Received: 30 June 2021 | Accepted: 13 June 2022

DOI: 10.1002/berj.3821

BERJ BERA

ORIGINAL ARTICLE

# Early detection of risk of reading difficulties using a working memory assessment battery

**Susan J. Atkinson**[1] | **Colin R. Martin**[2]

[1]Carnegie School of Education, Leeds Beckett University, Leeds, UK

[2]Institute of Health and Wellbeing, University of Suffolk, Ipswich, UK

**Correspondence**
Dr Susan J. Atkinson, Carnegie School of Education, Carnegie Hall, Leeds Beckett University, Church Wood Avenue, Leeds, UK, LS6 3QS.
Email: s.j.atkinson@leedsbeckett.ac.uk

**Funding information**
The corresponding author carried out data collection as part of a funded PhD studentship from the University of Central Lancashire.

## Abstract

Research suggests a role for aspects of the working memory system in reading. While much of the evidence points to a role for working memory capacity and the phonological loop, more recent work indicates a role for the central executive component, although findings remain unclear. There is an identified need for a reliable screening measure for risk of reading difficulties in children who are pre-readers. Recent research suggests that working memory measures may contribute to such a measure in addition to existing tests of phonological ability, nonverbal reasoning, motor skills and language. One hundred and two children aged between 56 and 69 months were assessed on measures of reading accuracy, working memory capacity, processing speed and a range of measures designed to assess central executive functioning. Linear regression, discriminant function analysis and confirmatory factor analysis were conducted to determine the predictive qualities and factor structure of the working memory assessment battery. Discriminant function analysis indicated that the working memory assessment battery was able to significantly discriminate between children who were at risk and not at risk of reading difficulties (as indicated by an independent measure).

Linear regression and Confirmatory Factor Analysis revealed that the tests were a good predictor of the later outcome, particularly the alliteration fluency and sentence verification tasks. Therefore, the working memory assessment battery promises to be a useful screening measure for potential reading difficulties in educational settings, facilitating early effective intervention.

**KEYWORDS**
assessment, early years, reading difficulties, working memory

**Key insights**

**What is the main issue that the paper addresses?**
Working memory is of increasing interest to teachers as a component of cognitive functioning which impacts on early reading development.

**What are the main insights that the paper provides?**
This paper argues that a battery of easily administered working memory tests can indicate children at risk of both working memory and reading difficulties.

## INTRODUCTION

The early detection of potential reading difficulties in young pre-readers remains an area of intense educational and clinical concern (Mathes & Denton, 2002). Contemporary assessment measures emphasise phonological awareness skills (Muter et al., 1996) or a range of skills associated with dyslexia such as motor skills and letter identification (Fawcett et al., 1998; Nicolson & Fawcett, 1996). However, these tests are not without their critics; for example, Simpson and Everatt (2005) argue that testing letter knowledge, sound order and rapid automatised naming predict later literacy skills more accurately than combining measures or using phonological tests alone.

There is compelling evidence for the role of working memory, the cognitive skills needed for storing, rehearsing and manipulating information (Baddeley & Hitch, 1974; Just & Carpenter, 1992), in the development of reading, and for a deficit, particularly in central executive functioning (a key component of working memory), in dyslexia (Palmer, 2000a; Pickering, 2004; Reiter et al., 2004; Smith-Spark & Fisk, 2007). Early intervention for those at risk of reading difficulties is important for facilitating engagement with the full school curriculum and preventing the educational underachievement that is likely to accompany reading difficulties (Cunningham & Stanovich, 1991; Stanovich, 1986). It also has a significant therapeutic benefit in enhancing self-esteem and preventing feelings of failure (McNulty, 2003). Therefore, the identification of early effective predictors of reading difficulties is of considerable importance. The assessment of components of working memory then may be useful in identifying early on which children may be at risk of reading difficulties as a result of poor working memory capacity or functioning.

Baddeley and Hitch's (1974) model of working memory continues to be influential. According to the model, working memory underpins many complex cognitive tasks and allows us to store and manipulate information for short time durations (Baddeley, 2017; Baddeley et al., 2021). Working memory is a multi-component model, with the central executive (CE) considered to be the most important component. The CE is modality free; it controls and coordinates the phonological loop and the visuo-spatial sketchpad. These slave systems store and rehearse short-term verbal and visual/spatial information in separate domain-specific components. The CE enables the performance of complex span tasks involving both the maintenance and the manipulation of information from short-term and long-term memory (Unsworth & Engle, 2007). The functions of the CE have been identified as comprising focusing, dividing and switching attention, inhibiting information and stimuli that are not relevant to a task, controlling and coordinating the subsystems of working memory and accessing long-term memory (Baddeley, 1996; Smith & Jonides, 1999). There is discussion in the literature about the precise role of the CE, but there is evidence for attention allocation and inhibitory processes. The episodic buffer has since been added to the model (Baddeley, 2000). It is argued that the episodic buffer is responsible for integrating information from long-term memory with information from the subcomponents of WM and combining information from different modalities. The components of the working memory system have been shown to be present from an early age (Alloway et al., 2004; Gathercole et al., 2004), although capacity, as evidenced by immediate serial recall, increases with age throughout childhood (Hulme et al., 1984; Nicolson, 1981), owing to increasing processing speed (Barrouillet et al., 2009; Bayliss et al., 2005) or changes in rehearsal strategy (Hulme et al., 1984). Capacity also shows wide individual differences throughout childhood which are persistent (Alloway, 2006).

