition post test

The child sees pictures of the three aliens from the training and three new aliens. They hear the alien names and are asked to point to the pictures of them. The pictures are randomly shuffled into different positions each time. They score one point per correct response.

Dynamic assessment of vocabulary learning

Did the dynamic assessment of vocabulary learning correlate with SES and/or English language proficiency?

- We conducted principal components analysis on the phases of the dynamic task: namely, vocabulary training, definition, and recall (recognition was excluded due to ceiling effects). We extracted two components, with vocabulary training and the recall post-test loading highly on a first 'phonological' factor, and the definitions post-test loading highly on a second 'semantic' factor.
- The dynamic assessment of vocabulary semantic scores were not statistically significantly correlated with SES (r=.09) but were moderately correlated with English language proficiency (r=.44). The dynamic assessment of vocabulary phonology scores were very weakly correlated with SES (r=.13) and weakly correlated with English language proficiency (r=.37).
- The static tests of reading were weakly correlated with SES (Reading accuracy r=.21; Reading comprehension r=.32), but moderately (Reading accuracy r=.49) and strongly (Reading comprehension r=.78) correlated with English language proficiency.

Hypothesis 1 supported

Did the dynamic assessment of vocabulary learning predict growth in vocabulary?

- The dynamic assessment of vocabulary semantic and phonology scores both predicted additional variance (2% and 4% respectively) in vocabulary growth in the whole sample after the predictive value of the static tests had been accounted for.
- When the data from the 197 monolingual children were analysed, the dynamic assessment of vocabulary scores continued to predict an additional unique variance in vocabulary growth (semantic 2%, phonology 3%).
- When the data from the 123 children with EAL were analysed, the dynamic assessment of vocabulary scores predicted additional unique variance to a greater extent (semantic 7%, phonology 9%).

Hypothesis 2 a supported

Did the dynamic assessment of vocabulary learning predict growth in reading comprehension?

- The dynamic assessment of vocabulary semantic post test scores predicted additional variance (<1%) in reading comprehension growth in the whole sample after the predictive value of all the static tests had been accounted for (SES, nonverbal ability, vocabulary knowledge, reading accuracy).
- When the data from the 197 monolingual children who took part in this work package were analysed as a subgroup, the dynamic assessment of vocabulary post test scores did not predict unique variance in reading comprehension growth.
- When the data from the 123 children with EAL were analysed as a subgroup, the dynamic assessment of vocabulary semantic scores predicted a small but significant amount of additional unique variance (1%) in reading comprehension growth, again after the predictive value of all the static tests had been accounted for.

Hypothesis 2 b supported

Can dynamic assessments of vocabulary learning accurately screen for reading difficulties?

- 20 children were identified as poor readers showing a poor comprehender profile, 14 of these were children with EAL and 6 were monolingual.
- From Figure 5 it can be seen that for the whole sample the dynamic assessment of vocabulary achieved a high level of classification accuracy (as indicated by the AUC, .864), and the static tests achieved a slightly higher level (.878). When added to the static tests, the DA of vocabulary did not improve classification accuracy.
- From Figure 6 it can be seen that for the 123 children with EAL, when the dynamic assessment of vocabulary phonology post test scores were added to the static tests they improved classification accuracy by a small but statistically significant amount (from .841 to .873). Together they were able to identify *all* 14 children with EAL who went on to have reading comprehension difficulties.

Hypothesis 3 supported

Dynamic assessment of vocabulary learning

	Sensitivity	Specificity	AUC
DA Phonology	.750	.820	.843
DA Semantic	.850	.727	.864
Static Reading Accuracy	1.000	.607	.862
Static Reading Comprehension	1.000	.623	.878

Whole sample

Figure 5 Sensitivity, specificity and AUC values from statistically significant predictors, obtained using DA vocabulary learning measures, static tests, and data from the whole sample (n=320).

	Monolingual			EAL		
	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
DA Phonology	*Not applicable as not significant*			.929	.615	.825
DA Semantic	*Not applicable as not significant*			*Not applicable as not significant*		
Static Reading Accuracy	.833	.864	.902	.714	.881	.841
Static Reading Comprehension	1.000	.770	.928	*Not applicable as not significant*		
Static + DA Phonology	*Not applicable as not significant*			1.000	.697	.873

Figure 6 Sensitivity, specificity and AUC values from statistically significant predictors, obtained using DA vocabulary learning measures, static tests, and both test types combined and data from the two subgroups (Monolingual n=197, EAL n=123).

