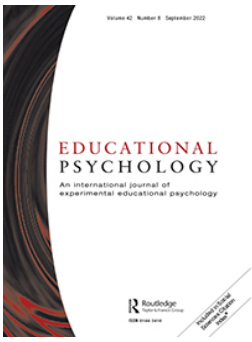

Routledge
Taylor & Francis Group

Educational Psychology

An International Journal of Experimental Educational Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cedp20

Underachievement at school relative to potential: links between reasoning, phonological decoding, short-term memory, and complex grammar

Julia R. Badger, Benjamin C. Howarth, Elena Svirko & Jane Mellanby

To cite this article: Julia R. Badger, Benjamin C. Howarth, Elena Svirko & Jane Mellanby (2022) Underachievement at school relative to potential: links between reasoning, phonological decoding, short-term memory, and complex grammar, Educational Psychology, 42:8, 952-971, DOI: 10.1080/01443410.2022.2115978

To link to this article: https://doi.org/10.1080/01443410.2022.2115978

EDUCATIONAL PSYCHOLOGY
2022, VOL. 42, NO. 8, 952–971
https://doi.org/10.1080/01443410.2022.2115978

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Underachievement at school relative to potential: links between reasoning, phonological decoding, short-term memory, and complex grammar

Julia R. Badger[a], Benjamin C. Howarth[b], Elena Svirko[a] and Jane Mellanby[a]

[a]Department of Experimental Psychology, Radcliffe Observatory Quarter, University of Oxford, Oxford, UK; [b]Academic Centre, Level 2 John Radcliffe Hospital, University of Oxford Clinical Medical School, Oxford, UK

ABSTRACT
Around 6% of U.K. children are underachieving at school relative to their potential (URP). We explored whether difficulties with phonological decoding, short-term memory (STM), and complex grammar may be responsible. We compared school-based reading test data or formal SATs, and verbal reasoning in 2462 children (150 URP and 2312 non-URP children) and administered a 7-min follow-up test to a matched subgroup of 106 URP and 106 non-URP children. Thirty-three of our original 150 URP children (22%) scored in the top 10% nationally for verbal reasoning, compared with 46 of our 2312 non-URP children (2%). Phonological decoding, STM, and complex grammar acquisition made independent URP contributions in children aged 7–9; in ages 10–12, only phonological decoding contributed. URP children were more likely to show multiple difficulties. The 7-min test enables school staff to quickly identify areas for targeted support. Some URP children not previously detected as capable may be encouraged to take up and succeed in further education.

ARTICLE HISTORY
Received 1 June 2020
Accepted 18 August 2022

KEYWORDS
Underachievement; potential; reasoning; learning; language

Introduction

In England, children at most state-maintained schools are tested on their knowledge and educational progression in line with the National Curriculum—a government-approved programme of study. Since 1995, primary school children aged 10–11 years have completed formal Statutory Assessment Tests (SATs) in English, Mathematics, and Science. In addition, many secondary schools use the Cognitive Abilities Tests (CAT) (GL Assessment, 2012; Thorndike et al., 1986), a verbal, quantitative, and non-verbal reasoning test, to give a more nuanced measure of academic potential than SATs. The validity of such tests has been established by showing that they are highly predictive

CONTACT Julia R. Badger julia.badger@psy.ox.ac.uk Department of Experimental Psychology, Radcliffe Observatory Quarter, University of Oxford, Anna Watts Building, Oxford, OX2 6GG, UK

of achievement in GCSEs (General Certificate of School Education), taken five or six years later. Strand (2006) showed that a combination of the two measures accounts for 86% of the variance in GCSE scores. These tests allow the identification of those children who are underachieving at school relative to the overall expectation of children their age. This has led to the use of SATs and the CAT for setting school targets at the start of secondary school, and individual pupil targets for GCSE attainment.

However, sometimes SATs and the CAT scores are an underestimate of what a child may be capable of achieving since the tests require more than raw 'ability': SATs measure knowledge and require reading and writing which are of course taught; the CAT, although less dependent on knowledge, also require reading. Furthermore, many children find SATs highly anxiety-promoting since so much is thought to depend on them, both for national comparisons of schools (league tables) and for the children's own destiny (see McDonald, 2001 for a review on the effects of test anxiety).

Underachieving relative to overall *expectations* (as measured by SATs and the CAT) is however only one kind of underachievement; there is another way of looking at this: identifying children who are underachieving relative to their *potential* when a discrepancy is identified between an individual's general ability (fluid intelligence) and their academic achievement.

Academic potential

Cattell proposed two elements of intelligence: fluid intelligence (Gf) which defines an individual's underlying ability to reason and think flexibly, and crystalised intelligence (Gc) which defines an individual's acquired information and skill, such as reading and writing. This was adjusted by Carroll (1993) to form a 3-level hierarchical model with very specific abilities making up the first level, Gf and Gc making up the second level, and a combined over-arching general intelligence making up the top level (in line with Spearman's 1904 'g'). Although more recent intelligence theories and models question the necessity or accuracy of an overarching 'g' level (Horn & Blankson, 2014), most academic tests, and indeed many reasoning tests, still require a mixture of Gf and Gc ('g') for a top score to be achieved. However, this does not provide a true reflection of an individual's academic potential (Gf) since this depends on environmental factors. A measure of academic potential needs to be an approximation to Gf (fluid intelligence) uncontaminated by Gc (crystallized intelligence).

It has been argued that the Verbal and Spatial Reasoning test for Children (VESPARCH; Mellanby et al., 2016) *is* an approximation to Gf (Badger & Mellanby, 2018a). The online reasoning test suitable for individual or whole class administration measures the ability to solve novel problems using simple categories and analogies—a core part of fluid intelligence (Schneider & McGrew, 2012)—without reliance on crystalised intelligence. A score on VESPARCH provides a score of putative academic potential (see Measures for further details).

The VESPARCH tests (one verbal and one spatial; one suitable for Yr3-4 children and one suitable for Yr6-7 children) have good test-retest reliability, construct validity, and conformity to the Rasch Model and are fully standardised (see Badger & Mellanby, 2018a). They were developed in line with the Verbal and Spatial Reasoning test for adults (VESPAR; Langdon & Warrington, 1995), but for the general population of children. The VESPAR test was designed for a *clinical* adult population and sought not only to disentangle the verbal and spatial abilities of individuals but allow for direct comparison of these abilities across matched tests rather than comparing scores from two separate tests with different designs and scoring. It was also designed to create a level playing field whereby only high-frequency words and simple concepts were used and no time limit was enforced. The test has been used in many clinical settings including exploring sequential learning in dyslexia (Kelly et al., 2002), and considering the hemispheric role in reasoning with unilateral hemisphere lesion patients (Langdon & Warrington, 2000).