Working memory predicts academic success throughout the school years (Alloway et al., 2014; Alloway & Alloway, 2010; Gathercole, 1999; Gathercole & Alloway, 2008; Gathercole & Pickering, 2001; Jarvis & Gathercole, 2003; Pham & Hasson, 2014). There is robust evidence that weaker phonological working memory skills are associated with poor vocabulary acquisition (Gathercole & Baddeley, 1989, 1990; Service, 1992) and developmental dyslexia (Beneventi et al., 2010; Smith-Spark et al., 2003; Smith-Spark & Fisk, 2007; Snowling & Hulme, 1989). There is also an increasing body of research linking CE performance in particular to, for example, mathematics performance (Bull et al., 2008; Holmes & Adams, 2006; Mammarella et al., 2018), reading skill (de Jong, 2006) and dyslexia (Pickering, 2004; Smith-Spark & Fisk, 2007). Palmer (2000a, 2000b) argues for a deficit specifically in the inhibition function of the CE rather than a general CE deficit. To read efficiently and fluently, cognitive resources such as attention have to be focused on recoding printed visual stimuli into phonological representations and matching them against activated representations from long-term memory whilst inhibiting the dominant competing or interfering visual code. Ineffective inhibition of competing activated visual representations when reading may lead to working memory becoming overloaded. As a result, children with underdeveloped inhibition skills may fail to automatise word recognition skills so that reading continues to be a conscious and effortful task. Children with poor inhibition skills at the age of 7/8 years show more positive indicators for dyslexia (Palmer, 2000a), and this deficit persists in teenagers with dyslexia (Palmer, 2000b). More recently, Booth and Boyle (2009) investigated the role of executive functions in reading and found that non-verbal measures of inhibition, but not planning, predict reading skill in boys aged between 9 and 11 years. However, it is not clear which tests of executive function are the most reliable and discriminating. Booth et al. (2010) conducted a meta-analysis of studies on the association between reading difficulties and executive impairments, concluding that there is an association, but effect sizes vary between studies depending on the assessment tasks used. They state that studies show that inhibition and working memory are involved in reading difficulties; children with reading difficulties score significantly lower on the Working Memory Test Battery-Children (Pickering & Gathercole, 2001) CE measures of backward digit recall, counting recall and listening recall than typically developing children.

The incidence of poor working memory skills is thought to be around 10% (Alloway et al., 2009), therefore affecting three or four children in each primary class in the UK, although some research suggests that the incidence could be as high as 35% (Atkinson & Mitchell, 2015), particularly at the moment when children have been affected by the Covid-19 pandemic and concomitant stress and anxiety, which affect working memory performance (Klein & Boals, 2001; Lavigne-Cerván et al., 2021). Children with weaker working memory skills find many classroom activities (such as following instructions, keeping place or remembering sequences) difficult owing to the task working memory load (Gathercole et al., 2006). Pupils with working memory problems tend to be identified by teachers as having attentional or behavioural difficulties (Alloway et al., 2009), which are more likely to be attributed to personality factors than poor working memory (Alloway et al., 2012). However, more recent research suggests that awareness of working memory is increasing amongst education practitioners (Atkinson et al., 2021), although there is wide variability in the accuracy of knowledge and strategies to support pupils. Working memory is now included in the Department for Education Core Content for initial teacher training courses (DfE, 2019) in Teacher Standard 2 ‘How Pupils Learn’, but it is not explicitly referred to in Teacher Standard 5, ‘Adaptive Teaching’. This suggests that a knowledge of working memory is useful for teachers and school staff, and that a time- and cost-efficient method for assessing components would be extremely useful.

As awareness and knowledge of the importance of working memory increase in schools, there will be a growing need for measures to quickly and easily identify children who are struggling or at risk of working memory difficulties so that schools can intervene early to support them. Normand and Tannock (2014) argue that there is a lack of effective and easy to administer tools. Schools tend to be time and resource poor: accessing commercial screening, diagnostic and intervention measures is often beyond their resources in cost and time, in terms of training staff, carrying out tests and finding time within the school day for interventions. The Working Memory Rating Scale (Alloway et al., 2008) is intended for use by teachers to screen children for behavioural difficulties associated with poor working memory. The Scale is a 20-item questionnaire where each behaviour (for example, ‘abandons activities before completion’, ‘does not follow classroom instructions accurately’, ‘depends on neighbour to remind them of the current task’) is rated from not at all typical of that child to very typical. The Scale was normed with the Automated Working Memory Assessment (Alloway, 2007), a computerised assessment comprising tests of four memory components: verbal Short-term memory (STM), Verbal CE, visuo-spatial STM and visuo-spatial CE. This test and the earlier Working Memory Test Battery for Children (Pickering & Gathercole, 2001), a non-computerised assessment, are no longer available. While being quick and easy to administer, the Working Memory Rating Scale indicates the severity of a child's difficulties overall but does not indicate which specific components of WM are stronger or weaker.

Given the contribution of aspects of working memory to reading competence (Palmer, 2000a, 2000b) and the problems with the established methods of assessment of risk for reading difficulties (Simpson & Everatt, 2005), the development of an accurate risk identification tool based on working memory is both timely and highly desirable, and is the focus of the study reported here.

An assessment battery was developed comprising measures of reading ability, phonological memory capacity, processing speed, and components of working memory functioning. Working memory measures were selected to address both fluent retrieval from long-term memory and inhibition (holding and manipulating information while suppressing the prepotent response), both of which have been identified as important in the development of fluent, automatic reading skills (Palmer, 2000a, 2000b, 2000c). This set of tests has advantages over other measures of working memory in that it ensures that any differences in central executive functioning are not due to differences in the amount of information that can be remembered or the speed at which it can be processed, as hypothesised by phonological processing deficit or double-deficit models of dyslexia (Denckla & Cutting, 1999; Snowling, 1995). There is increasing evidence that the central executive plays a greater role in dyslexia than these phonological processing deficit models predict (Palmer, 2000a, 2000b, 2000c), and therefore inclusion of tests in the battery which establish the role of the central executive is crucial.

The purpose of the present study was to determine if the devised test battery was effective in predicting reading ability, a key attribute in risk identification.

