What are the long-term prospects for children with comprehension weaknesses? A registered report investigating education and employment outcomes

Emma James[1,2], Paul A. Thompson[3], Lucy Bowes[1], & Kate Nation[1]

1. Department of Experimental Psychology, University of Oxford

2. Department of Psychology, University of York

3. Centre for Research in Intellectual and Developmental Disabilities, University of Warwick

Author Version

Accepted for publication in *Journal of Educational Psychology*

15th May 2024

What are the long-term prospects for children with comprehension weaknesses? A registered report investigating education and employment outcomes

Emma James[1,2], Paul A. Thompson[3], Lucy Bowes[1], & Kate Nation[1]

1. Department of Experimental Psychology, University of Oxford

2. Department of Psychology, University of York

3. Centre for Research in Intellectual and Developmental Disabilities, University of Warwick

Author note

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grant funding is available on the ALSPAC website: (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). This publication is the work of the authors, who will serve as guarantors for the contents of this paper.

This research was funded by ESRC (ES/S016333/1).

E.J. moved institutions part-way through this project. Correspondence should be addressed to: Emma James, Department of Psychology, University of York, York, YO10 5DD; Email: emma.james@york.ac.uk

Abstract

Reading is a key gateway to learning, enabling independent access to a range of educational materials. Thus, reading difficulties leave a child particularly vulnerable to academic problems in later schooling and beyond. However, while there is good awareness of children with word reading difficulties within the education system, much less is known about the children who struggle to comprehend texts despite having adequate word reading skills. In this registered report, we investigated the later education and occupational outcomes of 947 children initially identified as having poor reading comprehension at 8-9 years from the Avon Longitudinal Study of Parents and Children (ALSPAC), compared to peers not identified as having a specific reading difficulty (n = 4,516) and those with word reading weaknesses (n = 1,383). We observed that children with comprehension weaknesses (1) were less likely than typical readers to meet UK national educational targets as they progressed through education, with more marked differences for mathematics and science than English; (2) had poorer qualifications at the point of leaving compulsory education, which were comparable to children with word reading weaknesses; and (3) were the group at greatest risk of being out of employment, education and training at age 20. However, there was considerable variability in outcomes, with much of the risk shared with socio-demographic factors. The results address an important gap in knowledge regarding the functional consequences of reading comprehension difficulties in

mid-childhood, and inform discussions concerning the need for identification and targeted support in classroom settings.

Educational impact and implications statement

Some children struggle with reading comprehension beyond what would be expected based on their word reading skills, yet comparatively little targeted support is available for these children. In this study, we used a large UK sample to show that comprehension weaknesses are associated with poorer performance in national assessments at ages 10-11, 13-14, and 15-16 years, with outcomes comparable to children with word reading weaknesses. Children with comprehension weaknesses were also more likely to be out of education and employment at age 20, and much of this risk was shared with socio-demographic factors. The findings inform discussions over the need and nature of support required for reading comprehension weaknesses in the classroom.

Keywords

Poor comprehenders; educational attainment; occupation; longitudinal; ALSPAC

Introduction

Reading problems identified in childhood are associated with educational challenges at school and poor academic outcomes (Smart et al., 2017; Willcutt et al., 2007). This is not surprising as much of education depends on the ability to digest information from written texts. Reading is a complex behaviour that depends on a range of cognitive and linguistic abilities (Castles et al., 2018) and a consequence of this is that reading comprehension may fail for different reasons. Some children find it difficult to identify and process written words leading to reading that is slow, effortful, and error prone. Other children struggle specifically with comprehending text, despite being able to read accurately and fluently. Our focus is with this latter group, often referred to as 'poor comprehenders' (Catts et al., 2006; Nation & Snowling, 1997; Yuill et al., 1989) or 'specific reading comprehension deficit' (S-RCD; Landi & Ryherd, 2017)[1]. Although there is a sizeable literature investigating the nature of their difficulties in mid-childhood, we know very little about the educational and employment outcomes of students who were identified as having reading comprehension weaknesses earlier in development. This is a striking omission given academic achievement depends on the ability to comprehend and arguably, difficulties with comprehension may limit attainment across the curriculum. This study sought to address this knowledge gap by examining educational outcomes of 947 young people with reading comprehension weaknesses in childhood, using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a birth cohort growing up in the UK.

Reading comprehension difficulties are typically identified in research contexts using standardised assessments of reading ability, reflecting individuals with poor text comprehension alongside age-appropriate word reading ability. It is generally accepted that

[1] Note that while our questions are motivated by research into individuals described previously as "poor comprehenders", in this paper we take a person-centred approach in both analysis (capturing probabilistic patterns of weaknesses rather than cut-score determined categories) and terminology (referring to children with comprehension weaknesses).

children with comprehension difficulties form a heterogeneous group with varied strengths and weaknesses that affect the severity of their difficulties (Hayiou-Thomas et al., 2021; James et al., 2023). A consistent and well-replicated finding is that difficulties are not restricted to the comprehension of written text. Instead, these children perform less well than peers on measures of listening comprehension, vocabulary, and aspects of spoken language processing (e.g., Cain & Oakhill, 2006; Nation et al., 2004; Nation & Snowling, 1997). These difficulties are evident before children learn to read, as shown by retrospective studies that look back at the early development of children identified as having comprehension weaknesses on the basis of reading assessments in mid-childhood (Catts et al., 2006; Elwér et al., 2015; Nation et al., 2010). Numerous studies have shown that children with comprehension weaknesses have difficulties with making inferences when reading and listening, and are poor at comprehension monitoring (see Cain, 2022, for review). The strength of this basic research evidence led to a concerted effort to begin investing in classroom translation (Pearson et al., 2020), and the development of promising interventions that target children's reading comprehension (Clarke et al., 2010; Language and Reading Research Consortium (LARRC) et al., 2019; McMaster et al., 2015). Despite this progress, there lacks a clear consensus from the literature on if and how children with comprehension weaknesses despite accurate and fluent word reading should be supported in the classroom.

One reason for this might be that we know very little about educational achievement and employment outcomes in those children identified as having relatively circumscribed difficulties with reading comprehension. There are however many reasons to consider why comprehension weaknesses might impact educational attainment. First, difficulties with reading comprehension may directly affect children's ability to learn from texts (Kintsch, 1986) and therefore perform well in exams. Second, shared associations between reading comprehension and educational outcomes could also reflect that reading comprehension performance is the product of many cognitive skills that are likely relevant for academic attainment. Beyond decoding and linguistic knowledge, these include working memory, executive, reasoning, and inferential skills (Castles et al., 2018; Perfetti & Stafura, 2014; van den Broek & Kendeou, 2022), as well as domain-specific knowledge relating to the text and tasks (Catts & Kamhi, 2014). Thus, reading comprehension performance may be a good predictor of educational outcomes as both draw heavily upon the same set of skills. Third, children with reading comprehension difficulties show low motivation across school subjects and poorer school wellbeing (Torppa et al., 2020), such that this reduced engagement may lead to lower attainment. Yet, while each of these possibilities are intuitive, it might be that the risk for serious and/or broad educational difficulties is not elevated in this population: these children are often considered to have "hidden" reading and language impairments (e.g., Nation et al., 2004), which may be sufficiently mild to not affect longer-term outcomes. Children with comprehension weaknesses may instead be able to capitalise upon their word reading strengths and other cognitive skills to make academic progress. Understanding the implications for comprehension weaknesses across different aspects of educational attainment is thus an important step in helping us understand the nature of poor reading comprehension in this population, and determining whether there is a need for targeted support.

To date, investigations into this question have been small-scale in nature. Cain and Oakhill (2006) followed 17 children with comprehension weaknesses identified in mid-childhood through to 11 years of age and captured their performance on statutory national assessments of English, mathematics, and science. These tests, known as Standardised Assessment Tests (SATs), are used to monitor educational progress in the UK, as children transition from primary to secondary school at age 11 (and again at age 14). Although those previously identified as having comprehension weaknesses performed less well than good comprehenders across all three domains (English, mathematics and science), their attainment was in line with national target levels—that is, they reached the minimum expected level that the majority of children should reach. Ricketts et al. (2014) similarly traced the educational achievement of 15 children with comprehension weaknesses first identified in mid-childhood. While their sample later performed at UK target levels for mathematics and science at age 11, they were less likely to achieve target levels in the reading component of the English assessment. Together, these studies suggest that any educational difficulties experienced at 11 years by children with comprehension weaknesses identified earlier in time are relatively mild and relatively specific to literacy-focused assessments. Clearly, however, these sample sizes are not adequate for properly assessing educational outcomes.

We also need to consider markers of educational attainment beyond 11 years as the educational impact of having poor reading comprehension might become more apparent as young people progress through secondary education, given that learning across a wide range of school subjects becomes increasingly reliant on being able to read and understand complex material independently. We know that there remains substantial variability in reading comprehension through adolescence (Ricketts et al., 2020); potentially, this may extend beyond English and assessments of literacy and serve to constrain performance across the entire curriculum. Indeed, studies of individual differences make clear the high correlations between reading and science achievement (Cromley, 2009; Reed et al., 2017; Zhu, 2022), as well as between measures of text comprehension and mathematical abilities (e.g., Björn et al., 2016; Fuchs et al., 2018). Evidence to assess this hypothesis in children with a history of poor reading comprehension is so far extremely limited. Ricketts et al. (2014) continued to trace their sample through to 16 years of age, the end of compulsory education in the UK. At this time, students sit national assessments known as GCSEs (General Certificate of Secondary Education), comprising a wide range of subjects across the curriculum. Ricketts et al. found that students previously identified as having poor comprehension in primary school were less likely than the national average to attain at least five GCSE pass grades. This is an important observation as it is a key performance benchmark that enables access to further education and training opportunities. As GCSE examinations cover a wide range of academic subjects, Ricketts et al.'s findings suggest that children with comprehension weaknesses may go on to experience educational difficulties across the curriculum. This remains a tentative conclusion at best as GCSE data were available for only 11 of the sample. This group also performed more poorly than a comparison group of 10 skilled comprehenders (initially matched for decoding and nonverbal ability) on a range of GCSE indices, but it is important to note that none of these differences were statistically significant. It is impossible to draw reliable conclusions from these findings given the inadequate sample size. Nevertheless, there is a clear need for further investigations given that children identified with comprehension weaknesses in primary school performed numerically less well than a tightly matched comparison group on national assessments at the end of formal education at 16 years, especially when coupled with their below-average performance against national benchmarks.

The possibility that students with a history of comprehension weaknesses are more likely to leave school without adequate qualifications has long-term implications. In the UK, young people who leave school without five pass grades at GCSE are nearly three times more likely to spend 12 months not in education, employment or training (NEET) in early adulthood (Gadsby, 2019). Being NEET is associated with poor health and wellbeing for the individual (Public Health England, 2014), and presents a significant cost to public spending. To our knowledge, no studies have examined the long-term outcomes of comprehension weaknesses in terms of access to further/higher education and to employment. However, some insight into the importance of comprehension skills for these outcomes can be gained from the wider literature. For example, Kortteinen et al. (2021) found that adults with a history of reading disability (defined by poor decoding) were less likely to pursue further education and more likely to spend a significant time in unemployment. Importantly, those who did not experience such adverse outcomes had higher reading and verbal comprehension abilities, indicating that it may be the comprehension aspect of reading that drives the relationship between reading difficulties and life outcomes. Similarly, Armstrong et al. (2017) found that children with a history of poor vocabulary had a two-fold increase in odds of being NEET at age 21, rising to a three-fold increase if vocabulary problems had persisted. The reading and language difficulties that characterise children with comprehension weaknesses may thus place them at risk of adverse education and employment outcomes.

Research questions

The aim of our study was to examine the education and employment outcomes of students identified as having comprehension weaknesses in mid-childhood. In an earlier study, we used a data-driven approach to identify 947 children with relatively weak reading comprehension skills at age 9 years in the Avon Longitudinal Study of Parents and Children (James et al., 2023). In the current study, across three research questions, we used two sets of comparisons to understand the implications of reading comprehension weaknesses for education and employment outcomes. First and foremost, we compared the outcomes of this group to those of children not identified as having a reading weakness in mid-childhood. These comparisons allowed us to test predictions regarding poorer outcomes for children with comprehension weaknesses, as specified in the hypotheses below. Second, we explored how the outcomes of children with comprehension weaknesses compared to those of the group identified as having word reading weaknesses in mid-childhood. While these were planned comparisons, the specific predictions were less clear; rather, they are provided to benchmark outcomes against reading difficulties that are more widely recognised in the education system.

Question 1) Are children with comprehension weaknesses less likely to meet national assessment targets as they progress through schooling?

We analysed later educational attainment using linked data from the National Pupil Database: a government-maintained and mandatory record of pupil performance in state schools. Our first research question asked whether students with a history of comprehension weaknesses identified in mid-childhood go on to perform below national target levels in English, mathematics, and science in SATs in Year 6 (age 10-11) and Year 9 (age 13-14). As described above, the SATs provide a broad indicator of pupil attainment, primarily serving as a means for monitoring school and pupil performance. They also serve as key predictors of performance in school-leaving assessments (e.g., Rimfeld et al., 2019), demonstrating their importance as an early marker of educational difficulty. We know from past research that reading ability is key to success: one report found that over half of pupils who were not making expected progress in these assessments later met screening criteria for dyslexia (The Dyslexia-SpLD Trust, 2009). The linked data allowed us to test three overarching hypotheses regarding how students with a history of comprehension weaknesses fare in academic assessments relative to their peers. Compared to those without reading weaknesses earlier in time, we predicted that students with comprehension weaknesses would be less likely to meet national attainment targets (H1), have disproportionate difficulties in assessments of English versus mathematics and science (H2), and academic difficulties that become broader and/or more severe in Year 9 (age 13-14) than in Year 6 (age 10-11) (H3). We then explored how these patterns of attainment differed to those of children with word reading weaknesses.