Underachieving relative to potential (URP)

Reports have long stated the need to identify and support vulnerable children through mandatory and optional education in an attempt to reduce the achievement gap and widen participation in further education (Department for Children Schools and Families, 2009; DES, 1985; Gov.UK, 2019). Although reasons for academic underachievement can vary enormously, they broadly fall under the categories of biological or environmental origin. Developments in cognitive neuroscience have allowed us to better understand the links between learning and brain development and identify biological markers for increased cognitive and educational difficulties (see Blakemore & Frith, 2005; Byrnes, 2001). At the same time, research has reported ethnicity, gender, and socioeconomic status as playing significant environmental roles in the achievement gap (Strand, 2014a, 2014b). In fact, more recently, evidence was provided to show that for the most socially disadvantaged children, this gap may actually be increasing (Andrews et al., 2017). However, children falling into either, both, or neither of these categories could be underachieving relative to their *own* potential, regardless of where they place on the Government's overall expected achievement levels: they may not be marked as underachieving at school at all but that does not mean that they are reaching their potential. It is therefore essential that we identify children's potential, and work to create tailored interventions to support those who are URP.

Badger and Mellanby (2018a) directly compared U.K. children's scores on VESPARCH (a measure of potential) against their scores on SATs or primary school reading tests (school-based achievement data) and identified that 6% of pupils were underachieving at school relative to their potential (URP; achievement data on school tests was at least 1.5 *SD* lower than academic potential data showed they were capable of achieving. This arbitrary cut-off of at least 1.5 *SD* follows the approach of the regression model used to calculate underachievement; see Lau and Chan (2001) and Phillipson and Tse (2007) for thorough discussions of methods used to identify underachievers). This 6% of 'academically missed' (URP) children have their potential hidden by current school tests. As a consequence, the future academic performance of URP children may well be depressed by reduced aspirations and self-fulfilling prophesy. Those who show the average, or even good performance, may actually be URP if their true potential has not been recognised. Identification of those URP, and who may well also be part of the achievement gap, is the first step. What is then needed is to find reasons for their underperformance so that identified problems can be addressed and if possible remediated.

Identifying reasons for URP

In Badger and Mellanby's (2018a) paper they stated that 6% of U.K. children are URP, but did not explore the possible reasons why. When schools are given information on which of their children are URP, the response is to seek further information on what the school might do about this. Therefore, the next step is to try to identify the causes of the underachievement. There is potentially a range of reasons why children may be URP but to test everything would be incredibly time-consuming for already very busy teaching staff and may not be necessary. As a starting measure, and in an attempt to extend Badger and Mellanby's work, we devised a short follow-up test (about 7 min) to measure three skills. The test can be administered by teaching assistants (or teachers).

The three skills detailed below are by no means the only potential difficulties that may hold children back from reaching their potential when biological and environmental factors are controlled for. However, they are three skills deliberately removed from, and not required during, the VESPARCH tests (see the Method section on VESPARCH for more detail) but which are required during school-based tests. These skills were removed from VESPARCH to create a level playing field for all children and therefore measure only fluid intelligence: these three skills are often associated with learning (and learning difficulties) but are three of the more easily addressed in school. If a child scores 1.5 *SD* higher on VESPARCH than on school tests, it is very likely that one of these three skills is the barrier.

The purpose of the short follow-up test is to measure, phonological decoding, short-term memory, and complex grammar acquisition. Difficulties in any of these three skills can lead to misunderstanding or misinterpretation during lessons and in examinations which in turn could lead to underachieving relative to potential.

Phonological decoding

Reading is the key to most educational achievements. Identifying a reading disability, particularly where its presence has not previously been suspected, is a prerequisite for its remediation. There is a wide literature on methods of remediating reading (see e.g. Hulme et al., 2015; Snowling, 1980; Snowling & Hulme, 2007). Poor reading ability is often unacknowledged if the reading age of a child corresponds to his/her chronological age. However, a child's reading level may be below what his/her potential would predict is possible to achieve. Phonological problems have been shown to be a core deficit in the majority of children with a reading disability (Melby-Lervåg et al., 2012) and play a significant role in underachievement (Snowling, 2013). If children struggle with word-reading then they may not be able to understand specific words within a test question which will lead to misunderstanding and incorrect answering based on reading ability rather than general ability. Proficiency in phonological decoding, or reading, is not required during the VESPARCH tests.

Complex grammar acquisition

Poor oral language level has also been shown to be of importance in reading (Melby-Lervåg et al., 2012; Snowling & Melby-Lervåg, 2016), but additionally will be of importance in *understanding* instructions and following arguments. Children from deprived backgrounds are likely to have less sophisticated levels of oral language (including more complex forms of grammar) than those from more advantaged backgrounds (Hart & Risley, 1992; Hoff, 2003; Huttenlocher et al., 2010), and if a particular linguistic structure is not acquired early on (see Bortfield & Whitehurst, 2001), it can lead to long-term problems in language processing. Differences and difficulties are seen in adulthood (Ross, 1979). Complex grammar in particular, such as conditionals, gives an indication of such problems because they form an important part of the language for studying subjects, such as Science and History where hypothesis-making is important. It has been shown that if a child has not acquired the production and/or comprehension of complex grammar then they will be likely to misunderstand the true meaning of certain discussions, questions, passages of text, and reasoning problems (Svirko et al., 2019). This misunderstanding can lead to incorrect responses due to a lack of grammar knowledge rather than general knowledge. Complete complex grammar acquisition is not required during the VESPARCH tests.

Short-term memory (STM)

Phonological tests share much variance with STM (see Melby-Lervåg et al., 2012), however, STM alone has been found to be a significant predictor of both early and later academic achievement (Schneider & Niklas, 2017). Problems with STM will interfere with understanding and/or remembering instructions, following arguments, comprehension, and reasoning (Baddeley, 1986). Recent work has shown the impact of poor STM on academic achievement as it directly affects reasoning and flexible thinking (Vock et al., 2011). If a child struggles with their STM then they may not be able to hold in mind important information within an examination question, especially from long sections of text. This can lead to confusion, misunderstanding, and incorrect responses because of poor memory of the question rather than poor general information knowledge. Good STM is not required during the VESPARCH tests.

Current study

We tested a large sample of children aged 7–12 (UK years 3–7) on VESPARCH and compared their standardised age scores (SAS) against a measure of school attainment (school-based reading tests or formal SATs). A SAS converts a raw score into a standardised score based on a child's age in years and months. This indicates how a child is performing relative to the national average. Based on the scoring system of the school reading attainment test, data was converted into a point-scale so it could be directly and quantitatively compared against the VESPARCH data, using z-scores, to identify URP. The two tests were typically taken within six months of one another.

In line with Badger and Mellanby (2018a), those whose z-score for VESPARCH was higher by 1.5 *SD* or more than their attainment score were classified as URP. From the URP group, those for whom an individual match could be found in the non-URP group were selected for further analysis. This led to two groups of 106 children each, all of whom completed the three follow-up tests. We predicted that URP children would score lower than non-URP children on tests of phonological decoding, acquisition of complex grammar, and STM.

Method

Participants

A total of 2547 children aged 7–12, with English as their first language, from 14 non-selective primary and secondary schools were tested and identified as URP or non-URP (Figure 1). This study received university ethical approval, and study information letters were sent to the parents of children who participated. Parents could opt-out of their children and children could opt-out on the day. Any opt-out requests were submitted to the class teacher.