# METHOD

## Design

The study used a mixed group design. The children were allocated to one of three age-matched groups in their Reception year (aged 4–5 years) following screening using the Dyslexia Early Screening Test (DEST; Nicolson & Fawcett, 1996) and Raven's Coloured Progressive Matrices (Raven, 1995). This test (Nicolson & Fawcett, 1996) is designed to identify children aged between 4 years 6 months and 6 years 5 months at risk of developing reading difficulties. It consists of 10 sub-tests administered according to the manual instructions. The sub-tests assess knowledge (letter and digit naming), motor skills (bead threading, shape copying, postural stability), phonological awareness (phonological discrimination, rhyme detection, sound order discrimination, rapid naming) and memory capacity (forward digit span). Scores for each sub-test are compared with standardised norms for different age groups which, when totalled, give an overall ‘at risk’ quotient (ARQ). An ARQ ≥0.6 is considered as indicative of borderline risk of dyslexia (Nicolson et al., 1999).

Three groups were identified: (1) ‘at risk of reading difficulties’; (2) an ‘intermediate group’;[1] and (3) ‘not at risk of reading difficulties’. Further assessment using the study test battery (detailed below) was carried out in school years 1, 2 and 3 (ages 5–8 years).

## Participants

One hundred and two (54 female) children (mean age = 62.25 months, SD = 3.66, range = 56–69 months) participated in the study. All the children attended mainstream schools and were not identified as having special educational needs by the schools at the beginning of the study. The study had ethical approval from the relevant university faculty ethics committee, and parents of the children involved gave their informed consent.

## Measures and procedure

The study test battery comprised the following tests:

1. *Reading* – the Wide Range Achievement Test (Jastak & Wilkinson, 1984), measuring single letter and word identification. This was administered according to the manual instructions. The children were first asked to identify 15 letters of the alphabet by sound or letter name, then to attempt to read a list of 42 words increasing in difficulty. Testing stopped after 10 consecutive errors or omissions. The score used was the total number of letters and words read correctly.

2. *Working memory capacity* – forward digit span where children repeat back strings of numbers presented at the rate of 1 digit per second. Testing began with a list length of two digits with two trials at each increasing length and stopped when the child made errors in both trials at one length. The score used was the child's span: the number of digits that can be repeated back without error.
3. *Processing speed* – The Space Ships sub-test from the Test of Everyday Attention for Children (Manly et al., 1999) was used, where matching pairs of spaceships have to be marked as quickly as possible on an A3 array of distractor pairs. The score for this task was the time taken in seconds divided by the number of correctly marked pairs, controlling for motor speed.

## Central executive measures

1. *Fluency* – rapid naming (naming pictures of single-syllable line drawings on an A4 array as quickly as possible; Fawcett & Nicolson, 2004), semantic fluency (naming as many category exemplars as possible in 1 min, e.g. 'animals'; Frederickson et al., 1997) and alliteration fluency (naming items beginning with a particular sound in 1 min, e.g. /m/; Frederickson et al., 1997). These tasks measure the ability of the participant to retrieve information from long-term memory and to inhibit responses that are incorrect or have already been given (Henry, 2012). Baddeley (1996) argues that accessing and retrieving relevant information from long-term memory is one of the roles of the CE.
2. *Inhibitory processes* – these tasks require the manipulation and recall of information together with the inhibition of conflicting or incorrect information. Tasks used were reverse digit span (repeating back strings of digits in reverse order; Fawcett & Nicolson, 2004); the sentence verification task (identifying sentences as true or false, then recalling the final words in each block of sentences, e.g. 'grass grows in the house' *false … house*. Henry, 2001); Odd One Out (identifying the odd one out in a series of three shapes, then recalling its position, with trial length increasing (Henry, 2001), which is often seen as a measure of spatial CE functioning, but is included here as a measure of CE inhibition because it mimics the sentence verification task in format – recalling an increasing list of shapes and their position requiring the inhibition of incorrect spatial information – but with less reliance on verbal or linguistic skill); and the shape sorting task, where plastic shapes are sorted first on the basis of shape, then on colour, based on a task developed by Carlson and Moses (2001).

The forward digit span task and Rapid Naming task were included in the assessment battery as well as in the initial screening. These tests are useful and reliable indicators of phonological loop capacity and long-term memory retrieval respectively. The rapid naming task was identified by Simpson and Everatt (2005) as a more useful predictor of dyslexia than other sub-tests of the DEST. Both tests were scored differently in the DEST: there, the child's score is compared with standardised norms for their age group, giving an 'at risk index' score. The sub-test 'at risk' scores are then totalled to give an overall 'at-risk' quotient. In addition, if these tasks were removed from the initial screening, only two children from the whole sample would move groups. Neither of the tasks was a significant predictor of later at risk status at the end of the study.

The assessments were administered to the children individually in a quiet area of the school classroom. Location is important as performance in the classroom is likely to be worse than in a quiet room free from distractions and therefore more akin to daily behaviour (Friso-van den Bos & van de Weijer-Bergsma, 2020). Presentation was counterbalanced in two sessions lasting approximately 15 min each delivered on consecutive days to reduce fatigue.

## Data analysis

The DEST sub-test scores for each sub-test for each child were compared with standardised norms for the relevant age group to give the ARQ. Children were allocated to one of the three groups as indicated in the design section above. One-way between-groups ANOVAs were conducted for each sub-test to establish whether the groups differed on the sub-tests.

A discriminant function analysis was conducted to determine the efficacy of the battery test set at the first observation point in discriminating reading ability group membership at the third observation point.

A linear regression was conducted to determine the efficacy of the battery elements at the first observation point in predicting reading ability group membership at the third observation point.