Question 2) How do children with comprehension weaknesses differ from their peers at the point of leaving compulsory education?

Our second research question asked whether children with comprehension weaknesses in mid-childhood go on to leave compulsory education with fewer qualifications according to national statutory assessments. UK pupils work towards a range of GCSE qualifications in Years 10-11 (age 14-16). The number of GCSEs taken is decided by teachers and students, with most young people taking upwards of eight subjects, to include English, mathematics, and science. GCSEs are more formal qualifications than the SATs assessments, with key implications for further education and employment. For example, access to further study post-16 years usually requires a minimum of five passes at Grade A*-C, to include English and mathematics. This criterion is used to assess school performance in government reports, and the grades achieved play an important role in admission to university and applications for employment. Thus, poor GCSE performance leaves a young person vulnerable to being out of education and employment (e.g., Gadsby, 2019; Sadler et al., 2015). The size and nature of the ALSPAC dataset thus allowed us to test two further hypotheses: that students with a history of comprehension weaknesses would be less likely to achieve 5A*-C grades (including English and mathematics) (H4), and would achieve lower grades overall than peers not previously identified as having a reading weaknesses (H5). The additional comparisons to children with word reading weaknesses allowed us to explore school outcomes in relation to different types of reading difficulty.

Question 3) What are the higher education and employment outcomes for children with comprehension weaknesses when they reach age 20?

To complement our analyses on SATs and GCSE attainment, we investigated NEET (Not in Education, Employment or Training) status at age 20. Information about occupational activities has been obtained from ALSPAC participants via regular questionnaires into adulthood. The age 20 questionnaire is an ideal time to assess education and employment status as it follows significant opportunities to engage with university study or further training, or gain employment. The final hypothesis was that children with comprehension weaknesses would be more likely to be NEET at age 20 than individuals who were not identified as having a reading weakness in mid-childhood (H6). We did not have specific predictions about how this risk would compare to children with word reading weaknesses.

Implications

In summary, very little is known about the educational outcomes of children identified as having relatively circumscribed comprehension weaknesses in mid-childhood. While the experimental literature surrounding "poor comprehenders" has revealed possible implications for education, we need a more valid and reliable evidence base as a foundation for more applied work, and to better understand how variation manifests over time and across different outcomes. Evaluating our six hypotheses reveals whether children with comprehension weaknesses are at risk of poor educational outcomes, as assessed by high-stakes national assessments and post-school destinations. If they are at higher risk than children without reading weaknesses, this would be consistent with their difficulties interfering with academic performance, in line with DSM-5 criteria for diagnosing children with Specific Learning Disorder (American Psychiatric Association, 2013). This information would then inform discussions about the need to advance efforts to identify comprehension weaknesses and deliver appropriate intervention. If they are not at significant risk, this might indicate that previous small-scale experimental studies reflect a rare group that should not be considered in the classification of Specific Learning Disorder, aligned with the perspective that the most significant reading problems comprise both word reading and comprehension difficulties. We additionally compared outcomes to those of children with word reading weaknesses for whom more formal recognition and intervention is already available, providing a benchmark for considering allocation of resources and support. By addressing these questions in a large sample using registered hypotheses and analysis plans in advance of exploring the data, this study provides a robust investigation into the functional consequences of often-missed reading comprehension difficulties in childhood.

Methods

Transparency and openness

At Stage 1, we reported how we determined our sample size, the planned data exclusions, and all the pre-specified measures for this set of analyses, which was registered at https://doi.org/10.17605/OSF.IO/TH9QW. The data associated with this paper belong to ALSPAC and thus cannot be made openly available. The ALSPAC website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/), alongside information on how to request access. Our own annotated analysis scripts were available via the associated Open Science Framework page at Stage 1, and output files were added at Stage 2 submission: https://doi.org/10.17605/OSF.IO/ZVJW4. Data were analysed using R version 4.2.3, using *mice* and *miceadds* packages for multiple imputation (Buuren & Groothuis-Oudshoorn, 2011; Robitzsch et al., 2022), *ordinal* and *lme4* for analysis (Bates et al., 2014; Christensen, 2015), and *emmeans* (Lenth et al., 2018) for follow-up contrasts.

Sample

ALSPAC recruited 14,541 pregnant women in the former Avon area (UK) between April 1991 and December 1992, from whom 13,988 offspring were alive at one year. A further 913 eligible children were recruited into the study at age 7, increasing the total sample size to 14,901. The offspring have been studied ever since via a wide range of questionnaires and clinic assessments (Boyd et al., 2013; Fraser et al., 2013; Northstone et al., 2019). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethical and Law Committee at the time. At age 18, study children were sent 'fair processing' materials describing ALSPAC's intended use of their educational records and were given clear means to consent or object via a written form. Data were not extracted for participants who objected, or who were not sent fair processing materials.

Our analysis is based on a subset of participants who completed the Neale Analysis of Reading Ability (NARA-II; Neale, 1997) during a clinic visit at age 9 years (n = 6,935; n = 6,846 following the removal of twin pairs to address non-independence). We previously conducted a latent profile analysis to identify profiles of readers based on measures of reading accuracy, reading comprehension, and listening comprehension (James et al., 2023). The preregistered model did not support the existence of subgroups with specific decoding or comprehension weaknesses. However, by including overall reading ability as a covariate, we extracted profiles with relative strengths and weaknesses in word reading and comprehension domains. The focus of this registered report is the group of children who had weak comprehension skills relative to their reading accuracy skills (n = 947). We compared their educational and employment outcomes to children not identified as having reading difficulties (n = 4,516), as well as the group identified as having word reading weaknesses in the latent profile model (n = 1,383). Table 1 displays the descriptive statistics for these three groups on the variables used in selection: the children with comprehension weaknesses were the poorest performers in the comprehension tasks with relative strengths in word reading; the children with word reading weaknesses were the poorest on word reading with relative strengths in comprehension. For all groups, these weaknesses reflect relative patterns of performance across tasks and span the whole range of reading ability (Table 1, Figure S1). Further demographic information is presented in Table 2.

To further characterise the sample, we also inspected whether there were group differences in the proportion of children identified as having special educational needs (SEN). Using parent questionnaires from age 10 years (Year 6), teacher questionnaires from Year 3 and Year 6, and records from the National Pupil Database (census data from 2003-2005; data from GCSEs at age 16), we documented whether each child was ever recorded as having a statement of special educational needs. This likely underestimates prevalence as not all records from all time points were complete, but some data were available for approximately 95-96% of each group. Of these, 15.42% of children with comprehension weaknesses were recorded as having a statement for support at some point during their education, a slightly higher proportion than the 12.99% of the typically developing readers but less than the children with word reading weaknesses (23.4%). The linked records do not provide further details on the nature of SEN, but we additionally inspected the type of needs reported in the mother's questionnaire at age 10. Children with comprehension weaknesses were more likely than the other two groups to be recognised as needing support for a speech or hearing problem at school, but these numbers were still very low (1.28% and 2.17%, respectively). In contrast, the group with word reading weaknesses showed the highest proportions of recognised learning (9.58%), reading (10.56%), or other (4.88%) needs. A further breakdown of parental and school reported SEN details are available in Supplementary Materials (Table S1, S2).

The authors certify that they had not conducted any analyses on the outcome variables in question prior to Stage 1 acceptance. The national assessment variables were stored in a separate file spanning the whole cohort sample, and had not been filtered for the subgroups of interest.

Outcome measures

National assessments at ages 11 and 14

SATs are administered in Year 6, the final year of primary school when children are 10-11 years old, and again in Year 9 when children are 13-14 years old. At the time of data collection, children were awarded a national curriculum level for each of English, mathematics, and science[2]. These national curriculum levels differ from traditional "grades" in the sense that they reflected broad standards that the majority of children were expected to meet, rather than finely graded levels of attainment. The government expected most children to reach Level 4 at age 10-11, and Levels 5-6 at age 13-14. Sample assessment papers are linked in Appendix S2.

English. The English assessment comprised reading and writing components at each age. At age 11, the reading component required pupils to provide multiple choice and written answers to questions about short passages (45 mins); one writing paper included a short writing task (20 mins) and narrated spelling test, and a second writing paper included a longer writing assessment (45 mins). At age 14, the reading component required shorter and longer written responses to a previously unseen passage (75 mins), as well as an essay on a Shakespeare play (45 mins). The writing component included a shorter writing task (30 mins, also assessed for spelling) and longer writing assessment (45 mins). At each age, a national curriculum level was awarded across both reading and writing components, spanning levels 2-5 at age 11, and 3-7 at age 14.

Mathematics. The mathematics assessment at each time point consisted of two written papers (one with and one without calculators allowed) and a timed mental mathematics test to a recorded audio tape (20 minutes). A national curriculum level was awarded on the basis of marks across all tests, spanning levels 2-5 at age 11 and 2-8 at age 14. Sample questions from across the two time points are provided in Appendix S2.

Science. Children sat two science tests to assess different parts of the curriculum at each age, requiring short written answers and graph drawing. A national curriculum level was awarded across both tests. The tests spanned levels 2-5 at age 11, and 2-7 at age 14.

[2] Note that this grading differs from current SATs in the UK, which now award children a numeric score scaled from 80-120.

National assessments at the end of compulsory schooling (age 16)

The final two years of UK schooling comprise formal GCSE qualifications across a wide range of chosen subjects, with English, mathematics, and science usually considered compulsory subjects. The nature of each assessment varies depending on the subject matter. Each qualification is awarded a grade from A*-G, with grades A*-C considered a 'good' pass[3]. We assessed school leaving achievement in the following two ways, based on routinely used summary statistics.

National target. We assessed whether participants met the criterion of five GCSE (or equivalent) passes at grades A*-C, including English and mathematics. This criterion is often a requirement to progress into further education post 16-years, and is used as a performance benchmark for schools.

Overall attainment. To monitor school performance, the government developed a metric in which each qualification is converted to a points score according to grade (e.g., A* = 58; A = 52, B = 46, etc.), and summed to produce a total points score that reflects individual differences in attainment. A capped points score is calculated from each pupil's top eight grades only, representing the fairest way to capture differences in achievement beyond the number of examinations pupils entered (which may otherwise vary according to opportunity as well as ability). We used this capped points score in our analyses.

Employment outcomes at age 20

Education and employment status was assessed in a questionnaire sent to participants at approximate age 20 years, which was completed by 4,348 participants in the ALSPAC cohort. Participants were asked "*Are you currently in employment or doing any education or training?"*, and provided a Yes/No response. Any group differences can be further described in terms of the proportion reporting full-time education or employment.

[3] Passes at grades A*-C correspond to grades 4-9 on the current grading system introduced in 2014.

Covariates

Given established socio-demographic influences on educational attainment and occupational outcomes, we also considered whether group differences remain once maternal education and free school meal (FSM) eligibility is controlled for. Maternal education data was collected via questionnaire from participants' mothers during pregnancy. The highest reported educational qualification was recoded into three categories to reduce model complexity, using two contrasts to capture variation in maternal educational attainment (Geulayov et al., 2016): CSE/none versus O-level/vocational, CSE/none versus A-level/degree[4].

Eligibility for free school meals is based on household income, and is an effective proxy for socio-economic disadvantage (Gorard, 2012; Ilie et al., 2017). Several studies support its association with educational attainment (e.g., Sammons et al., 2014; Strand, 2014). It is also a readily available metric within schools, facilitating the interpretation of educational risk in wider context (beyond ALSPAC parent-report variables). ALSPAC contains FSM variables collected from the government Pupil Level Annual School Census 2001-2004 (i.e., around the time that the children completed the age 11 assessments), as well as from the database of GCSE results. We combined these measures to create a binary variable of whether children were recorded as being eligible for FSM at any of the available time points.

Planned analyses

All hypotheses were tested within a regression framework, allowing for hierarchical clustering where appropriate and for the examination of possible covariates. We report statistical significance at the conventional $\alpha = .05$; however, with this large sample we note that even small differences are likely to be statistically significant. Thus, our primary focus is on

[4]Note that CSE and O-level are national qualifications that are no longer awarded, having been phased out in the late 1980s. CSE captured lower levels of educational attainment than O-level at 16 years. A-level qualifications are typically taken at 18 years and are the gateway to further and higher education.

reporting and interpreting odds ratios, and using 95% confidence intervals to quantify uncertainty in the estimates.

For each analysis (detailed below), we ran two models: one with the key predictors of interest, and a second model that examined how estimates change once controlling for sex (male = -.5; female = .5), FSM eligibility (no = -.5; yes = .5), and maternal education (CSE/none = -0.33; O-level/vocational = 0.67 (contrast 1); A-level/degree = 0.67 (contrast 2)). The models included the subordinate interactions between these covariates and the predictors where appropriate (i.e., for Q1).

Question 1) Are children with comprehension weaknesses less likely to meet national assessment targets as they progress through schooling?

An ordinal mixed effects regression model was used to examine the performance of children with comprehension weaknesses in national assessments at age 11 and 14, allowing us to capture the nesting of pupils within schools. For each assessment, national curriculum levels were recoded as not meeting age-expected levels (-1), meeting age-expected levels (0), and exceeding age-expected levels (1); with age-expected levels reflecting Level 4 at age 11 and Level 5[5] at age 14. Predictors included group (comprehension weakness vs. typical readers; comprehension weakness vs. word reading weakness), assessment point (age 11 vs. age 14), and subject (English vs. mathematics; English vs. science), alongside all interactions between them. All predictors were simple-coded such that the intercept reflects the grand mean. Given that pupils typically change schools between the two assessment points, but that each of these assessment points is associated with a separate measurement, we used a cross-classification random effects structure to capture school clustering (Goldstein, 2011). Thus, the model incorporated random intercepts for participant ID and school ID.