Measures

Verbal and spatial reasoning for children (VESPARCH)

There are two versions of the online VESPARCH tests (Mellanby et al., 2016; www.cambridgeassessment.org.uk/vcm): one suitable for children aged 7–9 years (years 3–4, hereafter Yr3-4), and one suitable for children aged 10–12 years (years 6–7, hereafter Yr6-7). Each test has two sections, one verbal and one spatial (40 questions per section in the Yr3-4 test; 50 questions per section in the Yr6-7 test) and is completed individually (but can be administered in a group setting, such as a classroom; see Figure 2 for question examples).

The verbal battery has been designed to be answered without the requirement of several skills important during schooling including the three being measured in this paper's follow-up. Removing these skills (reading, memory, and grammatical acquisition) focuses on a child's fluid rather than their crystalised intelligence. The test does

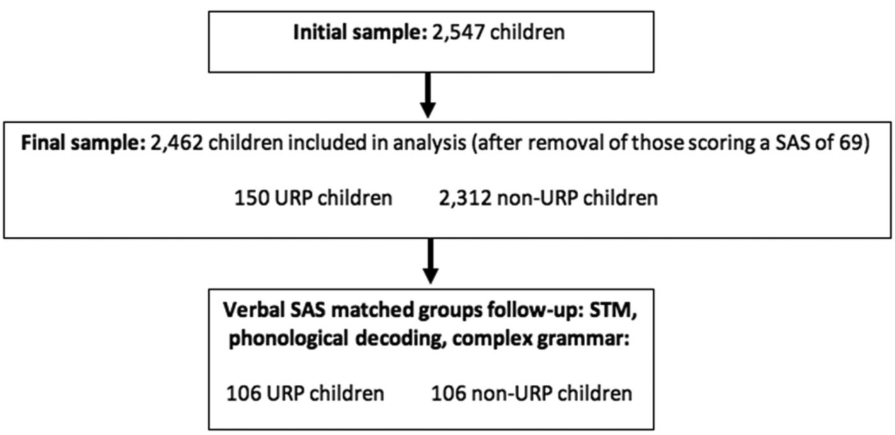

Figure 1. Flow diagram of the number of participants that were tested and included at each stage of the study.

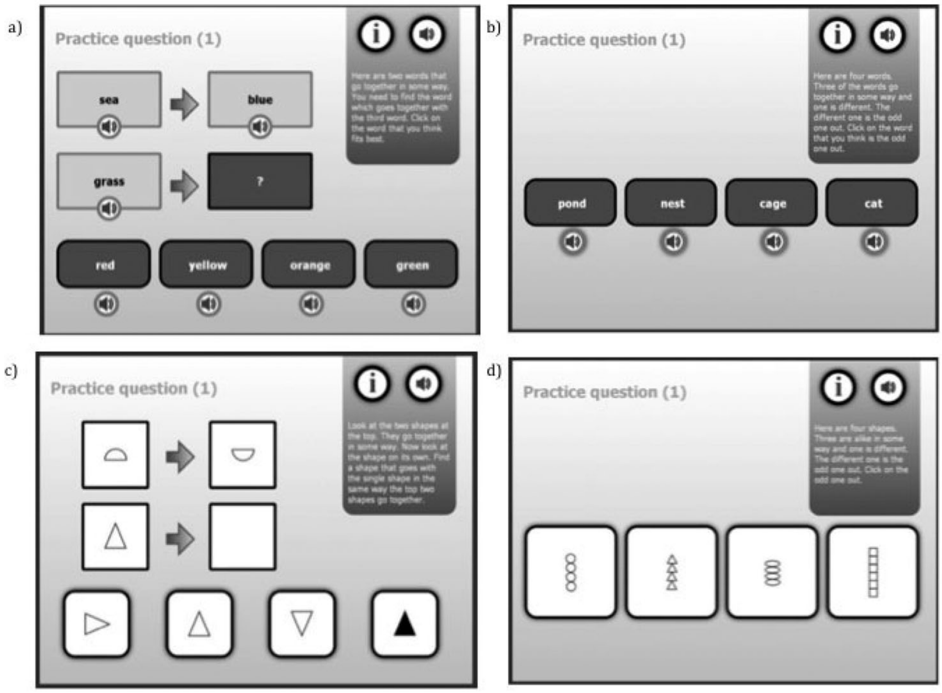

Figure 2. Examples of VESPARCH questions: verbal analogical (a), verbal categorical (b), spatial analogical (c), and spatial categorical (d). Image taken from Badger and Mellanby (2018a).

not depend on level of vocabulary, since all the words are highly familiar; it does not depend on any sophisticated knowledge; it does not require reading or writing as the questions are presented as singular words rather than a whole text and are presented simultaneously through headphones whilst being highlighted on the computer screen; answering just requires clicking on the chosen response. All the instructions are available on screen, are delivered through the headphones, and can be played as many times as needed so the test-taker does not need a good memory. There are extensive practice questions with detailed feedback at the start of each part of the test to provide a 'level playing field' for children who are likely to have variable experience of test-taking in general and in particular of analogies. The spatial battery has been designed to tap into an individual's ability to identify relational patterns between shapes (rather than between words) which in some cases also require visualising movements of shapes or parts of shapes. Spatial tests identify an individual's spatial ability. Both tests are also designed to minimise factors, such as time pressure (they are untimed) which tend to raise anxiety and hence interfere with problem-solving (Eysenck et al., 2007). Raw scores result in four standardised age scores (SAS) in the range 69–141: verbal SAS Yr3-4, spatial SAS Yr3-4, verbal SAS Yr6-7, and spatial SAS Yr6-7. The spatial test—which uses shapes rather than words—can therefore be helpful to identify potential in children whose English is limited, without the restraint of language. Many traditional tests of reasoning are reliant on verbal ability, written and spoken, which can result in inaccurate scores for individuals who struggle with the language of the test rather than the test itself. This can lead to difficulties in accurate class ability setting and providing the correct support to reach one's potential. The spatial test can also identify children who show greater reasoning strengths in a spatial direction which can help to provide information for the best methods of teaching in the classroom for particular children, for example, children with a strong spatial ability may benefit from practical experience, diagrams, and images, as well as the more traditional educational method of reading, writing, and discussion. However, the verbal part of the reasoning test is the best predictor of URP in educational attainment due to the verbal nature of educational assessment. Therefore, this study focusses on Yr3-4 and Yr6-7 SAS data from the verbal part of the reasoning test only. The national verbal VESPARCH SAS average for both Yr3-4 and Yr6-7 is 100 ($SD = 15.00$).

School attainment data

The most recent school attainment data was collected for every participant. These came in the form of school-based reading comprehension data for Yr3 and Yr4 in line with the National Curriculum expectations, which measured the same underlying concept, and Key Stage 2 reading SATs data for the Yr6 and Yr7 children. The data was converted into a comparable point-scale so it could be combined and quantitatively compared against VESPARCH.