Confirmatory factor analysis was conducted to determine the most parsimonious structural model of the test battery based on contemporary models of central executive function. Three models were tested. These were: (1) a uni-dimensional model of central executive function; and (2) two-factor correlated and (3) two-factor uncorrelated models based on factors of (1) inhibition and (2) long-term memory retrieval. Multiple goodness of fit tests (Akaike, 1987; Bentler, 1990; Bentler & Bonett, 1980; Bozdogan, 1987; Browne & Cudeck, 1993; Carmines & Mciver, 1981; Dunbar et al., 2000; Hu & Bentler, 1995; Joreskog & Sorbom, 1993; Kline, 1988; Marsh et al., 1988) were used to evaluate the two models. The characteristics of the models tested are shown in Figure 1.

# RESULTS

## Characteristics of the population

For this study, the children were allocated to one of three groups according to their DEST 'at-risk' quotient (ARQ): 'at risk' of reading difficulties group, ARQ ≥0.6; 'intermediate' group, ARQ between 0.3 and 0.5; and 'not at risk' group, ARQ ≤0.2. A series of one-way between-groups ANOVAs were conducted to see whether there were significant differences between the groups for each DEST sub-test. Using a Bonferroni adjusted alpha level of 0.01 for multiple comparisons, all of the tests reached significance. *Post-hoc* comparisons using Tukey's highest significant difference test indicated that the 'at risk' group performed significantly worse than the other groups on all sub-tests.

There were 29 (female = 10) participants in the 'at risk' group, 40 (female = 24) in the 'intermediate' group and 33 (female = 20) in the 'not at risk' group. No significant difference was observed as a function of gender and group type, $\chi^2_{(\text{d.f.}=2)} = 5.54$, $p = 0.06$. A 2 ×3 analysis of variance of age data revealed no main effect of either group type, $F_{(2,96)} = 0.05$, $p = 0.95$, or gender, $F_{(1,96)} = 0.37$, $p = 0.55$. No interaction was observed between group type and gender, $F_{(2,96)} = 0.18$, $p = 0.83$. The mean score on each test battery element at each observation point is shown in Table 1.

## Discriminant function analysis

Disciminant function analysis revealed the test battery at first observation to comprise a test set that significantly discriminates between group types at the third observation point (all tests, $F_{(2,99)} = 4.50$–$20.44$, $p = 0.01$–$0.001$). The test battery successfully predicted 'at

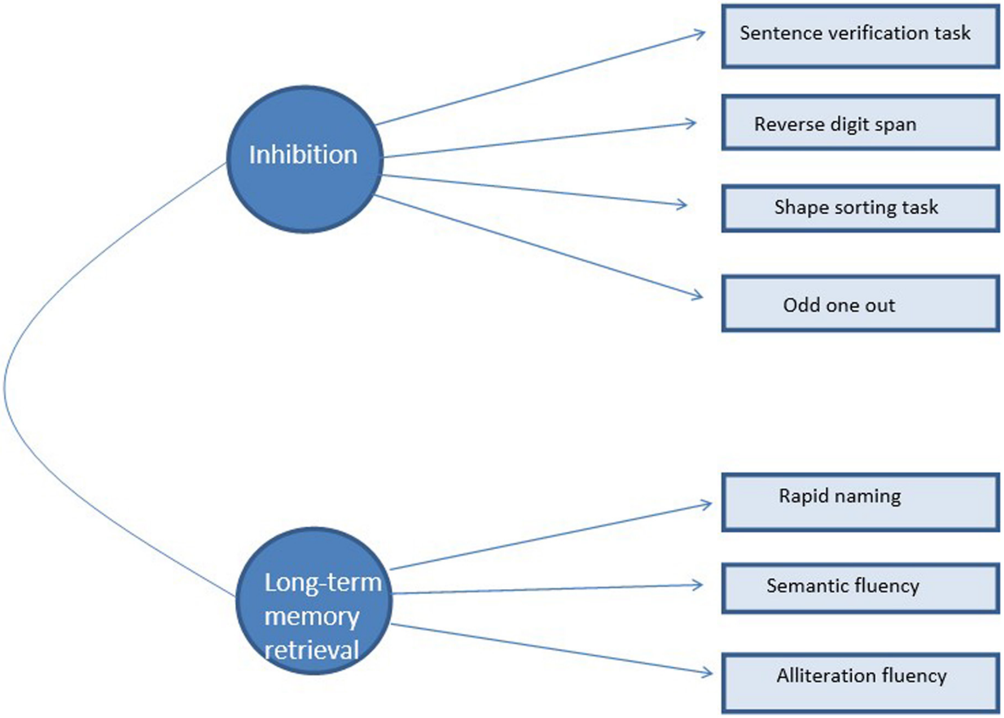

**FIGURE 1** Diagrammatic representation of the factor models evaluated
*Note*: The single factor model is obtained by setting the correlation between factors Inhibition and Long-term memory retrieval to 1. The two-factor uncorrelated model is obtained by setting the correlation between factors Inhibition and Long-term memory retrieval to 0. Finally, the two factor model is obtained by allowing the correlation between factors Inhibition and long-term memory retrieval to be freely estimated

risk' group membership in 64% of actual cases and in the 'not at risk' group in 86% of actual cases.

## Linear regression

Linear regression revealed that the test battery measures predicted approximately 34% of the variance in group type stratification ($R = 0.62$, $R^2 = 0.39$, adjusted $R^2 = 0.34$). Despite the modest proportion of variance explained, the regression analysis of variance (ANOVA) revealed that the model performed well, $F_{(7,94)} = 8.44$, $p < 0.001$. Examination of individual elements of the test battery revealed that the sentence verification test (standardised $\beta = 0.39$, $t = 3.40$, $p = 0.001$) and the alliteration fluency test (standardised $\beta = 0.19$, $t = 2.21$, $p < 0.05$) were the most important predictive elements in the test battery with all other tests non-significant ($p > 0.05$). *A posteriori* linear regression with the sentence verification test and the alliteration fluency test only included in the model revealed similar findings to the whole battery observations with approximately 35% of the variance in group type stratification ($R = 0.60$, $R^2 = 0.36$, adjusted $R^2 = 0.35$). Consistent with this observation, the regression ANOVA found this model to perform well, $F_{(2,99)} = 28.99$, $p < 0.001$. Similarly, the sentence verification test (standardised $\beta = 0.50$, $t = 5.93$, $p < 0.001$) and the alliteration fluency test (standardised $\beta = 0.21$, $t = 2.50$, $p = 0.01$) retained status as significant predictive elements.