[5] Although age-expected grades incorporate both Levels 5-6 at age 13-14, we chose to use Level 5 as it reflects the most commonly used standards in government reports. This target may therefore be more lenient than the one used at age 10-11, but any differences will be reflected in the assessment point contrast to facilitate interpretation.

H1 predicted that children with a history of comprehension weaknesses would be less likely to meet expected levels of performance in national assessments at ages 10-11 and 13-14 years, relative to children who were not identified as having a reading weakness. This hypothesis would be supported by a positive effect of group in the typical reader contrast. The comparability of educational outcomes to those of children with word reading weaknesses was provided by the second group contrast.

H2 predicted that the educational disadvantage of children with a history of comprehension weaknesses would be most apparent in English. This hypothesis would be supported if the typical reader group contrast shows an interaction with both subject contrasts. The possibility that only one subject contrast predicts performance will inform us about the patterning of educational weaknesses for children with comprehension difficulties. Any significant interactions were further interpreted by calculating the odds ratio for the effect of group for each subject separately.

H3 predicted that educational disadvantage would become increasingly severe and increasingly broad across development for children with history of comprehension weaknesses. More severe difficulties would be supported by an interaction between the typical reader group contrast and assessment point; increasingly broad difficulties would be supported by three-way interactions between group, assessment point and academic subjects. As above, significant interactions were further interpreted by calculating the odds ratio for the effect of group at each assessment point separately (and for each academic subject, if relevant).

Question 2) How do children with comprehension weaknesses differ from their peers at the point of leaving compulsory education?

Two mixed effects regression models were used to examine the performance of children with comprehension weaknesses in school leaving assessments: whether they met target attainment of 5A*-C grades (binomial outcome: no = 0, yes = 1), and overall attainment (continuous outcome). The main predictor in each analysis was group (comprehension weakness vs. typical readers; comprehension weakness vs. decoding weakness), alongside covariates in the sensitivity analysis. Random intercepts were included for school ID as above, to account for the clustering of pupils within schools.

H4 predicted that children with a history of comprehension weaknesses would be less likely to meet the target of 5A*-C grades including English and mathematics. This hypothesis would be supported by a positive effect of the typical reader group contrast in the binomial model.

H5 predicted that children with a history of comprehension weaknesses would show poorer attainment at the end of compulsory education, as indicated by capped points scores. This hypothesis would be supported by a positive effect of the typical reader group contrast in the continuous outcome model.

As for the analyses in Question 1, the comparability of educational outcomes to those of children with word reading weaknesses were provided by the second group contrast in each model, allowing us to consider the results in the broader context of reading difficulties.

Question 3) What are the higher education and employment outcomes for children with comprehension weaknesses when they reach age 20?

A binomial regression was used to examine whether group differences exist in the likelihood of reporting not being in education or employment (NEET) at age 20. The outcome variable was recoded to reflect NEET (1) versus in education/training (0), with the main predictor of group. No school clustering was accounted for in this model.

H6 predicted that children with a history of comprehension weaknesses would be over-represented in individuals categorised as NEET at age 20. This hypothesis would be supported by a positive effect of the typical reader group contrast, with the second group contrast providing a further comparison to children with a history of word reading weaknesses.

Missingness

For the education variables, we anticipated the majority of missingness to span all assessments. The primary reasons for this are likely: (a) the parent or child having withdrawn consent for data linkage (only $n = 15$ at the time of data linkage); or (b) the child was not in the database as they were educated outside of the English state-maintained system (e.g., they might have attended a state-maintained school outside of England, a private school, or been educated at home). A smaller proportion of data would be missing from individual assessments if the child did not sit a given exam (e.g., due to illness). For the employment variable, we anticipated a much higher proportion of missingness due to general sample attrition and dependence on questionnaire return rates. There was also potential for missingness in the covariates.

Prior to conducting analyses, we inspected the proportion and structure of missing data for each of the analysis variables. We inspected the distribution of missingness according to comprehension group, key covariates (sex, maternal education, free school meal status), alongside other variables that might predict either attrition and/or the values of our key variables (age at KS2 assessment point, private school attendance, ethnicity, reading scores, verbal and nonverbal IQ at age 8, teacher-reported literacy and mathematics ability groups, SEN recognition, parental home ownership, maternal age at birth, maternal marital status, maternal social class, maternal depression score, financial difficulties during first three years of life, deprivation indices, university attendance, NEET at age 21). We report potential biases in missingness to aid in interpretation, and similarly inspected whether each one correlated with the values of key analysis variables themselves.

Missingness was primarily dealt with via multiple imputation. For analyses with education data (Questions 1 and 2), our main analysis excluded participants that did not have a school ID so that school-level clustering could be accounted for in the data. Thus, it is primarily the covariates that were imputed as missingness on educational outcomes largely aligns with missingness in school ID. A supplementary analysis was conducted without school-level clustering, thereby allowing for the imputation of all covariate and outcome variables. The results of these non-clustered models are presented in Supplementary Materials (Appendix S4, S5), and any differences in interpreting the findings discussed in the main text. For Question 3 (which did not require school-level clustering), all analysis variables were imputed. Imputation models contained all analysis variables, alongside the auxiliary variables listed above (although we note one technical limitation in our ability to incorporate the three-way interactions between each socio-demographic covariate and the two within-subjects predictors for Q1; i.e., [variable]*subject*assessment point). If convergence issues arose, we simplified the imputation model according to the following principles: (a) removing variables that have substantial missingness themselves; (b) prioritising variables identified as having strongest relationships with missingness and missing values (as described above); and (c) prioritising variables that predict missing values rather than missingness itself (Spratt et al., 2010). We used $m = 30$ imputations in the first instance, and then added further imputations if the loss of precision (Fraction of Missing Information / m) was > 1% (White et al., 2011; Woods et al., 2021).

Statistical power

The sample size for this study was already determined by the groups extracted from our prior analyses. At this large sample size, we had 90% power to detect very small group differences that may not be educationally meaningful (e.g., $d = 0.12$ for comparison with typically developing readers; $d = 0.14$ for comparison with children with word reading weaknesses). Thus, while we report frequentist statistics and use them as a guide for pursuing interaction effects, our focus is on interpreting the odds ratios and corresponding estimates of uncertainty.

Results

Summary of data availability

Missingness in each of the analytic variables is summarised by reading group in Table S3, and we briefly describe patterns of missingness here. Beginning with the covariates, missingness in maternal education (8.69%) was associated with several maternal report measures pre/post pregnancy (higher maternal depression, financial difficulties, not married, home rental), lower performance in participants' clinic assessments, lower educational attainment, and higher likelihood of FSM status. Given that FSM status is derived from the state school-linked education data, the primary predictor of missingness (16.64%) was private school attendance.

Inspection of the SATs data indicated that data at each time point were more likely to be missing if the child attended private school (14.4%, 25.49%, for Year 6 and Year 9 respectively). Aligned with this, missingness in Year 9 was associated with maternal education and social class, slightly higher reading scores, and higher GCSE attainment. Similar patterns of missingness were observed for GCSE data (17.54% missing).

As anticipated, a much higher proportion of data were missing for NEET outcomes (55.86%). Missingness was associated with being male, SEN status, lower reading and IQ scores, lower educational attainment, as well as maternal factors (education, marriage, home ownership, financial difficulties, depression). These factors align with overall attrition in the ALSPAC dataset, as documented in clinic attendance reports and reported elsewhere (e.g., Spratt et al., 2010; Cornish et al., 2021).

In sum, data were deemed to be missing at random conditional on auxiliary variables included within the imputation models. We used m = 30 imputations for Q1 and 2, and increased to m = 80 for Q3 given high missingness in the NEET variable. The supplementary analyses for Q1 and Q2 without school clustering (and thus incorporating imputed outcomes) are presented in Appendix S4. However, the covariate-adjusted model for Q1 suffered convergence problems that could not be resolved, and we supplement the findings with evidence from an alternative ordinal regression model with cluster-robust standard errors (not preregistered).

Question 1) Are children with comprehension weaknesses less likely to meet national assessment targets as they progress through schooling?

The results of the ordinal mixed effects regression models are presented in Table 3. The analyses supported H1 that children with a history of comprehension weaknesses would be less likely to meet expected levels of performance in SATs assessments: a higher proportion of their assessments were below target (15.11%) than for children without a history of reading weaknesses (9.56%), and a smaller proportion exceeded target grades (44.61% versus 54.06%). Statistically, the odds of typically developing readers achieving at/above national targets were 2.38 times of that of children with comprehension weaknesses (95% CI [1.87, 3.02]), although this difference was smaller once controlling for socio-demographic factors (OR = 1.62, 95% CI [1.1, 2.38]). Children with comprehension weaknesses performed comparably to children with word reading weaknesses overall (OR = 0.93, 95% CI [0.7, 1.23]), but much of this risk was shared with sociodemographic factors and larger group differences were observed once these were controlled for (OR = 0.56, 95% CI [0.35, 0.89]).

The interactions between reader group and curriculum subject indicated that the difficulties experienced by children with comprehension weaknesses (relative to typically developing readers) were not uniform across assessments (English-Maths OR = 1.62, 95% CI [1.33, 1.99]; English-Science OR = 2.29, 95% CI [1.88, 2.79]). However, counter to H2 that they would demonstrate most significant difficulties in English assessments (OR = 1.53, 95% CI [1.18, 2.00]), greater differences were observed in mathematics (OR = 2.49, 95% CI [1.90, 3.26]) and science (OR = 3.52, 95% CI [2.69, 4.69]). Indeed, once adjusting for socio-demographic factors, group differences remained only in mathematics and science assessments (Figure 1; Table S4)[6]. Comparisons to the group with word reading weaknesses also revealed differences across subjects (English-Maths OR = 1.78, 95% CI [1.4, 2.26]; English-Science OR = 3.21, 95% CI [2.54, 4.05]). While children with comprehension weaknesses outperformed children with decoding weaknesses in English (OR = 0.52, 95% CI [0.38, 0.71]), they performed comparably in mathematics (OR = 0.92, 95% CI [0.67, 1.27]), and were less likely to do well in science (OR = 1.66, 95% CI [1.22, 2.27]). Again, the adjusted model demonstrated that much of this risk was shared with socio-demographic factors: children with comprehension weaknesses showed comparable science attainment to those with word reading weaknesses in the adjusted model and outperformed them in the other two subjects (Figure 1; Table S4).

Finally, we found limited support for the hypothesis that the educational attainment gap became more severe and/or broader across development (H3). One significant interaction emerged in comparing children with comprehension weaknesses with typically developing readers, indicating that the group difference in performance between English and Maths widened over time (OR = 1.62, 95% CI [1.08, 2.43]). However, we note that this interaction was not significant in the supplementary analysis with all participants (excluding school clustering; Appendix S4). The two-way interactions between group and assessment point were not significant, nor were any of the other three-way interactions.

[6] Aligned with this role for socio-demographic factors, the interactions between English-Maths and reader group were no longer statistically significant in the covariate-adjusted model in the supplementary analysis (Appendix S4; Table S5).

Question 2) How do children with comprehension weaknesses differ from their peers at the point of leaving compulsory education?

H4 predicted that children with comprehension weaknesses would be less likely to meet the national attainment target of 5A*-C grades including English and Mathematics. In support of this hypothesis, 56.73% of children with comprehension weaknesses achieved this target relative to 63.82% of typically developing readers. The odds of typically developing readers achieving the national target were 33% higher than for children with comprehension weaknesses (OR = 1.33, 95% CI [1.13, 1.58]), which remained comparable after controlling for socio-demographic factors (OR = 1.29, 95% CI [1.08, 1.54]). In contrast, children with comprehension weaknesses showed comparable GCSE outcomes to those who were previously identified as having word reading weaknesses, even once socio-demographic factors were controlled for (Table 4).

A similar pattern of results was observed for GCSE points score. Children with comprehension weaknesses scored an average of 329.28 points (*SD* = 80.23), which was comparable to those with word reading weaknesses (*M* = 328.57, *SD* = 86.14) but lower than that of typically developing readers (*M* = 347.09, *SD* = 77.16). The analyses showed that comprehension weaknesses were associated with a 12.35 [6.79, 17.90] decrease in points score relative to typically developing readers, which remained statistically significant once controlling for socio-demographic factors (10.73 [5.51, 15.95]). The grade to points conversion allows for 6 points per grade bounds (i.e., A = 52, B = 46, C = 40, etc.), meaning that this difference is equivalent to having two qualifications that are one grade lower (or one qualification two grades lower). Again, no significant differences were observed relative to children with word reading weaknesses in these school leaving assessments (Table 5).

Question 3) What are the higher education and employment outcomes for children with comprehension weaknesses when they reach age 20?

Based on the raw data, children with comprehension weaknesses had the highest proportion of respondents reported as being NEET (9.34%), followed by those with word reading weaknesses (7.84%) and those with no reading weaknesses (6.55%). These increased odds were reflected in the analysis, indicating that the odds of children with comprehension weaknesses being NEET at age 20 were 43% higher than those without previous reading difficulties (OR = 0.70; Table 6). However, the 95% confidence intervals were wide [0.50, 0.97], and differences were no longer significant once controlling for socio-demographic factors (OR = 0.75, 95% CI [0.53, 1.08]).

Exploratory analyses (Stage 2): To what extent do relative weaknesses predict outcomes beyond overall reading ability?

Our main research questions and preregistered analyses focused children with uneven reading profiles—namely, those who had comprehension poorer than would be anticipated based on their overall reading ability. However, each group spanned a wide range of ability (Table 1; Figure S1), which likely accounts for the considerable heterogeneity in outcomes. To explore this possibility, we conducted further analyses with overall reading ability as a covariate. This variable was a factor score computed from all four selection measures (item accuracy, passage accuracy, reading comprehension, listening comprehension; Figure S2), and had also been included as a covariate in the latent profile analysis that generated the groups. Thus, these exploratory analyses ask whether uneven profiles of reading skills are related to education and employment outcomes over and above continuous variation in reading ability. We describe overall patterns of performance and where they differ from the main preregistered analyses, with full model tables presented in Appendix S6.