Summary of findings

Key findings

- Performance on the dynamic assessments was, as hypothesised, only weakly to moderately correlated with socio-economic status. However, performance on the static tests was also only weakly to moderately correlated with socio-economic status, suggesting that the dynamic assessments did not offer any advantages over static tests with regards reducing socio-economic related bias.
- With regards bias associated with English language proficiency (specifically vocabulary knowledge) there was a noticeable difference between the dynamic assessments and static tests. Correlations between the dynamic assessments and vocabulary were typically weak to moderate, whereas those between static tests and vocabulary were moderate to strong. Given the growing population of primary school pupils who have EAL, the need to find assessments that are less biased with regards English language proficiency is an increasingly pressing issue.
- All of our dynamic assessments predicted unique growth in reading ability over periods of 10 to 15 months:
- The dynamic assessment of decoding predicted growth in early word reading.
- The dynamic assessment of sight word learning predicted growth in reading accuracy and fluency.
- The dynamic assessment of vocabulary learning predicted growth in reading comprehension.
- All of our dynamic assessments achieved excellent or outstanding levels of accuracy as screeners for later reading difficulties. Two showed potential to add value to static tests for children with EAL (decoding and vocabulary learning). The dynamic assessment of decoding has also shown potential for use with non-readers. This suggests that if we had been able to administer the decoding task earlier in the school year as we had planned (but which was delayed by the COVID-19 pandemic) we may have found that the task added to screening accuracy for the whole sample.

Other findings

- Our findings are consistent with evidence from previous research regarding static measures of reading-related skills as predictors of reading growth:

  1) They underscore the importance of letter knowledge and phonological awareness as key foundational skills for the development of early reading.

  2) They evidence the role of sight word learning in reading development, as the process of acquiring new orthographic representations was shown to predict growth in word reading accuracy and fluency.

  3) They provide further evidence of the important role of vocabulary learning in reading comprehension development.

The DART project work packages have successfully provided evidence of 'proof of concept'.

The dynamic assessments are practical to administer and provide unique information that is potentially useful for predicting progress in reading and identifying children at risk of reading difficulties.

Next steps

Approaches to screening

- We have shown that our dynamic assessments achieve excellent and outstanding levels of accuracy in identifying children at risk of reading difficulties when used alone. The dynamic assessments of decoding and vocabulary learning also have potential to add value when used in conjunction with static assessments. We now need to work in partnership with practitioners to establish how dynamic assessment may fit within existing assessment and monitoring practices in schools and identify which educators may be most likely to use them.
- We also need to pinpoint when in the screening process dynamic assessments might provide the most useful insights. Questions to consider include, should they be used before or after static tests? How often should they be repeated? What is a suitable interval between repeated assessments?
- The timing of the activities in the DART project were affected by the disruptions to schooling due to the COVID-19 pandemic, meaning that they may not all have been optimal. This was particularly true for the decoding task, which we had planned (based on pilot work) to administer after Christmas in the reception year. Our additional exploratory analyses have shown the dynamic assessment of decoding to have potential for use with children in the very early stages of learning to read. Specifically we found that for children who were non readers at the first time point, this dynamic assessment improved classification accuracy through increasing sensitivity. This means that it identified more of the children who had persistent difficulties. If we had been able to administered the decoding task earlier in the school year, as we had planned, we may have found that the task added to screening accuracy for the whole sample.
- Our data have shown the dynamic assessment of vocabulary learning to be an effective screening tool for children in Year 4 of primary school, however we believe it could be useful earlier in development to identify those at risk of reading comprehension difficulties sooner. With adaptation the learning task could be suitable for children as young as Year 2.
- In the DART project the dynamic assessments were all presented on a laptop and administration was guided by trained research assistants. More research is needed to determine the support and training that might be needed for educators to be able to use the assessments independently in schools, and how best to provide this.

Dynamic assessment design

- An additional cycle of design work is needed to refine the presentation and delivery of the dynamic assessments.
- The dynamic assessments as they are now can be quite time consuming with some children. We need to consider feasibility and acceptability within a school context, and explore ways to maximise the information that can be collected in a shorter period of time.
- Participatory co-design is needed with stakeholders (children and educators) to ensure that the dynamic assessments are as appealing and easy to use as possible.
- The dynamic assessments currently run on specialist software, we now need to create versions that can run on accessible, stable and low cost platforms suitable for school IT systems.
- Further consideration needs to be paid to the nature of the feedback provided, and the extent to which it needs to vary as the training unfolds. At present the feedback is a little repetitive and there is scope for improvement.
- Finally whilst the dynamic assessments have some elements that respond to the individual child depending upon the correct responses and errors that they make, we would like to explore ways to increase the extent to which they can be adaptive.