**TABLE 1** Means (and standard deviations) for the measures used for each group at each observation point

| | | At risk | Intermediate | Not at risk |
|---|---|---|---|---|
| Age (months) | | 62.83 (3.58) | 62.39 (3.43) | 62.03 (4.15) |
| Raven's | | 13.79 (3.24) | 15.78 (3.82) | 16.39 (3.14) |
| Reading | Y1 | 11.59 (5.68) | 17.80 (4.41) | 20.12 (4.11) |
| | Y2 | 19.59 (4.21) | 25.47 (4.63) | 30.12 (4.08) |
| | Y3 | 25.76 (4.49) | 30.15 (4.47) | 34.42 (5.34) |
| Forward digit span | Y1 | 4.41 (0.98) | 4.95 (0.85) | 5.30 (0.92) |
| | Y2 | 5.03 (0.87) | 5.38 (0.95) | 6.06 (1.06) |
| | Y3 | 5.17 (0.80) | 5.50 (0.75) | 6.00 (0.90) |
| Processing speed | Y1 | 17.07 (16.43) | 9.96 (6.30) | 8.50 (4.58) |
| | Y2 | 8.88 (4.32) | 7.67 (3.31) | 6.85 (2.66) |
| | Y3 | 6.78 (2.67) | 6.25 (3.47) | 5.28 (1.60) |
| Rapid Naming | Y1 | 66.63 (22.79) | 52.66 (11.99) | 51.97 (13.71) |
| | Y2 | 55.66 (14.26) | 47.13 (9.60) | 47.79 (8.85) |
| | Y3 | 47.14 (7.61) | 41.80 (7.12) | 41.33 (6.04) |
| Semantic fluency | Y1 | 11.87 (3.10) | 14.12 (4.78) | 15.38 (3.48) |
| | Y2 | 15.28 (4.51) | 17.27 (4.76) | 16.94 (4.76) |
| | Y3 | 16.72 (4.40) | 19.05 (4.98) | 19.52 (5.10) |
| Alliteration fluency | Y1 | 4.85 (3.49) | 6.71 (3.39) | 8.35 (3.29) |
| | Y2 | 7.41 (3.73) | 9.07 (3.14) | 11.09 (2.87) |
| | Y3 | 8.86 (4.08) | 9.30 (4.80) | 10.61 (3.51) |
| Reverse digit span | Y1 | 1.07 (1.16) | 1.90 (0.95) | 2.44 (0.95) |
| | Y2 | 2.00 (1.26) | 3.03 (0.62) | 3.27 (0.90) |
| | Y3 | 2.72 (0.88) | 2.93 (0.62) | 3.52 (1.00) |
| Sentence verification task | Y1 | 0.21 (0.45) | 0.99 (0.73) | 1.58 (0.93) |
| | Y2 | 1.23 (0.80) | 1.70 (0.58) | 2.00 (0.55) |
| | Y3 | 1.52 (0.65) | 2.11 (0.65) | 2.23 (0.55) |
| Odd one out task | Y1 | 0.47 (0.67) | 1.21 (0.99) | 1.31 (1.15) |
| | Y2 | 1.32 (1.32) | 1.96 (0.96) | 2.32 (1.27) |
| | Y3 | 1.98 (1.32) | 2.80 (1.19) | 3.21 (1.27) |
| Shape sorting task | Y1 | 2.31 (2.74) | 2.73 (3.20) | 4.44 (3.08) |
| | Y2 | 3.03 (2.71) | 4.07 (2.90) | 5.89 (2.80) |
| | Y3 | 3.10 (2.74) | 4.63 (2.96) | 5.70 (2.87) |

## Confirmatory factor analysis

The factor models tested and accompanying fit indices are shown in Table 2. The $\chi^2$ goodness of fit analyses for all models were highly statistically significant ($p < 0.01$) for the single-factor model and uncorrelated two-factor model, indicating that a significant proportion of the total variance was unexplained by these models. Examination of the fit indices of the two-factor correlated model revealed a very good fit to the data with little residual variance unexplained.

**TABLE 2** Model evaluation of the test battery by comparison of fit indices

| Model | $\chi^2$ | d.f. | *p*-Value | CFI | TLI | RMSEA |
|---|---|---|---|---|---|---|
| Single factor | 46.85 | 14 | <0.001 | 0.76 | 0.65 | 0.15 |
| Two factor uncorrelated | 43.03 | 14 | <0.001 | 0.79 | 0.69 | 0.14 |
| Two factor correlated | 16.99 | 13 | **0.20** | **0.97** | **0.95** | **0.05** |

*Note*: Bold indicates best model fit as a function of model fit index criterion. Abbreviations: CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean squared error of approximation.

## DISCUSSION

This investigation reveals that a simply administered working memory test battery has great promise as a risk identifier for later reading difficulty. The results showed that the full test battery performed effectively in predicting membership of both the ‘at risk’ and ‘not at risk’ groups. Whilst the model with all of the test battery measures was significant, of the individual measures, only the sentence verification task and the alliteration fluency task were significant predictors of reading difficulty. This suggests that these two measures could be of use in identifying weaker CE skills and therefore potential reading difficulties.