Unsurprisingly, overall reading ability was a significant predictor of outcomes in all analyses, and indeed accounted for some of the variation in group predictors. Yet with this continuous variation controlled for, the risk associated with comprehension weaknesses was *more* striking. In all analyses, the difference in odds (or attainment score) between children with comprehension weaknesses and typically developing readers increased, parcelling out risk associated with specifically weak comprehension versus general reading performance. With this, we also observed a widening gap in performance over time, with larger differences in SATs success at age 14 (OR = 6.15, 95% CI [4.95, 7.65]) than at age 11 (OR = 4.71, 95% CI [3.83, 5.78]). As in the main analyses, these changes were driven by increasing performance gaps in maths and science subjects, and not in English.

A second key difference was observed in relation to the group with word reading weaknesses. Where children with comprehension weaknesses generally showed similar outcomes to this group in the main analyses, they had poorer outcomes once variation in reading ability was accounted for. This likely reflects that reading accuracy measures were more highly loaded on our reading ability measure than comprehension measures (Figure S2), such that when these two groups are matched it is primarily variation in reading comprehension that remains. That children with comprehension weaknesses continued to show poorer outcomes across SATs, GCSEs, and education/employment beyond compulsory schooling emphasises the fundamental importance of comprehension skills beyond reading accuracy.

Discussion

Some children find reading comprehension difficult, despite being able to read accurately and fluently. While this profile of poor reader has been studied extensively in mid-childhood (Cain, 2022), little is known about their academic achievement through secondary school and their outcomes at the end of formal education. Given the centrality of comprehension to learning, this is a surprising gap in knowledge. This registered report was designed to address this gap using data from a longitudinal birth cohort study. We charted educational outcomes in 947 young people previously identified as having reading comprehension weaknesses in mid-childhood (James et al., 2023).

Our primary comparisons were on outcomes for this group relative to those of children without reading difficulties in mid-childhood. Previous small-scale work with 'poor comprehenders' (Cain & Oakhill, 2006, *n* = 17; Ricketts et al*.,* 2014, *n* = 15) indicated that those with a history of poor reading comprehension in primary school were at risk for lower educational achievement in subsequent school years, with some suggestion that their difficulties were relatively specific to outcomes in English rather than science or mathematics (Ricketts et al., 2014). We used SATs data in Year 6 (age 10-11) and Year 9 (age 13-14) to test three hypotheses stemming from these preliminary findings, namely that students with a history of comprehension weaknesses would be less likely to meet national attainment targets than typically-developing peers, have disproportionate difficulties in assessments of English versus mathematics and science, and academic difficulties that become broader and/or more severe in Year 9 than in Year 6. In line with the first hypothesis, typically developing students were 2.38 times more likely to achieve at or above national targets than those with a history of comprehension weaknesses. The second hypothesis was not supported in that those with comprehension weaknesses did more poorly in science and mathematics than English. There was some indication that the group difference widened over time: increasing in mathematics (relative to English) in the main analysis, and in both mathematics and science in exploratory analyses controlling for reading ability. However, this needs to be interpreted cautiously as no interactions with age were significant in the preregistered analyses that controlled for sociodemographic variables (and sociodemographic variables could not be incorporated in exploratory analyses). This guards against a more general conclusion of increasing severity.

Generally, our findings point to stability over time, in line with reading profiles across adolescence more generally (e.g., Ricketts et al., 2020).

Turning to GCSE outcomes at 16 years (national assessments taken at the end of formal schooling in the UK), we found that students with a history of comprehension weaknesses were less likely to meet the national target of 5A*-C grades (including English and Mathematics). Typically developing readers were 33% more likely to reach this target, and they achieved higher grades overall. These findings replicate those described by Ricketts et al. (2014) for 11 poor comprehenders and extend them across a much larger sample. Finally, data on NEET status at age 20 years allowed us to document how many young people were not in education, employment, or training at that time. Those with a history of comprehension weaknesses were over-represented in this group, being 43% more likely to be classed as NEET than those without a history of reading difficulty. Overall, our findings demonstrate that children with weak comprehension in mid-childhood underperform through secondary school relative to peers and are less likely to achieve national academic targets, with long-term societal consequences in early adulthood.

While statistical significance is not surprising given the large sample, the effect sizes are educationally meaningful; indeed, outcomes for those with comprehension weaknesses were generally similar to those with a history of poor word reading, a group known to be at significant risk for poor academic outcomes (Smart et al., 2017; Willcutt et al., 2007). Yet, those in the weak comprehension group were less likely than the poor readers to have ever been categorised as having special educational needs (15.42% vs. 23.4%). This finding fits with previous studies noting that the difficulties with language and literacy experienced by children identified as ‘poor comprehenders’ tend to be unnoticed in the classroom (e.g., Nation et al., 2004). This is unfortunate, given the poor educational outcomes experienced by some young people in this group. Having summarised our findings, we next consider two themes for discussion and future investigation: variation between individuals and difficulties across domains.

Individual variation in outcomes

First, despite the overall risk for poor attainment, our findings also show considerable variation: clearly some students with a history of comprehension weaknesses have good academic outcomes, meeting or exceeding national targets, and fare well in education or employment in early adulthood. Part of this variation reflects that this group has *relatively* weak comprehension across a wide span of reading ability. Importantly, however, controlling for this served to highlight greater risk associated with unexpectedly poor comprehension, both for poor educational outcomes and for NEET status. Another source of variation is likely to be associated with social disadvantage. Children with a history of weak comprehension were more likely to be eligible for free school meals and had lower levels of maternal education than typical readers and those in the poor word reading group. Controlling for these factors served to reduce group differences in educational outcomes but not eliminate them, and group differences in GCSE achievement were less impacted by social disadvantage.

In contrast to the educational results, increased risk for NEET status was entirely accounted for by socio-demographic factors in the main analysis. Tentatively, this could suggest shared risk factors that extend beyond those at play in educational settings, such as the social or financial support needed to continue in education or training. NEET status might also index more severe outcomes than the education variables, and it is perhaps less well-captured by our reading profiles alone as these incorporated a wide range of overall reading ability in each case. In line with this, when this variation was accounted for in exploratory analyses, comprehension weaknesses held as a risk factor even with socio-demographic variation controlled, consistent with poor comprehension being a risk factor for adverse outcomes.

Socio-demographic risk is itself complex to understand, possibly reflecting genetic influences on language and literacy as well as systemic social and environmental influences (e.g., Hart et al., 2021). There is a clear need to better understand heterogeneity, and having charted outcomes in this registered report, future work should look to model concurrent strengths and weaknesses across domains (language, literacy, cognitive ability and beyond) in interaction with family and neighbourhood factors. It will be important to determine how these extend beyond comprehension risk from a single timepoint in mid-childhood, using datasets that capture language and literacy in later development.

Difficulties across educational domains

The second theme concerns academic attainment across different curriculum subjects. Following previous findings with a small group of poor comprehenders (Ricketts et al., 2014), we had predicted that students with a history of comprehension weaknesses would be at greater risk for underachieving in English, rather than science or mathematics. This was not the case. While there was underachievement across the board relative to typically developing peers, group differences were more marked in science and mathematics. Interestingly, this finding aligns with the effect sizes described by Cain and Oakhill (2006) with smaller group differences reported for English (d = .78) than for mathematics (d = 1.05) and science (d = 1.25), but these were not statistically compared. A similar pattern is reflected in comparing the children with comprehension weaknesses in our study to those with word reading weaknesses: they outperformed them in English, underachieved in science, and obtained comparable results in mathematics.

Our analyses do not permit causal inferences, but we can speculate about this pattern of findings. As noted in the Introduction, individual differences in reading comprehension are associated with a range of cognitive and linguistic factors including working memory and other executive skills, inference making, problem solving, and integrating background knowledge as a mental model in constructed from text (e.g., Cain, 2022; van den Broek & Kendeou, 2022). It makes sense to consider these factors as constraints to progress in other academic domains too, in part explaining the high correlation between reading and achievement in mathematics and science seen in large datasets including international data such as the Programme for International Study Assessment (PISA; e.g., Cromley, 2009; Zhu, 2022). Using twin data, Harlaar et al. (2012) found that both phenotypic and genetic correlations between mathematics and reading comprehension were higher than those between mathematics and word decoding. They speculated that this genetic overlap might reflect the shared contribution of executive function processes that are implicated in mathematics as well as comprehension.

Alternative and not mutually exclusive possibilities relate to the more direct constraints that comprehension weaknesses might place on performance in science and mathematics. The ability to reason about and understand scientific and mathematical topics has been linked with reading ability (e.g., Björn et al., 2016; Reed et al., 2017), and there is some evidence showing that 'poor comprehenders' perform less well than typical readers on tests of mathematical reasoning (Pimperton & Nation, 2010). In addition, science subjects have their own disciplinary-specific vocabulary, including low frequency words and morphologically complex Latinate forms (Fang, 2006), posing additional challenges for children with language and literacy weaknesses. While empirical work is needed to tease out causal relations, our findings demonstrate that children with poor comprehension in mid-childhood are at risk for poorer education outcomes across the curriculum. It appears that the presence of relative strengths in word reading (cf. poor comprehension) does not seem to mitigate this risk, at least in science and mathematics. For English assessments, young people with a history of poor word reading fared less well than those with relative weaknesses in comprehension, perhaps because of the demands that reading extended text place on reading accuracy and fluency. It may be that the breadth of difficulty across subjects is what leaves children with comprehension weaknesses more vulnerable to NEET outcomes in adulthood relative to children with word reading weaknesses.

Strengths and limitations

Before reflecting on the educational implications of our study, it is important to note its strengths and limitations. It is the first large study to focus on the later education and employment outcomes of children who experienced relative reading comprehension difficulties in mid-childhood, with comprehension weaknesses defined by a data-driven approach (James et al., 2023). Previous explorations have been unacceptably small, but here the ALSPAC dataset provided a large sample of young people with a history of comprehension weaknesses, and data linkage with the National Pupil Database including achievement in national examinations at the end of formal schooling. The registered report format allowed us to preregister our hypotheses and analysis plans in detail in advance. This methodological transparency provides reassurance on the strength and utility of the findings we report. Turning to limitations, a key limitation is that our markers of reading weaknesses come from assessments administered ages 8 and 9 years, with no further indication of how reading difficulties developed or remediated into adolescence and adulthood. Second, while we accounted for school clustering in our analyses, we lack further detail on the school and home environment factors that might help in understanding the variation observed in outcomes. Finally, we note the limited nature of our exploration into higher education and employment analyses: the NEET self-report measure had high levels of missingness, and lacked further detail on the nature of education and employment in adulthood.

Implications for educational support

Despite these limitations, this study indicates that while children with reading comprehension weaknesses have comparably poor outcomes as the children with word reading weaknesses, they are not as readily recognised as needing additional support. It remains possible that children with word reading difficulties in this sample already had more educational support that mitigated more severe attainment difficulties. Indeed, since this cohort were in compulsory education, the UK curriculum has emphasised phonics teaching programmes and screeners to identify struggling word readers. However, our findings underscore the need to also attend to aspects of reading comprehension *beyond* word reading, and to consider children with comprehension weaknesses in the classroom.

Studies of traditionally defined poor comprehenders highlight poor language as an early risk factor (e.g., Catts et al., 2006; Nation et al., 2004, 2010). The introduction of programmes that target oral language (e.g., West et al., 2021) should help identify those in need of additional support at school entry and may help to improve outcomes for those at risk of comprehension difficulties. For children who reach secondary school with comprehension problems, our findings indicate that support across the curriculum might be needed. This will require greater awareness of the nature and complexity of comprehension and the consequences of comprehension weaknesses (e.g., Catts, 2022), and indeed the understanding that some young people find comprehension difficult, despite being able to read words. An immediate research priority is thus to unpack the nature of educational risk for this group, using a wide range of cognitive measures and fine-grained analyses of educational assessments. With this evidence, educators will be better positioned to mitigate risk for poor outcomes through classroom and small group interventions.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*.

Armstrong, R., Arnott, W., Copland, D. A., McMahon, K., Khan, A., Najman, J. M., & Scott, J. G. (2017). Change in receptive vocabulary from childhood to adulthood: Associated mental health, education and employment outcomes: Associated outcomes of receptive vocabulary change. *International Journal of Language & Communication Disorders*, *52*(5), 561–572. https://doi.org/10.1111/1460-6984.12301

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4* (arXiv:1406.5823). arXiv. https://doi.org/10.48550/arXiv.1406.5823

Björn, P. M., Aunola, K., & Nurmi, J.-E. (2016). Primary school text comprehension predicts mathematical word problem-solving skills in secondary school. *Educational Psychology*, *36*(2), 362–377. https://doi.org/10.1080/01443410.2014.992392

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., & Davey Smith, G. (2013). Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, *42*(1), 111–127. https://doi.org/10.1093/ije/dys064

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*, 1–67. https://doi.org/10.18637/jss.v045.i03

Cain, K. (2022). Children's Reading Comprehension Difficulties. In *The Science of Reading: A Handbook* (Second, pp. 298–322). Wiley-Blackwell.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*(4), 683–696. https://doi.org/10.1348/000709905X67610

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Catts, H. W. (2022). Rethinking How to Promote Reading Comprehension. *American Educator*, *45*(4), 26.