Test of non-word phonological decoding

We measured phonological decoding with a brief graded non-word reading test (taken from Nation et al., 2007; Snowling et al., 1996). Teachers' classroom time is limited therefore we chose to use this quick test rather than using longer, more in-depth measures of reading (e.g. Castles & Coltheart, 1993) or an established reading test.

This 24-nonword test increases in phonological complexity, for example, *kisp* to *pragendent*. Children were shown individual words and asked to read them aloud. If the child sounded the letters only, they were encouraged to blend the sounds to say the word as a whole. One point was given for every correctly spoken word, therefore interval data ranged from 0 to 24.

However, an arbitrary cut-off was used to create categorical data, based on the typical distribution of scores, whereby children scoring lower than the group mean were categorised as showing phonological difficulties: Children aged 7–9 years old, who got 15 or fewer words correct, and children aged 10–12 years old, who got 19 or fewer words correct, were recorded as having difficulties with phonological decoding. All other scores were categorised as showing no difficulty with phonological decoding.

The measure has a very strong internal consistency within our sample: Yr3-4 $\alpha = .90$ and Yr6-7 $\alpha = .94$. All errors were written down during testing and categorised into the following groups: (1) adding or removing letters from the nonwords, (2) letter switching, for example, 'drant' becomes 'darnt', or substituting letters, for example, 'prab' becomes 'prib', (3) an inability to blend the phonological sounds, (4) changing the nonwords into real words, for example, 'kisp' becomes 'crisps', or (5) a mixture of error types.

Test of complex language acquisition

The eight-sentence Complex Grammar Repetition Task (Badger & Mellanby, 2018b; Svirko, 2011) was used—the foundation for which is the elicited imitation method (Lust et al., 1996). This method identifies the general language level (Klem et al., 2015; Moll et al., 2015) and also gives insight into ability to reconstruct specific complex grammatical forms—a measure of underlying automatic processing of the grammar. Repetition of complex sentences, such as Type III conditionals, e.g. *'if Peter had picked some flowers, he would have given them to his mum'* requires the individual to reconstruct the grammatical structure of the sentence (e.g. Lombardi & Potter, 1992; Potter & Lombardi, 1990). If a child has not yet internalised this type of grammar, s/he would not be able to reconstruct the grammar of the sentence correctly. A recent paper details the developmental trajectory of Type I, Type II and Type III conditionals in children aged 4- to 11-years of age (Badger et al., 2020) and shows that by the end of Yr3, ∼95% of children can fully produce Type II conditionals. Whereas, ∼55% can fully produce Type III conditionals; this rises to ∼80% of Yr6 children. It is the early acquisition of these Type III conditionals that predicts a child's ability to reason scientifically and understand hypothesis testing (Svirko et al., 2019).

The test comprised four target (type III conditionals) and four control sentences—grammatically simple and of the same length (16–17 syllables). The sentence length of both control and target sentences was such that a child with typical memory capability should not struggle to recall. The score on the control sentences highlights any language or memory difficulties a child may have beyond complex grammar. Each sentence was read aloud to the child and s/he was then asked immediately, after each one, to repeat the sentence exactly. A grammatically complex sentence, such as, *'If Peter had picked some flowers, he would have given them to his mum'* was scored as correct if the child made no errors in the verbs. Omissions or substitutions for nouns or adjectives with words of similar meaning were not scored as a mistake. The fact that children do such substitutions is evidence that they are not just passively recalling the words but reconstructing the sentence. Scores were analysed in their raw form (a range of 0–4) as interval data.

The data was also looked at categorically: Children were recorded as having difficulties with this complex grammar if they scored an arbitrary total of fewer than three (the overall mean) correct out of four sentences. All other children were recorded as showing no difficulty with complex grammar.

The measure has good internal consistency within our sample: Yr3-4 $\alpha = .80$ and Yr6-7 $\alpha = .76$.

Short-term memory test

The grammatically simple (control) sentences in the eight-sentence Grammar Repetition Task (Svirko, 2011) were used as a measure of verbal STM. A grammatically simple sentence, such as *'Peter picked some lovely flowers and gave them to his mum and dad'* was scored as correct if the child made no errors or only one error, such as an addition or removal of a word that does not change the grammatical structure or the meaning of the sentence. Scores were analysed in their raw form (a range of 0–4) as interval data.

The data was also looked at categorically: Children were recorded as having difficulties with STM if they scored an arbitrary total of fewer than three (the overall mean) correct out of four sentences. All other children were recorded as showing no difficulty with STM.

The measure has reasonable internal consistency within our sample: Yr3-4 $\alpha = .57$ and Yr6-7 $\alpha = .59$.

Procedure

Participants completed the online group testing (VESPARCH) in one session in an ICT suite. The 7-min individual testing (for phonological decoding, complex grammar acquisition, and STM) was completed in one session just outside their classroom.

Analysis

Statistical analyses are those based on the SPSS (V.24) platform. Binary logistic regressions and Chi-squared tests were performed on our categorical data. Other analyses with continuous but not normally distributed data used Mann–Whitney tests.

Results

Underachievement relative to potential (URP)

A total of 85 children had a VESPARCH SAS of 69 and have been removed from further analysis: VESPARCH gives a SAS of 69 for all the lowest scores without further discrimination. In the final sample of 2462 children, 150 (6%) were identified as URP and 2312 (94%) children were identified as non-URP.

The URP group had a significantly higher mean verbal VESPARCH SAS of 111 ($SD = 16.78$) compared with the non-URP group which had a mean SAS of 102 ($SD = 13.92$). This difference was similar in the two age groups. Twenty-two percent of the 150 URP children scored 126 or above on verbal VESPARCH, placing them in the top 10% nationally, compared with only 2% of the 2312 non-URP groups.

In the sample of 2462 children, 54% were male. In the sample of 150 URP children, there were nearly twice as many males as females: 65% male compared to 35% female: X^2 $(1) = 8.23$, $p = .004$.

Phonological decoding, short-term memory, and complex grammar

Follow-up data (using the very brief test designed for the purpose of testing phonological decoding, STM, and complex grammar, see Methods) was collected from 106 of the original 150 children identified as URP, 49 Yr3-4 children, and 57 Yr6-7 children, along with 106 individually-matched non-URP children [matched on year group, sex, and verbal VESPARCH SAS: $M = 106$ (14.86) and $M = 106$ (14.89), respectively]. The additional 44 URP children could not be well-matched with a non-URP child based on the stated criteria and were therefore not included in these matched analyses.