These findings are of potentially profound importance for a variety of reasons. Firstly, early identification of risk facilitates the option of opportunistic intervention and monitoring to enhance the reading capacity and skill set of those at risk. Secondly, the ease of administration of the test battery affords an opportunity for endemic testing to not only systemically identify those at risk, but also gain additional insights and surveillance data regarding the epidemiology of later reading difficulty. This approach may thus also offer a cost-effective way to assess for later reading difficulty. Finally, given the potential stigma that may be associated with reading difficulty, above and beyond that of delays in reading milestones, this approach fosters the option to positively affect change and consequently reduce stigma in those at risk. The SEND Code of Practice (Department for Education/Department of Health, 2015) emphasises the importance of early identification of literacy difficulties, and these results suggest that this battery of working memory tests could contribute to the reliability of screening measures in a form which is easy for school teaching staff to administer. This was considered particularly important in the light of changes to the core content for initial teacher training as outlined in the introduction and the lack of easy to administer tests which identify skill in different components of working memory.

An interesting counter-intuitive finding was the observation of a well performing regression model but with a modest amount of variance explained within the model. One explanation for this concerns the limitations of a linear regression analysis on groups stratified into only three groups, although it should be emphasised that these need to satisfy the criteria of a linear relationship for the regression analysis to be appropriate. However, the ordered categorical aspect of the relatively small number of groups would be likely to result in a reduction of predictive discriminability inherent within the model evaluated; consequently less variance would be explained, although the model is acceptable. One perspective may be to examine reading scores purely along a continuum without group differentiation; however, although this has some statistical appeal, it unfortunately lacks ‘real world’ application since the reading veracity of a child will entail them being classified into one of a very small number of groups such as those operationalised in the current investigation. Reassurance over the accuracy of the test battery can be found by the impressive findings of the discriminant function analysis. These findings reveal impressive group discriminability inherent in the test battery, particularly given the extensive cognitive maturity occurring between the point of initial testing and the final group classification.

Scrutiny of the regression models also revealed that two tests performed particularly well at predicting group classification, these being the sentence verification test and the alliteration fluency test. The similarity of findings between the primary model comprising the whole test battery and the *a posteriori* model comprising just these two tests reveals the important contribution of these individual tests to the battery as a whole. A prima facie argument might be to simply utilise these two tests only as a risk detector for reading difficulty. However, there are arguments for retaining the whole test battery. The alliteration fluency test is usually assumed to be a measure of access to phonological information stored in long-term memory (Frith et al., 1995), and as such, fits together with the rapid naming and semantic fluency tests, both of which involve access to long-term memory. Success at the semantic fluency task, accessing members of a group from long-term memory, would support a diagnosis of a specific reading difficulty, whereas difficulties with it might suggest an alternative developmental disorder such as an Speech and Language Impairment (Fawcett & Nicolson, 2004). Those children who experience difficulties with both rapid naming and the alliteration fluency test fit the pattern of those with more severe indicators of dyslexia (Denckla & Cutting, 1999). Similarly, the sentence verification test, reverse digit span, odd one out and shape sorting tasks all measure the ability to hold and manipulate information, and in particular inhibit interfering activations or stimuli, but in different modalities or forms: words, digits or shapes. For these reasons, the whole battery of tests may be of more benefit in reaching an informed decision on a child's performance. Further support for this position can be found in the results of the Confirmatory Factor Analysis (CFA). The two-factor correlated model of the battery proved to be a very good fit to the data with little unexplained variance. The sentence verification test and the alliteration fluency test relate to different domains within this optimal two-factor correlated model. This suggests that these measures, although the best performing in the predictive analysis, are also couched within definable but related domains of cognitive performance and as such, are implicit to those domains. Since these domains are theoretically derived, and in light of structural support for the two-factor correlated model, there is considerable psychometric appeal in using the whole test battery. The findings from the CFA also suggest that consideration might be given to the formulation of domain scores as well as individual test result reporting.

Although the results indicate that the working memory test battery could profitably be used to identify children at the pre-reading stage who are at risk of reading difficulties, further work is needed to establish their reliability on a wider population sample. It would also be necessary to establish norms and standardised scores for each test and for the battery as a whole to enable it to be used confidently as a screening tool in schools and other educational establishments. In addition, the tests included in this battery are biased towards the verbal or linguistic. The Automated Working Memory Assessment (Alloway, 2007), delivered via a computer program, included central executive tasks which did not draw on verbal skills for processing, manipulation or retention. While the CE is seen as modality free, using measures which rely heavily on verbal fluency may introduce a confounding factor. It is important therefore to establish tasks with less dependence on oral language skills for younger children to give a more accurate picture of potential, particularly for those with limited or delayed language or who are not fluent in English.

The results presented here support existing research (Alloway et al., 2009; Gathercole et al., 2003) in implications for assessment and teaching. Assessment of working memory performance is important for establishing the reasons for children's underperformance from the beginning of their school careers. Such assessments would be helpful in addition to measures of reading or pre-reading ability because children with less favourable pre-school experience who have not had experience with rhyme detection or other phonological awareness activities may underperform on those tasks but show potential in working memory assessments. Conversely, those with good language skills and positive enhancing early experiences, for example, may appear to be achieving well at first, but with poorer working skills, may struggle later. The years from Reception to Year 2 are crucial years for children in establishing basic skills and feelings of self-competence for the future. Those with poorer working memory skills would benefit from shorter, clearer classroom instructions, greater routine and visual cues and reminders to aid in task completion (Alloway, 2006; Gathercole et al., 2006; Gathercole & Alloway, 2008). There is evidence that poor working memory skills are unlikely to improve over time as the result of normal schooling; early identification therefore needs to be a precursor to structured and effective classroom and pedagogical modifications (Elliott et al., 2010; Gathercole & Alloway, 2008). Research suggests that gains in remedial training programmes do not transfer reliably to classroom activities and academic outcomes such as reading and maths (Dunning et al., 2013; Melby-Lervåg & Hulme, 2013; St Clair-Thompson et al., 2010).