Catts, H. W., Adlof, S., & Ellis Weismer, S. (2006). Language Deficits in Poor Comprehenders: A Case for the Simple View of Reading. *Journal of Speech, Language, and Hearing Research : JSLHR*, *49*, 278–293. https://doi.org/10.1044/1092-4388(2006/023)

Catts, H. W., & Kamhi, A. G. (2014). Prologue: Reading Comprehension Is Not a Single Ability. *Language, Speech, and Hearing Services in Schools*, *45*(3), 73–76. https://doi.org/10.1044/2017_LSHSS-16-0033

Christensen, R. H. B. (2015). *Analysis of ordinal data with cumulative link models—Estimation with the R-package ordinal*.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties: A Randomized Controlled Trial. *Psychological Science*, *21*(8), 1106–1116. https://doi.org/10.1177/0956797610375449

Cromley, J. G. (2009). Reading Achievement and Science Proficiency: International Comparisons From the Programme on International Student Assessment. *Reading Psychology*, *30*(2), 89–118. https://doi.org/10.1080/02702710802274903

Elwér, Å., Gustafson, S., Byrne, B., Olson, R. K., Keenan, J. M., & Samuelsson, S. (2015). A retrospective longitudinal study of cognitive and language skills in poor reading comprehension. *Scandinavian Journal of Psychology*, *56*(2), 157–166. https://doi.org/10.1111/sjop.12188

Fang, Z. (2006). The Language Demands of Science Reading in Middle School. *International Journal of Science Education*, *28*(5), 491–520. https://doi.org/10.1080/09500690500339092

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Smith, G. D., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M., & Lawlor, D. A. (2013). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, *42*(1), 97–110. https://doi.org/10.1093/ije/dys066

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & N. Martin, B. (2018). Text Comprehension and Oral Language as Predictors of Word-Problem Solving: Insights into Word-Problem Solving as a Form of Text Comprehension. *Scientific Studies of Reading*, *22*(2), 152–166. https://doi.org/10.1080/10888438.2017.1398259

Gadsby, B. (2019). *Research Briefing 6: The long-term NEET population* (6). https://www.impetus.org.uk/assets/publications/Youth-Jobs-Gap-The-Long-Term-NEET-Population.pdf

Geulayov, G., Metcalfe, C., & Gunnell, D. (2016). Parental suicide attempt and offspring educational attainment during adolescence in the Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort. *Psychological Medicine*, *46*(10), 2097–2107. https://doi.org/10.1017/S0033291716000556

Goldstein, H. (2011). *Multilevel Statistical Models* (4th Edition). Wiley.

Gorard, S. (2012). Who is eligible for free school meals? Characterising free school meals as a measure of disadvantage in England. *British Educational Research Journal*, *38*(6), 1003–1017. https://doi.org/10.1080/01411926.2011.608118

Harlaar, N., Kovas, Y., Dale, P. S., Petrill, S. A., & Plomin, R. (2012). Mathematics is differentially related to reading comprehension and word decoding: Evidence from a genetically sensitive design. *Journal of Educational Psychology*, *104*(3), 622–635. https://doi.org/10.1037/a0027646

Hart, S. A., Little, C., & van Bergen, E. (2021). Nurture might be nature: Cautionary tales and proposed solutions. *Npj Science of Learning*, *6*(1), Article 1. https://doi.org/10.1038/s41539-020-00079-z

Hayiou-Thomas, M. E., Smith-Woolley, E., & Dale, P. S. (2021). Breadth versus depth: Cumulative risk model and continuous measure prediction of poor language and reading outcomes at 12. *Developmental Science*, *24*(1), e12998. https://doi.org/10.1111/desc.12998

Ilie, S., Sutherland, A., & Vignoles, A. (2017). Revisiting free school meal eligibility as a proxy for pupil socio-economic deprivation. *British Educational Research Journal*, *43*(2), 253–274. https://doi.org/10.1002/berj.3260

James, E., Thompson, P. A., Bowes, L., & Nation, K. (2023). Heterogeneity in children's reading comprehension difficulties: A latent class approach. *JCPP Advances*, *3*(4), e12177. https://doi.org/10.1002/jcv2.12177

Kintsch, W. (1986). Learning From Text. *Cognition and Instruction*, *3*(2), 87–108. https://doi.org/10.1207/s1532690xci0302_1

Kortteinen, H., Eklund, K., Eloranta, A.-K., & Aro, T. (2021). Cognitive and non-cognitive factors in educational and occupational outcomes—Specific to reading disability? *Dyslexia*, *27*(2), 204–223. https://doi.org/10.1002/dys.1673

Landi, N., & Ryherd, K. (2017). Understanding specific reading comprehension deficit: A review. *Language and Linguistics Compass*, *11*(2), e12234. https://doi.org/10.1111/lnc3.12234

Language and Reading Research Consortium (LARRC), Jiang, H., & Logan, J. (2019). Improving Reading Comprehension in the Primary Grades: Mediated Effects of a Language-Focused Classroom Intervention. *Journal of Speech, Language, and Hearing Research*, *62*(8), 2812–2828. https://doi.org/10.1044/2019_JSLHR-L-19-0015

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). *emmeans: Estimated marginal means, AKA least-squares means* [Computer software].

McMaster, K. L., van den Broek, P., Espin, C. A., Pinto, V., Janda, B., Lam, E., Hsu, H.-C., Jung, P.-G., Leinen, A. B., & van Boekel, M. (2015). Developing a reading comprehension intervention: Translating cognitive theory to educational practice. *Contemporary Educational Psychology*, *40*, 28–40. https://doi.org/10.1016/j.cedpsych.2014.04.001

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden Language Impairments in Children. *Journal of Speech, Language, and Hearing Research*, *47*(1), 199–211. https://doi.org/10.1044/1092-4388(2004/017)

Nation, K., Cocksey, J., Taylor, J. S. H., & Bishop, D. V. M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*, *51*(9), 1031–1039. https://doi.org/10.1111/j.1469-7610.2010.02254.x

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, *67*(3), 359–370. https://doi.org/10.1111/j.2044-8279.1997.tb01250.x

Neale, M. D. (1997). *Neale Analysis of Reading Ability—Second Revised British Edition*. NFER-Nelson.

Northstone, K., Lewcock, M., Groom, A., Boyd, A., Macleod, J., Timpson, N., & Wells, N. (2019). The Avon Longitudinal Study of Parents and Children (ALSPAC): An update on the enrolled sample of index children in 2019. *Wellcome Open Research*, *4*, 51. https://doi.org/10.12688/wellcomeopenres.15132.1

Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading*, *7*(3), 289–307. https://doi.org/10.1207/S1532799XSSR0703_6

Pearson, P. D., Palincsar, A., Biancarosa, G., & Berman, A. (2020). *Reaping the Rewards of the Reading for Understanding Initiative*. National Academy of Education. https://doi.org/10.31094/2020/2

Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Pimperton, Hannah., & Nation, Kate. (2010). Understanding words, understanding numbers: An exploration of the mathematical profiles of poor comprehenders. *British Journal of Educational Psychology*, *80*(2), 255–268. https://doi.org/10.1348/000709909X477251

Public Health England. (2014). *Review3_NEETs_health_inequalities.pdf* (Health Equity Evidence Review 3). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/356062/Review3_NEETs_health_inequalities.pdf

Reed, D. K., Petscher, Y., & Truckenmiller, A. J. (2017). The Contribution of General Reading Ability to Science Achievement. *Reading Research Quarterly*, *52*(2), 253–266. https://doi.org/10.1002/rrq.158

Ricketts, J., Lervåg, A., Dawson, N., Taylor, L. A., & Hulme, C. (2020). Reading and Oral Vocabulary Development in Early Adolescence. *Scientific Studies of Reading*, *24*(5), 380–396. https://doi.org/10.1080/10888438.2019.1689244

Ricketts, J., Sperring, R., & Nation, K. (2014). Educational attainment in poor comprehenders. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00445

Rimfeld, K., Malanchini, M., Hannigan, L. J., Dale, P. S., Allen, R., Hart, S. A., & Plomin, R. (2019). Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *Journal of Child Psychology and Psychiatry*, *60*(12), 1278–1288. https://doi.org/10.1111/jcpp.13070

Robitzsch, A., Grund, S., & Henke, T. (2022). *Some Additional Imputation Functions, Especially for mice* [Computer software]. https://cran.hafro.is/web/packages/miceadds/miceadds.pdf

Sadler, K., Akister, J., & Burch, S. (2015). Who are the young people who are not in education, employment or training? An application of the risk factors to a rural area in the UK. *International Social Work*, *58*(4), 508–520. https://doi.org/10.1177/0020872813515010

Sammons, P., Sylva, K., Melhuish, E., Siraj, I., Taggart, B., Toth, K., & Smees, R. (2014). Influences on students' GCSE attainment and progress at age 16 Effective Pre-School, Primary & Secondary Education Project (EPPSE). *DfE Research Brief*.

Smart, D., Youssef, G. J., Sanson, A., Prior, M., Toumbourou, J. W., & Olsson, C. A. (2017). Consequences of childhood reading difficulties and behaviour problems for educational achievement and employment in early adulthood. *British Journal of Educational Psychology*, *87*(2), 288–308. https://doi.org/10.1111/bjep.12150

Strand, S. (2014). 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

The Dyslexia-SpLD Trust. (2009). *No to Failure*.

Torppa, M., Vasalampi, K., Eklund, K., Sulkunen, S., & Niemi, P. (2020). Reading comprehension difficulty is often distinct from difficulty in reading fluency and accompanied with problems in motivation and school well-being. *Educational Psychology*, *40*(1), 62–81. https://doi.org/10.1080/01443410.2019.1670334

van den Broek, P., & Kendeou, P. (2022). Reading Comprehension I. In *The Science of Reading* (pp. 239–260). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119705116.ch11

West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan, H., & Hulme, C. (2021). Early language screening and intervention can be delivered successfully at scale: Evidence from a cluster randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *62*(12), 1425–1434. https://doi.org/10.1111/jcpp.13415

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. https://doi.org/10.1002/sim.4067

Willcutt, E. G., Betjemann, R. S., Pennington, B. F., Olson, R. K., DeFries, J. C., & Wadsworth, S. J. (2007). Longitudinal Study of Reading Disability and Attention-Deficit/Hyperactivity Disorder: Implications for Education. *Mind, Brain, and Education*, *1*(4), 181–192. https://doi.org/10.1111/j.1751-228X.2007.00019.x

Woods, A. D., Davis-Kean, P., Halvorson, M. A., King, K., Logan, J., Xu, M., Bainter, S., Brown, D., Clay, J. M., Cruz, R. A., Elsherif, M., Gerasimova, D., Joyal-Desmarais, K., Moreau, D., Nissen, J., Schmidt, K., Uzdavines, A., Dusen, B. V., & Vasilev, M. R. (2021). *Missing Data and Multiple Imputation Decision Tree*. PsyArXiv. https://doi.org/10.31234/osf.io/mdw5r

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology*, *80*(3), 351–361. https://doi.org/10.1111/j.2044-8295.1989.tb02325.x

Zhu, Y. (2022). Reading matters more than mathematics in science learning: An analysis of the relationship between student achievement in reading, mathematics, and science. *International Journal of Science Education*, *44*(1), 1–17. https://doi.org/10.1080/09500693.2021.2007552

Table 1

Sample-standardised descriptive statistics for the decoding and comprehension measures used in identifying reading profiles.

	Comprehension weakness (*n* = 947)			Word reading weakness (*n* = 1383)			No reading weakness (*n* = 4516)		
	M	*(SD)*	range	*M*	*(SD)*	range	*M*	*(SD)*	range
Item accuracy[a]	99.68	(15.63)	59-124	87.61	(15.40)	58-117	103.85	(12.42)	59-124
NARA passage accuracy[b]	105.87	(14.24)	57-124	95.56	(17.27)	54-125	100.13	(13.93)	51-125
NARA passage comprehension[b]	92.47	(12.43)	52-118	96.86	(16.19)	54-135	102.53	(14.39)	52-137
WOLD listening comprehension[c]	94.96	(14.09)	57-143	101.85	(14.61)	57-158	100.53	(15.06)	55-159

Note. Scores reflect performance age-standardised on the sample entered into the latent profile analysis (*n* = 6,846). (a) The item accuracy measure comprises real word and nonword reading scores (Nunes et al., 2003), with test-retest reliabilities of 0.8 and 0.73 respectively; (b) The NARA-II manual reports inter-form reliabilities of .84-.92 for accuracy and .87-.92 for comprehension; (c) The Wechsler Objective Language Dimensions reports listening comprehension test-retest reliabilities between 0.83-0.88 in children aged 6-11 years.

Table 2

Demographics for each group of readers

	Comprehension weakness (*n* = 947)	Word reading weakness (*n* = 1383)	No reading weakness (*n* = 4516)
Sex: Female	519 (54.80%)	646 (46.71%)	2325 (51.56%)
Ethnicity: Non-white	40 (4.65%)	42 (3.35%)	164 (4.07%)
>1 main language (incl. English)	4 (0.01%)	2 (<0.01%)	13 (<0.01%)
Main language not English	2 (<0.01%)	0 (0%)	3 (<0.01%)
Record of SEN statement	140 (15.42%)	307 (23.44%)	558 (12.99%)
SEN recognition reported by school in age 10-11 questionnaire	108 (19.22%)	248 (30.5%)	442 (17.62%)
Free school meal eligibility	89 (11.12%)	107 (9.49%)	328 (8.68%)
Maternal education:			
CSE/none	148 (16.89%)	201 (15.81%)	524 (12.77%)
vocational/O-level	404 (46.12%)	550 (43.27%)	1774 (43.23%)
A-level/degree	324 (36.99%)	520 (40.91%)	1806 (44.01%)

Note. Percentages reflect proportion of the group once missing data are excluded.