Acknowledgements

Research Assistants

We would like to thank the large team of research assistants who were involved in the three work packages. They made significant contributions to the data collection and data entry, visiting schools to work with the school staff and children. Their hardwork and commitment to the project is very much appreciated.

Schools

This research was made possible by the fantastic group of schools who supported the project during the unpredictable and challenging circumstances of the COVID-19 pandemic. The schools were incredibly welcoming, flexible and accommodating and we sincerely thank them for all of the help and support they provided.

Children

In total, 1118 children took part in the DART project of varying ages and at various points in their primary school journey. Their enthusiasm and energy was wonderful, and we thank them for all their hard work. It was an absolute privilege to meet such a brilliant group of children and we wish them all the best for the future.

The Nuffield Foundation

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-founder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org.

For more information about the project please contact Dr Hannah Nash via email h.nash@leeds.ac.uk

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Bailey, C. E., Manis, F. R., Pedersen, W. C., & Seidenberg, M. S. (2004). Variation among developmental dyslexics: Evidence from a printed-word-learning task. Journal of Experimental Child Psychology, 87(2), 125-154.

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. Psychological science, 24(8), 1398-1407.

Castles, A. & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), From inkmarks to ideas: Challenges and controversies about word recognition and reading. London: Psychology Press.

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. Journal of learning disabilities, 42(2), 163-176.

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.

Demie, F. (2018). English language proficiency and attainment of EAL (English as second language) pupils in England. Journal of Multilingual and Multicultural Development, 39(7), 641-653.

Department for Education (2022). Statistics: Schools, pupils and their characteristics.

Department for Education (2022). National Curriculum Assessments: Key Stage 2, October 2022.

Dixon, C., Oxley, E., Gellert, A. S., & Nash, H. (2022a). Dynamic assessment as a predictor of reading development: a systematic review. Reading and Writing, 1-26.

Dixon, C., Oxley, E., Nash, H., & Gellert, A. S. (2022b). Does dynamic assessment offer an alternative approach to identifying reading disorder? A systematic review. Journal of Learning Disabilities.

Dunn L, Dunn D, Styles B (2009). The British picture vocabulary scale. 3rd ed. GL Assessment

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior research methods, instruments, & computers, 35(1), 116-124.

Forum for Research in Literacy and Language, Institute of Education. (2012). The Diagnostic Test of Word Reading Processes. Swindon, UK: G L Assessment.

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.

Gellert, A. S., & Elbro, C. (2017). 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.

Gibbs, S., & Bodman, S. (2014). Phonological Assessment Battery Primary: PhAB2 Primary. GL Assessment Limited.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.

Hulme, C., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M.J. (2009). YARC York Assessment of Reading for Comprehension: Early Reading. London: GL Assessment.

References

Hutchinson, J. (2018). Educational outcomes of children with English as an additional language.

Li, M., Geva, E., D'Angelo, N.et al. Exploring sources of poor reading comprehension in English language learners. Ann. of Dyslexia71, 299–321 (2021).

Keenan, J. M., Hua, A. N., Meenan, C. E., Pennington, B. F., Willcutt, E., & Olson, R. K. (2014). Issues in identifying poor comprehenders. LAnnee psychologique, 114(4), 753-777.

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. Psychological Bulletin, 138(2), 322.

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.

Nash, H. M., Davies, R., & Ricketts, J. (2021). The contributions of decoding skill and lexical knowledge to the development of irregular word reading. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of Neuroscience Methods, 162(1-2), 8-13.

Peña, E. D., & Halle, T. G. (2011). Assessing preschool dual language learners: Traveling a multiforked road. Child Development Perspectives, 5(1), 28-32.

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., ... & Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. Journal of Abnormal Psychology, 121(1), 212.

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.

Raven, J.C., Court, J.H., & Raven, J. (2008). UK Norms for Coloured Progressive Matrices, Standard Progressive Matrices and the Crichton and Mill Hill Vocabulary test. Pearson.

Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., ... & Carroll, J. (2004). Sex differences in developmental reading disability: new findings from 4 epidemiological studies. Jama, 291(16), 2007-2012.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. Cognition, 55(2), 151-218.

Snowling, M. J., Stothard, S. E., Clarke, P.,Bowyer-Crane, C., Harrington, A., Truelove, E. &Hulme, C. (2009)YARC York Assessment of Reading for Comprehension. London: GL Assessment.

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.

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.

Torgesen J. K., Wagner R. K., Rashotte C. A. (2012). Test of Word Reading Efficiency–Second Edition. Austin, TX: Pro-Ed.

Wang, H. C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. Journal of Experimental Child Psychology, 109(1), 39-57.