In conclusion, the research presented here indicates that working memory skills and central executive functioning in particular, are an important predictor of later reading ability. Of the tests used, the sentence verification task and the alliteration fluency task are the best performing predictors, but it is argued that these should be considered as a part of the battery rather than screening tools on their own on theoretical grounds. The battery as a whole gives valuable additional information on which aspects of working memory are areas of strength or weakness, and therefore where to target support. It has the potential to be a valuable addition to the classroom assessment materials for both teachers and support staff.

## ACKNOWLEDGEMENT

Open access funding enabled and organized by ProjektDEAL.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ETHICS APPROVAL STATEMENT

This study was carried out in compliance with the British Psychological Society's ethical guidelines. Ethics approval was granted for the study as part of the corresponding author's PhD data collection. All data was anonymised by allocating each participant an identification number, stored separately from consent forms. The dataset is stored securely protected by passwords and is only available upon request.

## ORCID

*Susan J. Atkinson* https://orcid.org/0000-0003-1863-2164

## ENDNOTE

[1] The intermediary group showed some risk factors for dyslexia as determined by the DEST, but not sufficient to be classed as at risk.

## REFERENCES

Akaike, H. (1987). Factor analysis and the AIC. *Psychometrika*, *52*, 317–332.

Alloway, T. P. (2006). How does working memory work in the classroom? *Educational Research and Reviews*, *1*(4), 134–139.

Alloway, T. P. (2007). *Automated working memory assessment*. Pearson Assessment.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Alloway, T. P., Alloway, R. G., & Wootan, S. (2014). Home sweet home: Does where you live matter to working memory and other cognitive skills? *Journal of Experimental Child Psychology*, *124*, 124–131. https://doi.org/10.1016/j.jecp.2013.11.012

Alloway, T. P., Doherty-Sneddon, G., & Forbes, L. (2012). Teachers perceptions of classroom behaviour and working memory. *Educational Research and Reviews*, *7*(6), 138–142. https://doi.org/10.5897/ERR11.223

Alloway, T. P., Gathercole, S., Willis, C. S., & Adams, A.-M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*(2), 85–106.

Alloway, T. P., Gathercole, S. E., & Kirkwood, H. (2008). *The working memory rating scale*. Pearson Assessment.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development*, *80*(2), 606–621.

Atkinson, A. L., Allen, R. J., & Waterman, A. H. (2021). Exploring the understanding and experience of working memory in teaching professionals: A large-sample questionnaire study. *Teaching and Teacher Education*, *103*, 103343. https://doi.org/10.1016/j.tate.2021.103343

Atkinson, S.J. & Mitchell, N. (2015). *Working memory strengths and deficits: Incidence and coping strategies in a school population*. Proceedings of the social competences, Creativity and Wellbeing Symposium, Plovdiv.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, *49A*, 5–28.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–422.

Baddeley, A. (2017). *Exploring working memory: Selected works of Alan Baddeley*. Routledge.

Baddeley, A. & Hitch, G. (1974). Working memory. In Bower, G.A. (Ed.) *Recent advances in learning and motivation* (pp. 47–90). Academic Press.

Baddeley, A., Hitch, G., & Allen, R. (2021). A multicomponent model of working memory. In R. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: The state of the science*. Oxford University Press.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based, resource-sharing model account. *Developmental Psychology*, *45*, 477–490.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology*, *41*, 579–597.

Beneventi, H., Tønnessen, F. E., Ersland, L., & Hugdahl, K. (2010). Working memory deficit in dyslexia: Behavioral and FMRI evidence. *International Journal of Neuroscience*, *120*(1), 51–59. https://doi.org/10.3109/00207450903275129

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Booth, J. N., & Boyle, J. M. E. (2009). The role of inhibitory functioning in children's reading skills. *Educational Psychology in Practice*, *25*(4), 339–350.

Booth, J. N., Boyle, J. M. E., & Kelly, S. W. (2010). Do tasks make a difference? Accounting for heterogeneity of performance of children with reading difficulties on tasks of executive function: Findings from a meta-analysis. *British Journal of Developmental Psychology*, *28*, 133–176.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Sage Publications.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053.

Carmines, E. G., & Mciver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohmstedt & E. F. Borgatta (Eds.), *Social Measurement*. Sage Publications.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology*, *83*, 264–274.

de Jong, P. F. (2006). Understanding normal and impaired reading development: A working memory perspective. In S. J. Pickering (Ed.), *Working memory and education* (pp. 34–61). Academic Press.

Denckla, M. B., & Cutting, L. E. (1999). History and significance of rapid automatised naming. *Annals of Dyslexia*, *49*, 29–41.

Department for Education and Department of Health (2015) Special educational needs and disability code of practice: 0 to 25 years. Available at: https://www.gov.uk/government/publications/send-code-of-practice-0-to-25 (Accessed: 25 March 2021).

DfE/EEF. (2019). *The ITT Core content framework*. DfE. Crown Copyright.

Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). A confirmatory factor analysis of the hospital anxiety and depression scale: Comparing empirically and theoretically derived structures. *British Journal of Clinical Psychology*, *39*, 79–94.

Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science*, *16*(6), 915–925. https://doi.org/10.1111/desc.12068

Elliott, J. G., Gathercole, S. E., Alloway, T. P., Holmes, J., & Kirkwood, H. (2010). An evaluation of a classroom-based intervention to help overcome working memory difficulties and improve long-term academic achievement. *Journal of Cognitive Education and Psychology*, *9*(3), 227–250. https://doi.org/10.1891/1945-8959.9.3.227

Fawcett & Nicolson. (2004). *The dyslexia screening test – Junior*. The Psychological Corporation.

Fawcett, A. J., Singleton, C. H., & Peer, L. (1998). Advances in early years screening for dyslexia in the United Kingdom. *Annals of Dyslexia*, *48*(1), 57–88.

Frederickson, N., Frith, U., & Reason, R. (1997). *Phonological assessment battery*. Windsor & NFER-Nelson.