Table 3

Reading profile differences in meeting SATs target attainment levels according to subject and assessment point

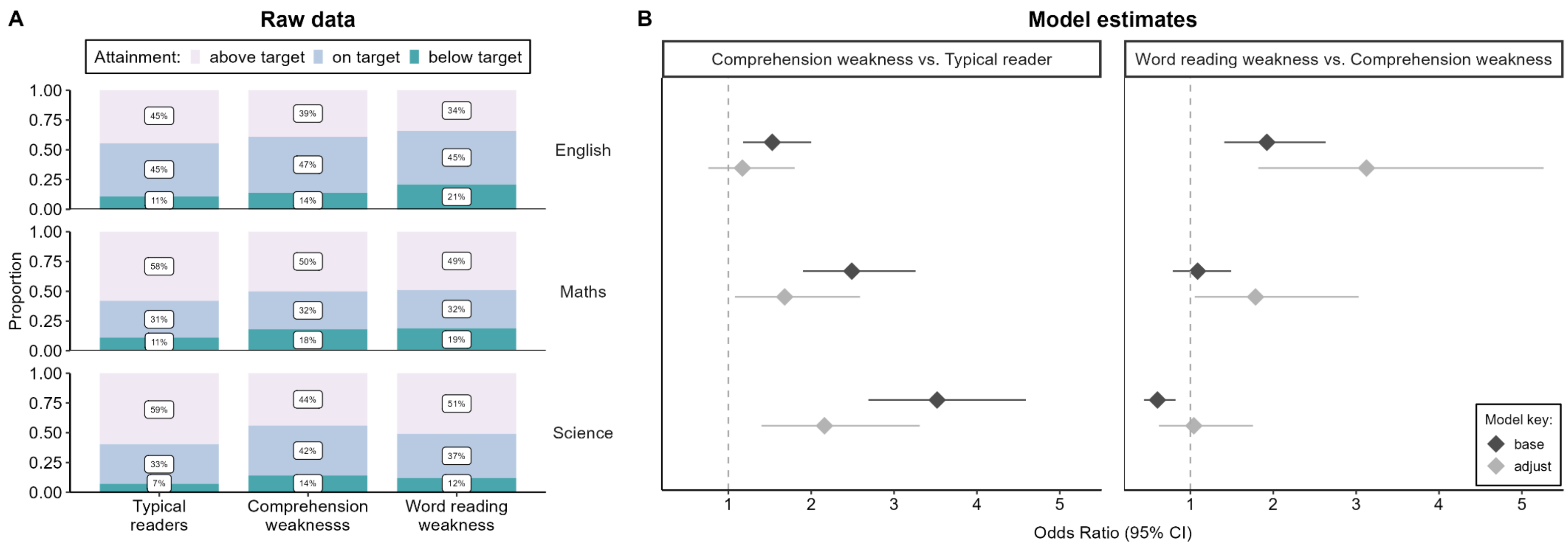
	Base model					Adjusted model				
	B	*SE*	*t*	*p*	*OR [95% CI]*	*B*	*SE*	*t*	*p*	*OR [95% CI]*
-1 | 0	-4.26	0.08	-53.42	<.001	0.01 [0.01, 0.02]	-3.33	0.10	-33.48	<.001	0.04 [0.03, 0.04]
0 | 1	0.11	0.07	1.64	.101	1.12 [0.98, 1.29]	1.21	0.10	12.67	<.001	3.37 [2.79, 4.06]
CW-DW	-0.08	0.14	-0.53	.594	0.93 [0.70, 1.23]	**-0.58**	**0.24**	**-2.44**	**.015**	**0.56 [0.35, 0.89]**
CW-TD	**0.87**	**0.12**	**7.06**	**<.001**	**2.38 [1.87, 3.02]**	**0.48**	**0.20**	**2.43**	**.015**	**1.62 [1.10, 2.38]**
Year (6-9)	**0.93**	**0.09**	**10.35**	**<.001**	**2.55 [2.13, 3.04]**	**0.67**	**0.10**	**6.94**	**<.001**	**1.96 [1.62, 2.37]**
Eng-Maths	**0.85**	**0.04**	**19.56**	**<.001**	**2.33 [2.14, 2.53]**	**0.89**	**0.07**	**11.98**	**<.001**	**2.43 [2.10, 2.82]**
Eng-Sci	**0.94**	**0.04**	**22.04**	**<.001**	**2.57 [2.36, 2.79]**	**1.10**	**0.07**	**14.99**	**<.001**	**3.02 [2.61, 3.49]**
CW-DW*Year	0.05	0.10	0.47	.637	1.05 [0.86, 1.29]	0.01	0.18	0.07	.944	1.01 [0.72, 1.43]
CW-TD*Year	0.11	0.09	1.27	.204	1.12 [0.94, 1.33]	0.10	0.14	0.68	.494	1.10 [0.83, 1.47]
CW-DW*Eng-Maths	**0.58**	**0.12**	**4.74**	**<.001**	**1.78 [1.40, 2.26]**	**0.56**	**0.21**	**2.69**	**.007**	**1.74 [1.16, 2.61]**
CW-TD*Eng-Maths	**0.48**	**0.10**	**4.70**	**<.001**	**1.62 [1.33, 1.99]**	**0.36**	**0.17**	**2.09**	**.036**	**1.43 [1.02, 2.01]**
CW-DW*Eng-Sci	**1.17**	**0.12**	**9.81**	**<.001**	**3.21 [2.54, 4.05]**	**1.09**	**0.20**	**5.33**	**<.001**	**2.98 [1.99, 4.45]**

CW-TD*Eng-Sci	**0.83**	**0.1**	**8.29**	**<.001**	**2.29**	**[1.88,**	**2.79]**	**0.61**	**0.17**	**3.63**	**<.001**	**1.84**	**[1.33,**	**2.56]**
Year*Eng-Maths	**1.66**	**0.09**	**19.09**	**<.001**	**5.24**	**[4.42,**	**6.21]**	**1.65**	**0.13**	**12.51**	**<.001**	**5.22**	**[4.03,**	**6.76]**
Year*Eng-Sci	**-1.13**	**0.08**	**-13.35**	**<.001**	**0.32**	**[0.27,**	**0.38]**	**-1.25**	**0.13**	**-9.64**	**<.001**	**0.29**	**[0.22,**	**0.37]**
CW-DW*Year*Eng-Maths	0.29	0.24	1.20	.230	1.34	[0.83,	2.16]	0.27	0.25	1.07	.283	1.31	[0.80,	2.14]
CW-TD*Year*Eng-Maths	**0.48**	**0.21**	**2.34**	**.019**	**1.62**	**[1.08,**	**2.43]**	**0.51**	**0.21**	**2.39**	**.017**	**1.66**	**[1.09,**	**2.52]**
CW-DW*Year*Eng-Sci	-0.04	0.24	-0.19	.850	0.96	[0.60,	1.52]	-0.12	0.24	-0.50	.617	0.89	[0.55,	1.42]
CW-TD*Year*Eng-Sci	0.2	0.2	1.02	.306	1.23	[0.83,	1.82]	0.20	0.20	0.98	.328	1.22	[0.82,	1.82]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Eng = English; Sci = Science; Maths = Mathematics. The adjusted model controls for sex, maternal education, and free school meal status, including in interaction with all main effects and subordinate interactions. Bold font is used to denote significant predictors ($p < .05$).

Figure 1

Group differences in SATs attainment for English, mathematics, and science subjects



Note. (A) The proportion of each reading profile attaining below, at, and above target levels for each subject, collapsed across Year 6 and Year 9 assessments. (B) Baseline and adjusted odds ratios for attaining at/above target for each subject. Odds ratios for the contrast between children with comprehension and word reading weaknesses (right panel) are presented as the inverse to ease comparison of effect sizes.

Table 4

Reading profile differences in meeting the national attainment target of 5 A-C GCSE qualifications (including English & Mathematics)*

	Base model					Adjusted model				
	B	*SE*	*t*	*p*	*OR [95% CI]*	*B*	*SE*	*t*	*p*	*OR [95% CI]*
(Intercept)	0.60	0.1	5.78	<.001	1.83 [1.49, 2.24]	-0.22	0.11	-2.02	.043	0.81 [0.65, 0.99]
CW vs. DW	-0.06	0.1	-0.64	.524	0.94 [0.77, 1.14]	-0.07	0.11	-0.67	.506	0.93 [0.76, 1.15]
CW vs. TD	**0.29**	**0.09**	**3.34**	**.001**	**1.33 [1.13, 1.58]**	**0.26**	**0.09**	**2.84**	**.004**	**1.29 [1.08, 1.54]**
Sex						0.38	0.06	6.02	<.001	1.46 [1.29, 1.66]
Maternal ed. 1						0.6	0.09	6.3	<.001	1.81 [1.51, 2.18]
Maternal ed. 2						1.41	0.1	13.58	<.001	4.10 [3.34, 5.02]
FSM eligibility						-1.27	0.11	-11.05	<.001	0.28 [0.23, 0.35]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest.

Table 5

Reading profile differences in capped GCSE points scores

	Base model						Adjusted model					
	B	*SE*	*t*	*p*	*[95% CI]*		*B*	*SE*	*t*	*p*	*[95% CI]*	
(Intercept)	345.11	3.91	88.28	<.001	[337.44,	352.77]	310	3.7	83.82	<.001	[302.75,	317.25]
CW vs. DW	-2.74	3.36	-0.81	.415	[-9.34,	3.85]	-2.38	3.17	-0.75	.454	[-8.59,	3.84]
CW vs. TD	**12.35**	**2.83**	**4.36**	**<.001**	**[6.79,**	**17.9]**	**10.73**	**2.66**	**4.03**	**<.001**	**[5.51,**	**15.95]**
Sex							16.86	1.86	9.08	<.001	[13.22,	20.51]
Maternal ed. 1							25.38	2.96	8.57	<.001	[19.57,	31.19]
Maternal ed. 2							54.11	3.18	17.02	<.001	[47.87,	60.35]
FSM eligibility							-54.62	3.35	-16.32	<.001	[-61.18,	-48.06]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest.

Table 6

Reading profile differences in self-reporting as NEET at age 20

	Base model						Adjusted model					
	B	*SE*	*t*	*p*	*OR*	*[95% CI]*	*B*	*SE*	*t*	*p*	*OR*	*[95% CI]*
(Intercept)	-1.96	0.07	-26.91	<.001	0.14	[0.12, 0.16]	-1.42	0.09	-15.34	<.001	0.24	[0.20, 0.29]
CW vs. DW	-0.2	0.21	-0.98	.330	0.82	[0.54, 1.23]	-0.19	0.23	-0.83	.408	0.83	[0.53, 1.30]
CW vs. TD	**-0.36**	**0.17**	**-2.13**	**.034**	**0.70**	**[0.5, 0.97]**	-0.28	0.18	-1.57	.119	0.75	[0.53, 1.08]
Sex							-0.39	0.14	-2.75	.007	0.68	[0.52, 0.90]
Maternal ed. 1							-0.52	0.19	-2.76	.007	0.59	[0.41, 0.86]
Maternal ed. 2							-0.99	0.21	-4.69	<.001	0.37	[0.24, 0.56]
FSM eligibility							1.41	0.19	7.48	<.001	4.10	[2.82, 5.95]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest.

Supplementary materials for

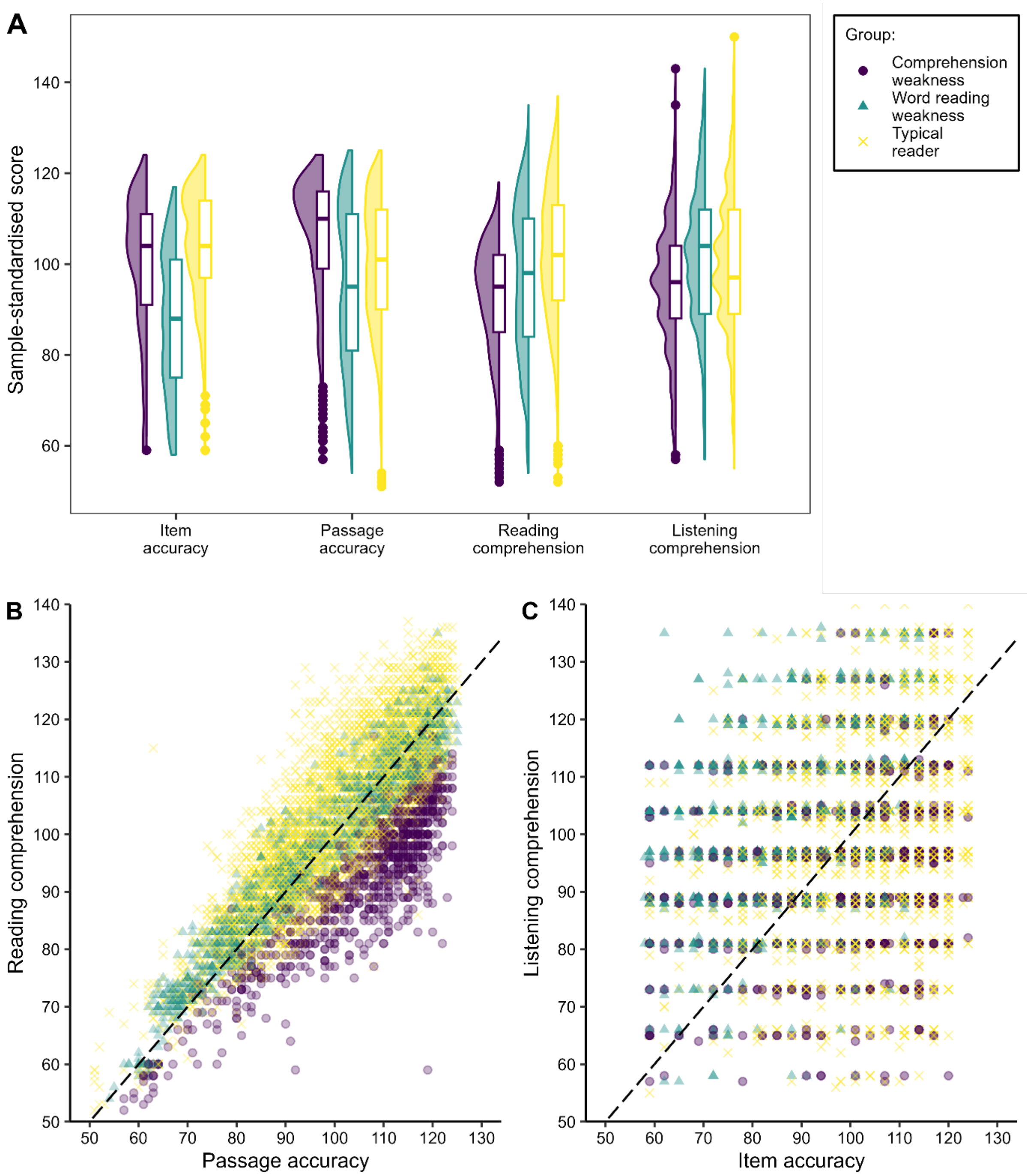
What are the long-term prospects for children with comprehension weaknesses? A registered report investigating education and employment outcomes

Contents

Appendix S1: Sample descriptors

Figure S1

Sample-standardised scores for each reading profile on each selection measure



Note. (A) Distribution of scores on the two word reading and two comprehension measures included in the latent profile analysis; (B) Distribution of scores on the NARA measures of passage accuracy and comprehension; (C) Distribution of scores on the item accuracy (word and nonword reading) and WOLD listening comprehension measure. The dashed lines represent equivalent performance on the word reading and comprehension measures.