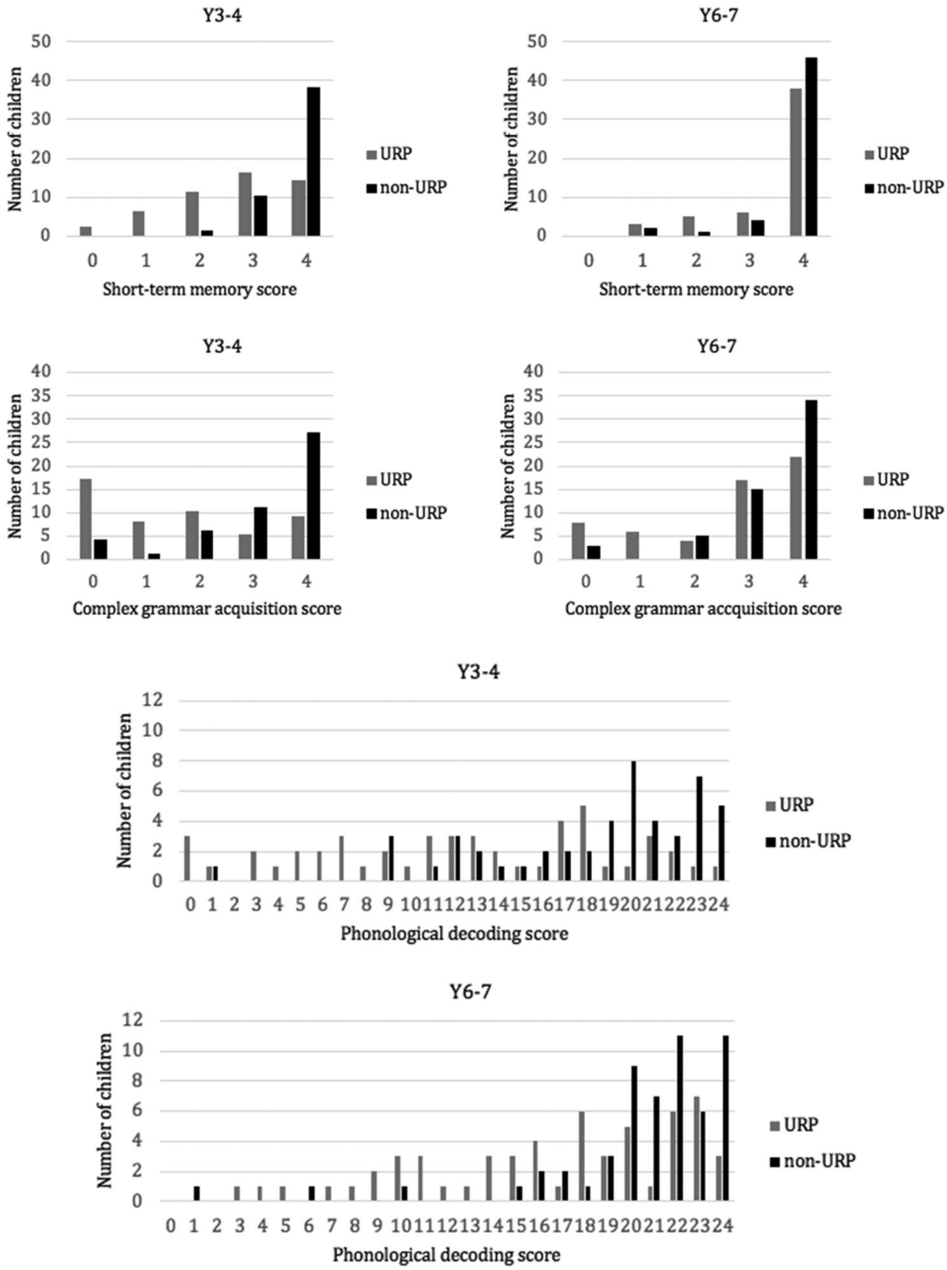

Figure 3. The raw scores are split by year group and whether URP or non-URP, for STM, acquisition of complex grammar, and phonological decoding. Data is from the matched sample of 212 children (49 Yr3-4 URP children *vs.* 49 Yr3-4 non-URP children, and 57 Yr6-7 URP children *vs.* 57 Yr6-7 non-URP children).

The data from the two different age groups have been analysed separately to reflect both the difference in VESPARCH test (one suitable for Yr3-4 and one suitable for Yr6-7 children, but also to reflect the way results are typically reported to UK teachers: start of KS2 (Yr3-4) or the transition from KS2 into KS3 (Yr6-7). For Yr3-4

Table 1. The number and percentage of children underachieving relative to potential (URP) and those non-URP showing difficulties on the follow-up tests are split by age group.

	Underachieving relative to potential			Achieving relative to potential			
URP (*N*)	Phono. decoding % (*N*)	STM % (*N*)	Complex grammar % (*N*)	Phono. decoding % (*N*)	STM % (*N*)	Complex grammar % (*N*)	Non-URP (*N*)
Yr3-4 (49)	61 (30)	51 (25)	71 (35)	24 (12)	2 (1)	22 (11)	Yr3-4 (49)
Yr6-7 (57)	61 (35)	14 (8)	32 (18)	21 (12)	5 (3)	14 (8)	Yr6-7 (57)

data, there was a marked difference in the raw scores between URP and non-URP groups for nonword phonological decoding, STM, and complex grammar acquisition (Figure 3), and this difference was highly statistically significant for each factor: (Mann–Whitney $U = 571$, 521, and 552, respectively, $p < .001$). For Yr6-7, there was also a significant difference in scores for phonological nonword decoding (Mann–Whitney $U = 929$, $p < .001$) and complex grammar (Mann–Whitney $U = 1195$, $p = .009$, see Figure 3) but STM between the URP and non-URP groups was not statistically significant ($p = .078$).

Incidence of specific difficulties

We have used cut-off criteria (see Methods) for classifying children as having or not having difficulty with each of our follow-up measures. In our work with schools, we have found this is the best way to make the identification of such children clear for teachers. It also follows the classification used by Badger and Mellanby (2018b) when using these STM and complex language measures. This creates categorical data.

Chi-square analyses demonstrated that for the Yr3-4 children there was a significantly higher incidence of difficulties within the URP group than within the non-URP group on all three measures: phonological decoding [X^2 (1) = 13.50, $p < .001$], STM [X^2 (1) = 20.35, $p < .001$], and complex grammar acquisition [X^2 (1) = 23.60, $p < .001$]. For Yr6-7 there were more children with phonological problems in the URP group [X^2 (1) = 19.15, p <. 001], and the difference for complex grammar was significant, [X^2 (1) = 4.98, $p = .043$]. But there was no significant difference for STM, [X^2 (1) = 2.52, $p = .113$]. See Table 1 for the breakdown by year groups.

For both age ranges there was a much higher incidence of multiple difficulties (that is difficult with two or three out of three follow-up measures) in the URP groups compared with non-URP: Yr3-4, 69.3 *vs.* 8.1%; X^2 (1) = 36.68; Yr6-7, 70.2 *vs.* 10.7%, X^2 (1) = 42.13, both $p < .001$ (see the number of difficulties encountered, collapsed across ages as comparable, in Figure 4). The data also shows that overall only 23 URP children (22%) were found to have no difficulty with any of the follow-up tests compared to 71 of the non-URP children (67%, see Figure 4): X^2 (1) = 44.04, $p < .001$.