Friso-van den Bos, I., & van de Weijer-Bergsma, E. (2020). Classroom versus individual working memory assessment: Predicting academic achievement and the role of attention and response inhibition. *Memory*, *28*(1), 70–82. https://doi.org/10.1080/09658211.2019.1682170

Frith, U., Landerl, K., & Frith, C. (1995). Dyslexia and verbal fluency: More evidence for a phonological deficit. *Dyslexia*, *1*, 2–11.

Gathercole, S., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. Sage.

Gathercole, S., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–418.

Gathercole, S. E., & Baddeley, A. (1989). Evaluation of the role of phonological short-term memory in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*, 200–213.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*, 439–454.

Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology*, *20*, 109–122.

Gathercole, S.E., Lamont, E. & Alloway, T.P. (2006). Working memory in the classroom. In S. Pickering (Ed.). *Working memory and education*. (pp. 220–238). Elsevier Press.

Gathercole, S. E., & Pickering, S. J. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education*, *28*(2), 89–97.

Henry, L. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, *9*, 4/5/6, 233–247,

Henry, L. (2012). *The development of working memory in children*. Sage Publications Ltd.

Holmes, J., & Adams, J. W. (2006). Working memory and Children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, *26*(3), 339–366.

Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues and applications*. Sage Publications.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, *38*, 241–253.

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and nonverbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, *20*, 123–140.

Jastak, J. J., & Wilkinson, G. (1984). *Manual: The wide range achievement test – Revised. Wilmington*. Jastak Associates.

Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8*. Scientific Software International.

Just, M. A., & Carpenter, R. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Klein, K., & Boals, A. (2001). The relationship of life event stress and working memory capacity. *Applied Cognitive Psychology*, *15*, 565–579.

Kline, R. B. (1988). *Principles and practice of structural equation modeling*. Guilford Press.

Lavigne-Cerván, R., Costa-López, B., Juárez-Ruiz de Mier, R., Real-Fernández, M., Sánchez-Muñoz de León, M., & Navarro-Soria, I. (2021). Consequences of COVID-19 confinement on anxiety, sleep and executive functions of children and adolescents in Spain. *Frontiers in Psychology*, *12*, 565516. https://doi.org/10.3389/fpsyg.2021.565516

Mammarella, I. C., Caviola, S., Giofrè, D., & Szűcs, D. (2018). The underlying structure of visuospatial working memory in children with mathematical learning disability. *British Journal of Developmental Psychology*, *36*(2), 220–235. https://doi.org/10.1111/bjdp.12202

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The test of everyday attention for children. Bury St*. Thames Valley Test Co. Ltd.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391–410.

Mathes, P. G., & Denton, C. A. (2002). The prevention and identification of reading disability. *Seminars in Pediatric Neurology*, *9*(3), 185–191.

McNulty, M. A. (2003). Dyslexia and the life course. *Journal of Learning Disabilities*, *36*(4), 363–381.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Muter, V., Hulme, C., & Snowling, M. (1996). *Phonological abilities test*. The Psychological Corporation.

Nicolson, R. (1981). The relationship between memory span and processing speed. In M. Friedman, J. P. Dass, & N. O'Connor (Eds.), *Intelligence and learning*. Plenum Press.

Nicolson, R., & Fawcett, A. (1996). *Dyslexia early screening test*. The Psychological Corporation.

Nicolson, R., Fawcett, A., Moss, H., & Nicolson, M. K. (1999). Early reading intervention can be effective and cost-effective. *British Journal of Educational Psychology*, *69*, 47–62.

Normand, S., & Tannock, R. (2014). Screening for working memory deficits in the classroom: The psychometric properties of the working memory rating scale in a longitudinal school-based study. *Journal of Attention Disorders*, *18*(4), 294–304.

Palmer, S. (2000a). Development of phonological recoding and literacy acquisition; a four-year cross-sequential study. *British Journal of Developmental Psychology*, *18*, 533–555.

Palmer, S. (2000b). Working memory: A developmental study of phonological recoding. *Memory*, *8*(3), 179–193.

Palmer, S. (2000c). Phonological recoding deficit in working memory of dyslexic teenagers. *Journal of Research in Reading*, *23*(1), 28–40.

Pham, A.V. & Hasson, R.M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, first published online May *31*, 2014, 477. https://doi.org/10.1093/arclin/acu024

Pickering, S.J. (2004). *Working memory in dyslexia: Beyond the phonological loop*. Paper presented at the sixth BDA international conference, march 2004.

Pickering, S. J., & Gathercole, S. E. (2001). *The working memory test battery for children*. The Psychological Corporation.

Raven, J. C. (1995). *Coloured progressive matrices, sets a, AB, B*. Oxford Psychologists Press.

Reiter, A., Tucha, O., & Lange, K. W. (2004). Executive functions in children with dyslexia. *Dyslexia*, *11*, 116–131.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, *45A*, 21–50.

Simpson, J., & Everatt, J. (2005). Reception class predictors of literacy skills. *British Journal of Educational Psychology*, *75*(2), 171–188.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657–1661.

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory*, *15*(1), 34–56.

Smith-Spark, J. H., Fisk, J. E., Fawcett, A. J., & Nicolson, R. I. (2003). Central executive impairments in adult dyslexics: Evidence from visuospatial working memory performance. *European Journal of Cognitive Psychology*, *15*(4), 567–587.

Snowling, M., & Hulme, C. (1989). A longitudinal case study of developmental phonological dyslexia. *Cognitive Neuropsychology*, *6*, 379–401.

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading*, *18*(2), 132–138.

St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203–219. https://doi.org/10.1080/01443410903509259

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–406.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search in secondary memory. *Psychological Review*, *114*, 104–132.

**How to cite this article:** Atkinson, S. J. & Martin, C. R. (2022). Early detection of risk of reading difficulties using a working memory assessment battery. *British Educational Research Journal*, *00*, 1–15. https://doi.org/10.1002/berj.3821