Table S1

Parent-reported needs for special arrangements at school at age 10

	Comprehension weakness (82.58%)[a]	Word reading weakness (81.49%)[a]	No reading weakness (82.06%)[a]
Overall needs	11.25	18.9	11.31
Emotional or behavioural problem	1.53	1.86	1.73
Eyesight problem	0.77	1.24	0.76
Hearing problem	2.17	1.69	1.16
Learning difficulty	5.63	9.58	4.21
Physical problem	0.77	1.15	0.78
Reading difficulty	3.71	10.56	3
Speech problem	1.28	0.89	0.51
Other	2.17	4.88	3.53

Note. a) Proportion of subgroup with data available

Table S2

School report of current special educational needs in Year 6 (age 10-11 years)

	Comprehension weakness (59.34%)[a]	Word reading weakness (58.50%)[a]	No reading weakness (55.47%)[a]
Developmental delay	1.43	1.11	0.48
Emotional and behavioural difficulties	2.67	5.18	3.47
Learning difficulties	9.79	18.29	7.7
Medical conditions	3.03	2.09	1.12
Physical disabilities	1.07	1.36	0.68
Specific learning difficulties	3.92	8.92	4.15
Speech and language difficulties	3.39	1.97	1.28
Sensory impairment (hearing)	0.71	0.62	0.44
Sensory impairment (visual)	0.36	0.25	0.52
Other	8.39	14.57	8.87

Note. a) Proportion of subgroup with data available

Appendix S2: UK SATs assessments

Children in the ALSPAC cohort span three academic years. Year 6 SATs were completed by members of the cohort in the 2001/2002, 2002/2003, and 2003/2004 academic years. Year 9 SATs were completed by members of the cohort in the 2004/2005, 2005/2006, and 2006/2007 academic years.

Some relevant past SATs papers can be found on websites, such as https://www.testbase.co.uk/past-papers/ and https://www.sats-papers.co.uk/.

Of particular interest to readers may be the reading and language comprehension demands of the mathematics assessments. Sample first/last questions from Year 6 and Year 9 SATs papers are provided for reference below.

KS2 Mathematics (Year 6) – Paper 1 sample questions

From Qualifications and Curriculum Authority, 2003.

(1) Write in the missing numbers.

37 x ___ = 111

225 - ___ = 150

___ ÷ 4 = 21

(24) 30 children are going on a trip. It costs £5 including lunch. Some children take their own packed lunch. They pay only £3. The 30 children pay a total of £110. How many children are taking their own packed lunch?

KS2 Mathematics (Year 6) – Mental mathematics sample questions

From Qualifications and Curriculum Authority, 2003.

(1) "Write in figures the number one thousand and twenty."

(20) "Look at your answer sheet. What is the angle between the hands of a clock at four o'clock?"

KS3 Mathematics (Year 9) – Paper 1 sample questions

From Qualifications and Curriculum Authority, 2003.

(1) On each spinner right five numbers to make the statements correct.

It is certain that you will get a number less than 6.

It is more likely that you will get an even number than an odd number.

It is impossible that you will get a multiple of three.

(21) The scatter graph shows 15 pupils' coursework and test marks. To find a pupil's total mark, you add the coursework mark to the test mark.

Which pupil had the highest mark?

Look at the statement below. Tick True or False. *The range of coursework marks was greater than the range of test marks.* Explain your answer.

Pupils with total marks in the shaded region on the graph win a prize. What is the smallest total mark needed to win a prize?

KS3 Mathematics (Year 9) – Mental mathematics sample questions

From Qualifications and Curriculum Authority, 2006.

(1) "A teacher divides a class of thirty pupils into six equal groups. How many pupils are in each group?"

(30) "*m* squared equals one hundred. Write down the two possible values of *m* plus fifteen."

Appendix S3: Supplementary information for main analysis

Table S3

Data availability (n; %) in covariate and outcome variables for each reading group

	Comprehension weakness (*n* = 947)	Word reading weakness (*n* = 1383)	No reading weakness (*n* = 4516)
Covariates			
Sex	947 (100%)	1382 (99.93%)	4505 (99.76%)
Maternal education	876 (92.50%)	1271 (91.90%)	4104 (90.88%)
Free school meal status	800 (84.48%)	1127 (81.49%)	3780 (83.70%)
Outcome variables			
SATs (Year 6)[a]	807 (85.22%)	1171 (84.67%)	3882 (85.96%)
SATs (Year 9)[a]	722 (76.24%)	1007 (72.81%)	3372 (74.67%)
GCSEs[a]	794 (83.84%)	1113 (80.48%)	3738 (82.77%)
NEET status	396 (41.82%)	625 (45.19%)	2001 (44.31%)

Note. (a) Data availability for educational outcomes is reported based on the availability of a school ID, as this was necessary for inclusion in the main analysis which included school clustering.

Table S4

Odds ratios for group comparisons of SATs performance in each subject

	Base model			Adjusted model		
	B	*SE*	*OR [95% CI]*	*B*	*SE*	*OR [95% CI]*
CW vs. DW						
English	0.66	0.16	0.52 [0.38, 0.71]	1.13	0.27	0.32 [0.19, 0.55]
Maths	0.08	0.16	0.92 [0.67, 1.27]	0.58	0.27	0.56 [0.33, 0.95]
Science	-0.51	0.16	1.66 [1.22, 2.27]	0.04	0.26	0.96 [0.57, 1.61]
CW vs. TD						
English	-0.43	0.13	1.53 [1.18, 2.00]	-0.16	0.22	1.17 [0.76, 1.80]
Maths	-0.91	0.14	2.49 [1.90, 3.26]	-0.52	0.22	1.68 [1.08, 2.59]
Science	-1.26	0.14	3.52 [2.69, 4.59]	-0.77	0.22	2.16 [1.40, 3.31]
DW vs. TD						
English	-1.09	0.12	2.96 [2.36, 3.72]	-1.29	0.21	3.62 [2.42, 5.43]
Maths	-0.99	0.12	2.70 [2.14, 3.41]	-1.09	0.20	2.98 [2.00, 4.44]
Science	-0.75	0.12	2.11 [1.68, 2.67]	-0.81	0.20	2.24 [1.51, 3.32]

Appendix S4: Question 1 (SATs) without school clustering

Our primary analytical approach excluded the majority of children with missing data on the outcome variables, as they did not have an associated School ID to allow for clustering. To complement this analysis, our preregistered approach at Stage 1 was to conduct a supplementary analysis without school as random effect, thereby allowing all predictor and outcome variables to be imputed for all participants.

We preregistered the same mixed effects regression approach, including only participant ID as a random effect. However, when analysing the SATs data, the full model with covariates suffered convergence issues that could not be resolved, and no alternative was preregistered at Stage 1. For transparency, we present the base model using the initial mixed effects approach in Table S5 (as per Stage 1 plans). We then present both the base and covariate-adjusted models in Table S6, using a standard cumulative link regression model with cluster-robust standard errors to address for non-independence (not preregistered at Stage 1).

Table S5

Reading profile differences in meeting SATs target attainment levels according to subject and assessment point (supplementary analysis without school clustering, using a mixed effects model)

	Base model					
	B	*SE*	*t*	*p*	*OR [95% CI]*	
-1 | 0	-4.05	0.06	-63.8	<.001	0.02	[0.02, 0.02]
0 | 1	0.26	0.05	5.11	<.001	1.30	[1.18, 1.44]
CW-DW	0.00	0.14	0.02	.981	1.00	[0.76, 1.33]
CW-TD	**0.94**	**0.12**	**7.68**	**<.001**	**2.55**	**[2.01, 3.25]**
Year (6-9)	**0.57**	**0.03**	**17.35**	**<.001**	**1.78**	**[1.66, 1.89]**
Eng-Maths	**0.61**	**0.04**	**14.76**	**<.001**	**1.84**	**[1.70, 2.00]**
Eng-Sci	**0.77**	**0.04**	**18.35**	**<.001**	**2.17**	**[2.00, 2.36]**
CW-DW*Year	0.05	0.09	0.51	.607	1.05	[0.87, 1.26]
CW-TD*Year	0.10	0.08	1.25	.213	1.10	[0.94, 1.29]
CW-DW*Eng-Maths	**0.41**	**0.11**	**3.58**	**<.001**	**1.50**	**[1.20, 1.88]**
CW-TD*Eng-Maths	**0.35**	**0.09**	**3.70**	**<.001**	**1.42**	**[1.18, 1.71]**
CW-DW*Eng-Sci	**0.90**	**0.11**	**7.94**	**<.001**	**2.46**	**[1.97, 3.07]**
CW-TD*Eng-Sci	**0.66**	**0.09**	**6.94**	**<.001**	**1.93**	**[1.60, 2.32]**
Year*Eng-Maths	**1.19**	**0.08**	**14.22**	**<.001**	**3.29**	**[2.79, 3.88]**
Year*Eng-Sci	**-0.99**	**0.08**	**-12.47**	**<.001**	**0.37**	**[0.32, 0.43]**
CW-DW*Year*Eng-Maths	0.10	0.23	0.43	.664	1.11	[0.70, 1.75]
CW-TD*Year*Eng-Maths	0.28	0.20	1.44	.151	1.33	[0.90, 1.96]†
CW-DW*Year*Eng-Sci	-0.14	0.22	-0.63	.530	0.87	[0.57, 1.34]
CW-TD*Year*Eng-Sci	0.08	0.19	0.41	.682	1.08	[0.75, 1.57]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Eng = English; Sci = Science; Maths = Mathematics. Parameters were estimated separately in $m = 35$ imputations (increased from initial 30 to ensure loss of precision <.01), and pooled using Rubin's rules. The corresponding model adjusted for covariates did not converge. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest. †marks predictors that differ in statistical significance relative to the primary analyses.

Table S6

Reading profile differences in meeting SATs target attainment levels according to subject and assessment point (supplementary analysis without school clustering, using cluster-corrected standard errors)

	Base model				Adjusted model			
	B	*SE*	*p*	*OR [95% CI]*	*B*	*SE*	*p*	*OR [95% CI]*
-1 | 0	-1.95	0.03	<.001	0.14 [0.13, 0.15]	-1.56	0.05	<.001	0.21 [0.19, 0.23]
0 | 1	0.08	0.02	.001	1.08 [1.04, 1.14]	0.66	0.05	<.001	1.93 [1.77, 2.12]
CW-DW	0.01	0.07	.908	1.01 [0.88, 1.16]	**-0.28**	**0.13**	**.032**	**0.76 [0.58, 0.98]**
CW-TD	**0.43**	**0.06**	**<.001**	**1.54 [1.38, 1.73]**	**0.25**	**0.11**	**.019**	**1.28 [1.04, 1.58]**
Year (6-9)	**0.34**	**0.02**	**<.001**	**1.40 [1.35, 1.45]**	**0.20**	**0.03**	**<.001**	**1.22 [1.15, 1.30]**
Eng-Maths	**0.35**	**0.02**	**<.001**	**1.42 [1.35, 1.48]**	**0.35**	**0.04**	**<.001**	**1.42 [1.30, 1.54]**
Eng-Sci	**0.37**	**0.02**	**<.001**	**1.45 [1.39, 1.51]**	**0.44**	**0.04**	**<.001**	**1.55 [1.45, 1.68]**
CW-DW*Year	0.04	0.05	.449	1.04 [0.94, 1.14]	-0.01	0.09	.874	0.99 [0.83, 1.17]
CW-TD*Year	0.06	0.04	.168	1.06 [0.98, 1.15]	0.05	0.07	.445	1.05 [0.92, 1.21]
CW-DW*Eng-Maths	**0.20**	**0.06**	**.002**	**1.22 [1.07, 1.38]**	0.23	0.12	.056	1.26 [0.99, 1.60]†
CW-TD*Eng-Maths	**0.16**	**0.05**	**.003**	**1.17 [1.06, 1.31]**	0.15	0.09	.104	1.16 [0.97, 1.40]†
CW-DW*Eng-Sci	**0.43**	**0.06**	**<.001**	**1.54 [1.36, 1.73]**	**0.44**	**0.11**	**<.001**	**1.55 [1.26, 1.92]**
CW-TD*Eng-Sci	**0.32**	**0.05**	**<.001**	**1.38 [1.25, 1.51]**	**0.26**	**0.08**	**.002**	**1.30 [1.09, 1.52]**
Year*Eng-Maths	**0.63**	**0.04**	**<.001**	**1.88 [1.75, 2.03]**	**0.65**	**0.06**	**<.001**	**1.92 [1.70, 2.14]**

Year*Eng-Sci	**-0.42**	**0.03**	**<.001**	**0.66**	**[0.61,**	**0.70]**	**-0.52**	**0.05**	**<.001**	**0.59**	**[0.53,**	**0.66]**
CW-DW*Year*Eng-Maths	0.10	0.11	.350	1.11	[0.90,	1.36]	0.10	0.12	.391	1.11	[0.88,	1.39]
CW-TD*Year*Eng-Maths	0.15	0.09	.089	1.16	[0.98,	1.38]†	0.17	0.10	.083	1.19	[0.98,	1.42] †
CW-DW*Year*Eng-Sci	-0.02	0.10	.835	0.98	[0.80,	1.20]	-0.06	0.11	.597	0.94	[0.76,	1.17]
CW-TD*Year*Eng-Sci	0.05	0.08	.581	1.05	[0.89,	1.23]	0.04	0.09	.697	1.04	[0.87,	1.25]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Eng = English; Sci = Science; Maths = Mathematics. The adjusted model controls for sex, maternal education, and free school meal status, including in interaction with all main effects and subordinate interactions. Parameters were estimated separately in $m = 55$ imputations (increased from initial 30 to ensure loss of precision <.01), and pooled using Rubin's rules. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest; †marks predictors that differ in statistical significance relative to the primary analyses.