Binary logistic regression

We conducted a binary logistic regression with whether the child was categorised as URP or not as the dependent variable, to look at the contribution of the three follow-up measures in the two separate age ranges. The STM and complex grammar repetition measures had only four items, thus the variability in scores was not great. The

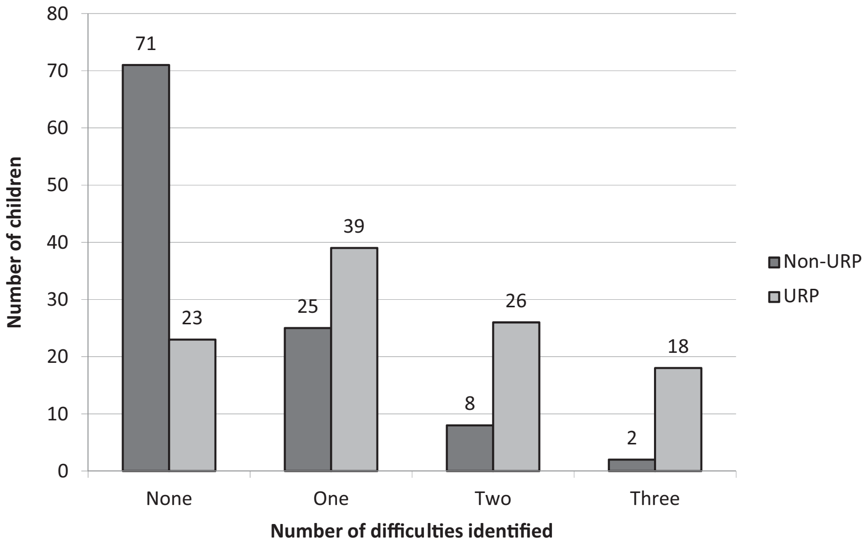

Figure 4. The number of children showing difficulties on one, two, or three of the follow-up tests was split by whether they were identified as underachieving relative to potential (URP) or non-URP.

Table 2. Binary logistic regression identifies that underachievement relative to potential (URP) can be predicted based on phonological decoding, STM, and complex grammar acquisition for Yr3-4.

	B	S.E.	Wald	Sig.	Exp(*B*)	95% C.I. for EXP(*B*) lower	95% C.I. for EXP(*B*) upper
Phono. decoding	−1.063	0.520	4.181	.041*	.345	.125	.957
STM	−2.793	1.099	6.464	.011*	.061	.007	.527
Complex grammar	−1.569	0.514	9.309	.002*	.208	.076	.571

*Denotes significance.

Table 3. Binary logistic regression identifies that underachievement relative to potential (URP) can be predicted by phonological decoding for Yr6-7.

	B	S.E.	Wald	Sig.	Exp(*B*)	95% C.I. for EXP(*B*) lower	95% C.I. for EXP(*B*) upper
Phono. decoding	−1.678	.439	14.63	.001*	.187	.079	.949
STM	−0.553	.876	0.398	.528	.575	.103	5.009
Complex grammar	−0.340	.605	0.317	.574	.712	.218	2.183

*Denotes significance.

performance on the STM measure was also very close to ceiling level in the older age group. Because of these concerns, all three were entered into the regressions as categorical variables: difficulty Yes/No (i.e. the same categorizations as were used for multiple difficulties analyses).

We entered categorical variables as difficulty Yes/No for phonological decoding, STM, and complex grammar. For Yr3-4 data, all three predictor variables contributed significantly to the model (see Table 2), while for Yr6-7 only phonological decoding was an independent significant predictor (see Table 3).

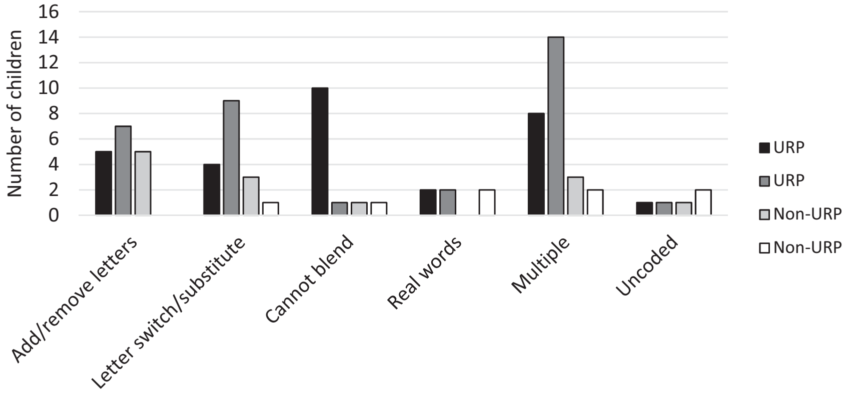

Figure 5. The number of children showing difficulties with different aspects of nonword phonological decoding is split by year group and URP status.

Further analysis of phonological test errors

Phonological decoding was significantly worse for URP children in both age groups and, as detailed in the methods, errors were placed into five categories (1) add/ remove letters, (2) letter switch/substitution, (3) cannot blend, (4) replaced with real words, (5) a mixture of the above four categories. Unfortunately, in five participant cases, the nature of errors was not recorded.

Of those highlighted as having phonological decoding difficulties, all Yr3-4 URP children were able to sound out the letters but a third were unable to blend them (see Figure 5) unlike in the Yr3-4 non-URP group, where only one child was unable to blend the sounds. The Yr6-7 URP children were most likely to show a mixture of difficulties with phonological decoding and when their difficulty was specific, it tended to be with adding/removing or switching/substituting letters within the word. Only eight of the Yr6-7 non-URP children showed any phonological difficulties at all (see Figure 5).

Reading nonwords as real words tended not to be an issue for any year group or URP/non-URP classification with only six children showing this error.

Discussion

This study aimed to investigate whether difficulties with phonological decoding, the acquisition of complex grammar, and STM may be why some children aged 7–12 years are underachieving at school relative to their potential (URP). As in a study by Badger and Mellanby (2018a), we identified 150 (6%) of our sample of 2462 children as URP.

Underachievement relative to potential (URP)

URP was identified when a pupil's score on VESPARCH (potential) was 1.5 *SD* higher than their score on SATs reading or primary school reading test (attainment).

Some of those with the highest scores on verbal reasoning (VESPARCH) were in the URP group. It is particularly striking that the mean verbal VESPARCH score for the URP group ($M = 111$) is about 10% higher than for the non-URP group ($M = 102$) and this difference is similar in Yr3-4 and Yr6-7. Indeed, although the participant numbers are small, 22% of the 150 URP children, compared with only 2% of the 2312 non-URP children, scored a SAS of 126+ (placing them in the top 10% nationally). This study points to VESPARCH as being a way not only to identify those children across the ability range who are underachieving relative to potential (URP) but also as a way to identify the very children that the government and highly selective universities are particularly keen to target—that is bright children from underprivileged backgrounds who have not been identified from school performance and who may ‘slip through the net’ with regards to completing further or higher education. Identifying, at an early age, such children with high verbal VESPARCH scores and encouraging them appropriately might be expected to boost the numbers of those going on to take and succeed in A-levels.

Interestingly, although the male-female divide was equal in our initial sample, we found that the proportion of males in the URP group was approximately twice that of females (65 *vs.* 35%, respectively). This is in line with the well-known finding that in most groups of children with problems in education there is a preponderance of males.

Possible difficulties resulting in URP

As we had predicted, the scores for all three follow-up measures were lower overall in the URP group, with both year groups showing a significantly higher incidence of multiple difficulties than non-URP children, with only 22% of URP children showing no difficulties compared with 67% of non-URP children. While for Yr3-4 children phonological decoding, STM, and complex grammar acquisition were independent significant predictors of URP categorisation, for Yr6-7 children, only phonological decoding was a significant independent predictor.

Phonological decoding

In the URP children, the commonest problem identified in our follow-up tests was in phonological decoding: a total of 60% of the URP children had a big deficit in their ability to decode non-words compared with 20% of non-URP children. This shows that decoding difficulties are still as salient at a later age than that at which children are nowadays tested on phonics (Yr1 and/or Yr2). However, the nature of the decoding problem differed between the year groups in that while with Yr3-4 URP children, their decoding problem was not with letter-sound connection but with blending those sounds, this problem was rarely encountered in the Yr6-7 URP children (or in non-URP children in either age group). Future research needs to look more closely at the reasons behind these deficits to better determine suitable remediation. Since letter-sound connection will necessarily precede the blending of those sounds in development, this suggests that the URP group is showing a delay in phonological development rather than a more enduring deficit. Needless to say, a longitudinal study rather than the present cross-sectional work is needed to investigate this point further: what percentage of these children ‘catch up’ to their peers’ phonological development and for the ones that do not catch up, are there clear delayed developmental patterns, which might indicate the most appropriate method of intervention?

The URP children in this study have been identified by their large discrepancy between VESPARCH verbal reasoning and scores on tests of school achievement of which reading was an integral part. As already stated, a majority of URP children had a phonological deficit, specifically in their ability to decode nonwords. Such a deficit is commonly seen in children diagnosed as having a specific reading disability, such as 'dyslexia' (Hulme et al., 2015; Melby-Lervåg et al., 2012; Snowling & Melby-Lervåg, 2016). However, specific reading disability has many putative causes and symptoms (see Mellanby & Theobald, 2014 for a brief review) and therefore it is unsurprising that this discrepancy between verbal reasoning ability and school tests requiring reading is not always associated with a decoding problem. Furthermore, if phonological decoding ability is viewed as a spectrum, the children in this study might have phonological problems that are not severe enough to merit a dyslexia diagnosis but still severe enough to hinder their reading performance. Whether remediation by a further emphasis on decoding or by employing alternative strategies will be the better route needs to be assessed for each child.

A further overlap with children considered to have a reading disability (Swanson et al., 2009) is the finding of problems with STM. Although STM was found to be a significant problem during our follow-up tests for the Yr3-4 URP children, it was not identified as a problem for the Yr6-7 URP children. This finding shows that STM is only a limiting factor for scholastic achievement relative to potential in the younger ages and suggests a developmental delay rather than a deficit.

Complex grammar acquisition was significantly poorer in the URP groups in both year groups, although only an independent predictor of URP in the youngest group. This is in agreement with previous work on the predictive validity of a composite of a wide range of grammar types (as measured with TROG, Bishop, 2005) and also of the acquisition of conditionals (a more focussed measure of complex grammar acquisition) on measures of literacy (Svirko, 2011), even when age, non-verbal ability, STM and working memory have been controlled. Furthermore, poor grammar understanding contributes to poor reading comprehension (Nation et al., 2007) which will make it hard for children to learn from books and it will also make understanding teachers' classroom language and hence subject content difficult. Indeed, it has recently been shown that understanding complex conditional sentences predict the ability to succeed in tests of scientific reasoning (Svirko et al., 2019). Where children are not exposed to much complex grammar in the home, it is likely that it will need to be acquired by reading and exposure to speech outside the home. Further investigation should be carried out to consider the impact of in-school targeted support for complex grammar acquisition at different ages via reading and oral conversation, to establish the best time for an intervention.

Limitations

Our brief follow-up assessment was designed to be quick for teachers to administer and score. However, owing to the limited nature of this quick follow-up test, it is not possible to identify *all* reasons leading to URP, and therefore for some children, additional, wider testing is required. Additional tests could move away from reasons associated with misunderstanding in the classroom and move to wider reasons, such as emotional or behavioural issues, difficulties with motivation, school attendance, academic self-concept, and home life situation, many of which are often singled out as contributing to lower levels of attainment (Barry et al., 2002; Biederman et al., 2004; McCoach et al., 2016; Reis & McCoach, 2000). This list, of course, is not exhaustive.

It is also important to note that testing of all data took place at single time-points. We suggest that initial data (VESPARCH and reading tests) is collected on several occasions which would allow a more accurate categorisation of participants by overcoming fluctuations and natural measurement error or variation.

Summary

Children whose academic attainment was substantially lower than their verbal reasoning were identified as underachieving relative to potential (URP; 6%). Twice as many were male. The commonest problem identified was in phonological decoding: a total of 60% of URP children. In both year groups, URP children were found to have a significantly higher incidence of multiple difficulties and fewer cases where no difficulty was identified. This data shows that underachievement relative to potential is not a single-cause issue and investigation at a case-by-case level is necessary to determine which difficulties may be causing this underachievement. Our 7-min individual identification test (measuring phonological decoding, complex grammar acquisition, and short-term memory) can be used quickly in schools and is effective in identifying potential difficulties in children reaching their potential at school. We propose that with our 7-min identification test and targeted support in reaching their true potential, URP children may be able to progress further in education than would otherwise have been the case, including the very children that the government is particularly keen to target (bright children from underprivileged backgrounds who are not identified from school performance).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Realising Potential Trust under number B1760.

ORCID

Julia R. Badger http://orcid.org/0000-0003-3060-1934
Benjamin C. Howarth http://orcid.org/0000-0001-7637-8147

References

Andrews, J., Robinson, D., & Hutchinson, J. (2017). *Closing the gap?: Trends in educational attainment and disadvantage.* Education Policy Institute. Retrieved August, 2017, from https://epi.org.uk/publications-and-research/closing-gap-trends-educational-attainment-disadvantage/

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Badger, J. R., & Mellanby, J. (2018a). Revealing hidden talents: The development, use, and benefit of VESPARCH. *The British Journal of Educational Psychology*, *88*(3), 380–395. https://doi.org/10.1111/bjep.12189

Badger, J. R., & Mellanby, J. (2018b). Producing and understanding conditionals: When does it happen and why does it matter? *Journal of Child Language Acquisition and Development*, *6*, 21–41.

Badger, J. R., Pascoe, E., Jagmetti, L., Baldwin, E., & Mellanby, J. (2020). The developmental trajectory of English conditional grammar in 4- to 11-year-old children. *Journal of Child Language Acquisition and Development*, *8*, 1–21.

Barry, T. D., Lyman, R. D., & Klinger, L. G. (2002). Academic underachievement and attention-deficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of School Psychology*, *40*(3), 259–283. https://doi.org/10.1016/S0022-4405(02)00100-0

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., Morgan, C. L., & Faraone, S. V. (2004). Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. *Journal of Consulting and Clinical Psychology*, *72*(5), 757–766. https://doi.org/10.1037/0022-006X.72.5.757

Bishop, D. V. M. (2005). *The test for reception of grammar* [Grammar Test]. Harcourt Assessment.