Appendix S5: Question 2 (GCSE) results without school clustering

The GCSE outcomes were analysed as planned, using mixed effects models.

Table S7

Reading profile differences in meeting the national attainment target of 5 A-C GCSE qualifications (supplementary analysis without school clustering)*

	Base model							Adjusted model						
	B	*SE*	*t*	*p*	*OR [95% CI]*			*B*	*SE*	*t*	*p*	*OR [95% CI]*		
(Intercept)	0.34	0.03	10.7	<.001	1.40	[1.32,	1.49]	-0.34	0.06	-5.89	<.001	0.71	[0.63,	0.80]
CW vs. DW	-0.03	0.09	-0.36	.716	0.97	[0.81,	1.15]	-0.06	0.10	-0.63	.531	0.94	[0.78,	1.14]
CW vs. TD	**0.31**	**0.08**	**4.12**	**<.001**	**1.37**	**[1.18,**	**1.58]**	**0.24**	**0.08**	**2.95**	**.003**	**1.27**	**[1.08,**	**1.49]**
Sex								0.50	0.06	8.62	<.001	1.64	[1.47,	1.84]
Maternal ed. 1								0.70	0.09	8.17	<.001	2.02	[1.70,	2.39]
Maternal ed. 2								1.50	0.09	16.68	<.001	4.49	[3.76,	5.35]
FSM eligibility								-1.30	0.11	-12.00	<.001	0.27	[0.22,	0.34]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Parameters were estimated separately in $m = 80$ imputations (increased from initial 30 to ensure loss of precision <.01 in the NEET analysis), and pooled using Rubin's rules. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest. All hypothesis tests aligned with the findings of the primary analysis.

Table S8

Reading profile differences in capped GCSE points scores (supplementary analysis without school clustering)

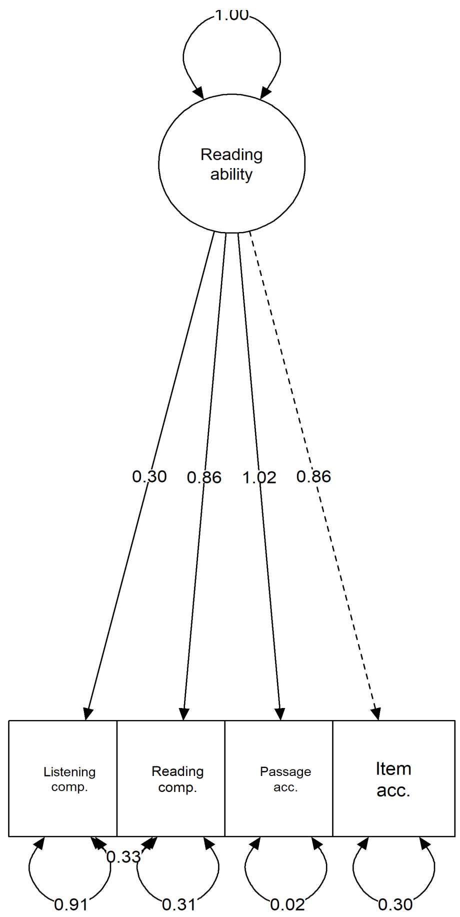
	Base model						Adjusted model					
	B	*SE*	*t*	*p*	*[95% CI]*		*B*	*SE*	*t*	*p*	*[95% CI]*	
(Intercept)	335.47	1.32	254.01	<.001	[332.88,	338.06]	300.78	1.93	156.19	<.001	[297.00,	304.56]
CW vs. DW	1.73	3.61	0.48	.633	[-5.36,	8.82]	0.28	3.27	0.08	.933	[-6.14,	6.69]
CW vs. TD	**19.6**	**3.06**	**6.41**	**<.001**	**[13.60,**	**25.59]**	**14.32**	**2.77**	**5.16**	**<.001**	**[8.88,**	**19.76]**
Sex							19.9	1.88	10.56	<.001	[16.20,	23.59]
Maternal ed. 1							32.01	3.14	10.21	<.001	[25.86,	38.16]
Maternal ed. 2							71.38	3.15	22.68	<.001	[65.20,	77.55]
FSM eligibility							-65.26	3.61	-18.07	<.001	[-72.35,	-58.17]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Parameters were estimated separately in $m = 80$ imputations (increased from initial 30 to ensure loss of precision <.01 in the NEET analysis), and pooled using Rubin's rules. Bold font is used to denote significant predictors ($p < .05$) that correspond to hypotheses of interest. All hypothesis tests aligned with the findings of the primary analysis.

Appendix S6: Exploratory analyses including reading ability as a covariate

Figure S2

Task loadings on the overall reading ability factor



Note. The model also include covariates for age at test, not depicted.

Table S9

Reading profile differences in meeting SATs target attainment levels according to subject and assessment point, controlling for overall reading ability

	Base model				
	B	*SE*	*t*	*p*	*OR [95% CI]*
-1 | 0	-4.24	0.07	-61.37	<.001	0.01 [0.01, 0.02]
0 | 1	0.21	0.06	3.57	<.001	1.23 [1.10, 1.37]
CW-DW	1.59	0.12	13.60	<.001	4.88 [3.88, 6.13] †
CW-TD	1.68	0.10	17.37	<.001	5.36 [4.44, 6.48]
Year (6-9)	1.03	0.08	12.38	<.001	2.80 [2.38, 3.30]
Eng-Maths	0.90	0.05	18.40	<.001	2.45 [2.23, 2.70]
Eng-Sci	0.92	0.05	19.76	<.001	2.51 [2.29, 2.75]
CW-DW*Year	0.46	0.12	3.89	<.001	1.58 [1.25, 1.98] †
CW-TD*Year	0.27	0.09	2.85	.004	1.30 [1.09, 1.57] †
CW-DW*Eng-Maths	0.66	0.14	4.83	<.001	1.93 [1.48, 2.51]
CW-TD*Eng-Maths	0.44	0.11	4.00	<.001	1.55 [1.25, 1.92]
CW-DW*Eng-Sci	1.00	0.13	7.60	<.001	2.72 [2.10, 3.52]
CW-TD*Eng-Sci	0.71	0.11	6.66	<.001	2.04 [1.65, 2.51]
Year*Eng-Maths	1.79	0.09	19.50	<.001	5.97 [4.99, 7.14]
Year*Eng-Sci	-1.04	0.09	-12.08	<.001	0.35 [0.30, 0.42]
CW-DW*Year*Eng-Maths	0.81	0.26	3.07	.002	2.24 [1.34, 3.75] †
CW-TD*Year*Eng-Maths	0.67	0.21	3.14	.002	1.96 [1.29, 2.99]
CW-DW*Year*Eng-Sci	0.47	0.25	1.89	.059	1.60 [0.98, 2.61]
CW-TD*Year*Eng-Sci	0.52	0.21	2.52	.012	1.68 [1.12, 2.51] †

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Eng = English; Sci = Science; Maths = Mathematics. This model controls for variation in overall reading ability (not included in table; full output available on the OSF). Parameters were estimated separately in $m = 30$ imputations (re-run from main analysis to include overall reading ability) and pooled using Rubin's rules. The corresponding model adjusted for sociodemographic covariates did not converge. †marks predictors that differ in direction and/or statistical significance relative to the preregistered analyses without controlling for overall reading ability.

Table S10

Reading profile differences in meeting the national attainment target of 5 A-C GCSE qualifications, controlling for overall reading ability*

	Base model					Adjusted model				
	B	*SE*	*t*	*p*	*OR [95% CI]*	*B*	*SE*	*t*	*p*	*OR [95% CI]*
(Intercept)	0.57	0.11	5.10	<.001	1.76 [1.42, 2.19]	-0.09	0.12	-0.74	0.461	0.91 [0.72, 1.16]
Reading ability	0.31	0.01	28.89	<.001	1.36 [1.34, 1.39]	0.29	0.01	26.53	<.001	1.34 [1.31, 1.37]
CW vs. DW	0.89	0.12	7.35	<.001	2.44 [1.93, 3.10] †	0.84	0.12	6.70	<.001	2.31 [1.81, 2.95] †
CW vs. TD	0.87	0.10	8.84	<.001	2.40 [1.97, 2.91]	0.81	0.10	8.00	<.001	2.26 [1.85, 2.75]
Sex						0.37	0.07	5.25	<.001	1.44 [1.26, 1.66]
Maternal ed. 1						0.42	0.10	4.04	<.001	1.52 [1.24, 1.87]
Maternal ed. 2						1.05	0.12	9.09	<.001	2.84 [2.27, 3.56]
FSM eligibility						-1.12	0.13	-8.86	<.001	0.33 [0.25, 0.42]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Parameters were estimated separately in $m = 30$ imputations (re-run from main analysis to include overall reading ability), and pooled using Rubin's rules. †marks predictors that differ in direction and/or statistical significance relative to the preregistered analyses without controlling for overall reading ability.

Table S11

Reading profile differences in capped GCSE points scores, controlling for differences in reading ability

	Base model						Adjusted model					
	B	*SE*	*t*	*p*	*[95% CI]*		*B*	*SE*	*t*	*p*	*[95% CI]*	
(Intercept)	340.22	3.15	107.92	<.001	334.04	346.40	315.55	3.15	100.03	<.001	[309.37,	321.74]
Reading ability	10.16	0.22	45.37	<.001	9.72	10.60	9.07	0.22	40.75	<.001	[8.64,	9.51]
CW vs. DW	24.99	2.96	8.45	<.001	19.19	30.79†	22.46	2.85	7.89	<.001	[16.88,	28.04] †
CW vs. TD	27.07	2.46	10.99	<.001	22.25	31.90	24.41	2.37	10.31	<.001	[19.77,	29.06]
Sex							14.4	1.64	8.78	<.001	[11.19,	17.62]
Maternal ed. 1							16	2.56	6.24	<.001	[10.97,	21.03]
Maternal ed. 2							34.37	2.8	12.29	<.001	[28.88,	39.86]
FSM eligibility							-41.9	2.94	-14.23	<.001	[-47.68,	-36.13]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Parameters were estimated separately in m = 30 (re-run from main analysis to include overall reading ability), and pooled using Rubin's rules. †marks predictors that differ in direction and/or statistical significance relative to the preregistered analyses without controlling for overall reading ability.

Table S12

Reading profile differences in self-reporting as NEET at age 20, controlling for differences in reading ability

	Base model					Adjusted model				
	B	*SE*	*t*	*p*	*OR [95% CI]*	*B*	*SE*	*t*	*p*	*OR [95% CI]*
(Intercept)	-2.02	0.08	-25.92	<.001	0.13 [0.11, 0.15]	-1.49	0.10	-15.28	<.001	0.22 [0.19, 0.27]
Reading ability	-0.12	0.02	-6.65	<.001	0.89 [0.86, 0.92]	-0.07	0.02	-3.73	<.001	0.93 [0.90, 0.97]
CW vs. DW	-0.55	0.21	-2.56	.011	0.58 [0.38, 0.88] †	-0.42	0.22	-1.86	.064	0.66 [0.42, 1.03]
CW vs. TD	-0.53	0.18	-2.98	.003	0.59 [0.41, 0.83]	-0.41	0.19	-2.21	.028	0.66 [0.46, 0.96] †
Sex						-0.42	0.13	-3.21	.002	0.66 [0.51, 0.85]
Maternal ed. 1						-0.45	0.18	-2.47	.015	0.64 [0.45, 0.91]
Maternal ed. 2						-0.80	0.21	-3.84	<.001	0.45 [0.30, 0.68]
FSM eligibility						1.34	0.17	7.67	<.001	3.82 [2.70, 5.39]

Note. CW = children with comprehension weaknesses; DW = children with decoding weaknesses; TD = typically developing children without reading weaknesses; Maternal ed. 1 = CSE/no qualification vs. O-level/vocational; Maternal ed. 2 = CSE/no qualification vs. A-level/degree; FSM = Free School Meal. Parameters were estimated separately in $m = 80$ (re-run from main analysis to include overall reading ability), and pooled using Rubin's rules. †marks predictors that differ in direction and/or statistical significance relative to the preregistered analyses without controlling for overall reading ability.