Blakemore, S.-J., & Frith, U. (2005). *The learning brain: Lessons for education*. Blackwell Publishing.

Bortfield, H., & Whitehurst, G. (2001). Sensitive periods to first language acquisition. In D. Bailey, J. Bruer, J. Lichtman & F. Symons (Eds.), *Critical thinking about critical periods: Perspectives from biology, psychology, and education* (pp. 173–192). Paul H. Brookes Publishing.

Byrnes, J. P. (2001). *Minds, brains, and learning: Understanding the psychological and educational relevance of neuroscientific research*. Guilford Press.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge University Press.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, *47*(2), 149–180. https://doi.org/10.1016/0010-0277(93)90003-E

DES (1985). *Education for all. The report of the committee of enquiry into the education of children from ethnic minority groups (Swann Report).* Her Majesty's Stationery Office (HMSO).

Department for Children Schools and Families. (2009). *Deparivation and education: The evidence on pupils in England, foundation stage to key stage 4.* DCSF.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Gov.UK. (2019). *Education Secretary: Do more for disadvantaged students*. Retrieved from https://www.gov.uk/government/news/education-secretary-do-more-for-disadvantaged-students

GL Assessment. (2012). *The cognitive abilities test* [Reasoning Test].

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, *28*(6), 1096–1105. https://doi.org/10.1037/0012-1649.28.6.1096

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Horn, J. L., & Blankson, A. N. (2014). Foundation for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual theories and issues* (3rd ed., pp. 73–98). Guilford Press.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at family risk of dyslexia. *Psychological Science*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, *8*(1), 43–52. https://doi.org/10.1002/dys.208

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A H., Gustafsson, J.-E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*(1), 146–154. https://doi.org/10.1111/desc.12202

Langdon, D. W., & Warrington, E. K. (1995). *Verbal and spatial reasoning test*. Psychology Press.

Langdon, D. W., & Warrington, E. K. (2000). The role of left hemisphere in verbal and spatial reasoning tasks. *Cortex*, *36*(5), 691–702.

Lau, K.-L., & Chan, D. W. (2001). Identification of underachievers in Hong Kong: Do different methods select different underachievers? *Educational Studies*, *27*(2), 187–200. https://doi.org/10.1080/03055690120050419

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713–733. https://doi.org/10.1016/0749-596X(92)90036-W

Lust, B., Flynn, S., & Foley, C. (1996). What children know about what they say: Elicited Imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee & H. Smith Cairns (Eds.), *Methods for assessing syntax* (pp. 55–76). MIT Press.

McCoach, B., Newton, S. D., Siegle, D., Baslanti, U., & Picho, K. (2016). Is having low motivation the same as not having high motivation? Comparing the CSAS-R and the SAAS-R. *High Ability Studies*, *27*(1), 61–81. https://doi.org/10.1080/13598139.2015.1103209

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, *21*(1), 89–101. https://doi.org/10.1080/01443410020019867

Mellanby, J., McElwee, S., & Badger, J. R. (2016). *Verbal and spatial reasoning tests for children: VESPARCH* [Reasoning Test]. Cambridge Assessment.

Mellanby, J., & Theobald, K. (2014). *Education and learning: An evidence-based approach*. Wiley Blackwell.

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), 322–352. https://doi.org/10.1037/a0026744

Moll, K., Hulme, C., Nag, S., & Snowling, M. J. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Applied Psycholinguistics*, *36*(2), 203–219. https://doi.org/10.1017/S0142716413000209

Nation, K., Snowling, M. J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech Language Pathology*, *9*(2), 131–139. https://doi.org/10.1080/14417040601145166

Phillipson, S. N., & Tse, A. K. (2007). Discovering patterns of achievement in Hong Kong students: An application of the Rasch measurement model. *High Ability Studies*, *18*(2), 173–190. https://doi.org/10.1080/13598130701709640

Potter, M., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654. https://doi.org/10.1016/0749-596X(90)90042-X

Reis, S. M., & McCoach, D. B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly*, *44*(3), 152–170.

Ross, H. (1979). Where's English? In C. Fillmore & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behaviour* (pp. 127–163). Academic Press.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment theories and issues* (3rd ed., pp. 99–114). Guilford Press.

Schneider, W., & Niklas, F. (2017). Intelligence and verbal short-term memory/working memory: Their interrelationships from childhood to young adulthood and their impact on academic achievement. *Journal of Intelligence*, *5*(2), 26–19. https://doi.org/10.3390/jintelligence5020026

Snowling, M. J. (1980). The development of grapheme-phoneme correspondence in normal and dyslexic readers. *Journal of Experimental Child Psychology*, *29*(2), 294–305. https://doi.org/10.1016/0022-0965(80)90021-1

Snowling, M. J., & Hulme, C. (2007). *The science of reading: A handbook* (2nd ed.). Wiley Blackwell.

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin*, *142*(5), 498–545. https://doi.org/10.1037/bul0000037

Snowling, M. J. (2013). Early identification and interventions for dyslexia: A contemporary view. *Journal of Research in Special Educational Needs*, *13*(1), 7–14. https://doi.org/10.1111/j.1471-3802.2012.01262.x

Snowling, M. J., Stothard, S., & McLean, J. (1996). *Graded nonword reading test*. Thames Valley Test Company.

Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–293. https://doi.org/10.2307/1412107

Strand, S. (2006). Comparing the predictive validity of reasoning tests and national end of key stage 2 tests: Which tests are the 'best'? *British Educational Research Journal*, *32*(2), 209–225. https://doi.org/10.1080/01411920600569073

Strand, S. (2014a). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it' for the white working class. *Research Papers in Education*, *29*(2), 131–171. https://doi.org/10.1080/02671522.2013.767370

Strand, S. (2014b). School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education*, *40*(2), 223–245. https://doi.org/10.1080/03054985.2014.891980

Svirko, E. (2011). *Individual differences in complex grammar acquisition: Causes and consequences* [Unpublished doctoral thesis]. University of Oxford.

Svirko, E., Gabbott, E., Badger, J., & Mellanby, J. (2019). Does acquisition of hypothetical conditional sentences contribute to understanding the principles of scientific enquiry? *Cognitive Development*, *51*, 46–57. https://doi.org/10.1016/j.cogdev.2019.05.008

Swanson, H. L., Zheng, X., & Jerman, O. (2009). Working memory, short-term memory and reading disabilities. A selective meta-analysis of the literature. *Journal of Learning Disabilities*, *43*, 260–287. https://doi.org/10.1177/0022219409331958

Thorndike, R. L., Hagen, E., & France, N. (1986). *Cognitive abilities test* (2nd ed.) [Reasoning Test]. NFER Nelson.

Vock, M., Preckel, F., & Holling, H. (2011). Mental abilities and school achievement: A test of a mediation hypothesis. *Intelligence*, *39*(5), 357–369. https://doi.org/10.1016/j.intell.2011.06.006