Ambition Institute



Evidence-based deployment of teaching assistants: a systematic review of interventions in the UK

Briony Banks, Clare Routledge, Jennifer Curran, Sam Sims

2025

Contents

Contributions	**3**
Acknowledgements	**4**
Abstract	**5**
Introduction	**6**
The role of teaching assistants	7
Support for pupils with SEND	7
The impact of teaching assistant support	8
The present study: aims and scope	10
Methods	**11**
Literature search	11
Screening	12
Data extraction and analysis	13
Risk of bias and limitations analysis	14
Results	**15**
Interventions overview	15
Evidence of intervention effectiveness	20
What knowledge and skills are interventions likely to improve?	30
Which interventions are effective for pupils with SEND?	32
Principles for implementing an intervention effectively	36
Implications for schools and school leaders	37
Discussion	**38**
Limitations	40
References	**42**
Appendix	**49**

Contributions

Conceptualisation

Sam Sims
Briony Banks

Methodology

Briony Banks
Clare Routledge
Sam Sims
Lucy Newman

Literature searches and screening

Clare Routledge
Jennifer Curran
Briony Banks
Sam Sims

Data extraction

Clare Routledge
Jennifer Curran
Briony Banks

Formal analysis

Briony Banks

Writing – original draft

Briony Banks
Clare Routledge

Writing – review and editing

Jennifer Curran
Sam Sims
Sarah Scott
Rebecca Curtis

Data visualisation

Briony Banks

Supervision

Briony Banks
Sam Sims
Rebecca Curtis

Project administration

Clare Routledge
Briony Banks
Rebecca Curtis

Funding

UBS
Ambition Institute

Acknowledgements

Thank you to UBS for their support in funding this Ambition Institute research project. Thank you to Lucy Newman and Sarah Scott at Ambition Institute for their valuable contributions, advice and guidance. Thank you to Emma Lark, Nick Rose, Kay Rimmington-Singh and Lisa Heys at Ambition Institute for helpful comments. Particular thanks to Dr Rob Webster, University of Greenwich, for reviewing this paper.

Abstract

Guidance from the UK Education Endowment Foundation (EEF, 2025) recommends deploying teaching assistants (TAs) to deliver evidence-based interventions. However, there is currently a lack of clear guidance on which TA-led interventions are effective. This systematic review and meta-analysis examines the evidence around one-to-one and small-group interventions that can be delivered by TAs in the UK. We searched for randomised controlled trials (RCTs) of interventions delivered by TAs which aimed to improve academic outcomes in any subject area, for pupils aged four to 18. From these searches, we identified 17 interventions that are currently available to schools in the UK. We reviewed their effectiveness and the quality of the evidence, and also conducted meta-analyses to obtain an average effect size for each intervention. We found that the quality and strength of evidence vary greatly, and only a limited number of interventions have reliable, consistent evidence of a positive effect on pupil outcomes. We present summaries of the evidence for each intervention and discuss the findings in relation to specific learning needs. We also provide a simple summary of effective interventions to help school leaders select the most appropriate intervention for their pupils' learning needs and context.

Keywords: teaching assistant, intervention, systematic review, meta-analysis

Introduction

Teaching assistants (TAs)[1] represent a substantial segment of the education workforce, playing an important role in supporting teaching and learning. Many countries have seen an increasing reliance on TAs in recent decades (Navarro, 2015). This is particularly true in the UK; in Northern Ireland, the number of classroom assistants increased by 63% between 2018 and 2024 (O'Connor et al., 2024). In England, TAs comprise around 30% of school staff in England (Department for Education, 2024a), and in Wales there are currently more registered learning support workers than registered teachers (Education Workforce Council, 2024).

Despite the increasing reliance on TAs in schools, evidence of the impact on pupils' attainment is mixed, and seems to depend on how TAs are deployed (Alborz et al., 2009; Farrell et al., 2010). For example, the large-scale, longitudinal Deployment and Impact of Support Staff (DISS) project in the UK found that TA support can actually have a negative impact on pupils' academic outcomes (Blatchford et al., 2009). The importance of improved training for TAs has also been widely recognised. For example, in a survey of 2,531 respondents in Scotland, access to training for support staff was highlighted as a key need by support staff themselves, and by senior leaders (Education Scotland, 2023).

These findings have informed a series of practical publications for school leaders on how to improve their deployment of TAs (for example, EEF, 2018, 2025; Webster et al., 2021). Such guidance is particularly important given the central role that TAs play in supporting pupils with the lowest attainment and with special educational needs and disabilities (SEND). Although better deployment of TAs may positively impact teachers' workload, there is a lack of evidence on how effective these recommendations are at improving pupils' academic outcomes. For example, a trial of the Maximising the Impact of Teaching Assistants (MITA) programme found that although pupil engagement improved, their academic (reading) outcomes did not (Dimova et al., 2021). One promising way to more effectively deploy TAs is for them to deliver structured, evidence-based interventions. Indeed, this is one of the recommendations in recent Education Endowment Foundation guidance for schools (EEF, 2025). The guidance also recommends that school leaders engage with the evidence base before selecting an intervention. However, this can be a complex task for school leaders. There has so far been no systematic review of the effectiveness of TA-led interventions, and it is unclear which are the most effective, particularly for pupils with SEND. This research therefore systematically reviews the evidence of interventions that can be delivered by TAs in the UK. Our goal is to improve the impact of TAs on pupil outcomes, by helping schools to choose evidence-based interventions appropriate for their pupils' learning needs and context.

1 There are many different terms for the role of teaching assistant (see 'The role of TAs') and the terminology differs between countries. We use the term 'TA' throughout to broadly encompass these different labels.

The role of teaching assistants

Despite the strong reliance on TAs in education, there is currently no widely accepted definition of their primary role, which varies across Organisation for Economic Co-operation and Development (OECD) member countries (Navarro, 2015). For example, in Germany and Ireland the role is more focused on facilitating pupils' care needs, rather than on pupils' learning needs as in the UK and other OECD countries (Webster & de Boer, 2022). The broad role of a TA consequently has a wide range of titles used nationally and internationally, including classroom assistant, paraeducator/paraprofessional, pupil support worker, teacher/instructional aide, higher level teaching assistant, and learning support assistant.

In many countries, TAs have seen their responsibilities increase and change over the last few decades, often due to policies of inclusive education (Navarro, 2015). Responsibilities have often moved away from general support in the classroom (for example, helping teachers with classroom organisation) to more direct involvement in teaching and learning, as well as first language support, and pastoral and non-classroom activities (Blatchford et al., 2007; Department for Education, 2024; B. Groom, 2006; Navarro, 2015). TAs can be deployed to support a whole class, or to provide small-group or one-to-one support both within and outside of class-taught lessons (Department for Education, 2024; Skipp & Hopwood, 2019). There is sometimes no clear distinction between the responsibilities of the teacher and the TA, particularly if TAs are used to cover teacher absences (Webster, 2024).

Increasingly, TAs provide considerable support to pupils with SEND, often without adequate training (Webster & Blatchford, 2013, 2017). Despite their increasing responsibilities, TAs often lack the same level of professional status or structured career pathways as teachers. In the UK, there is no specific training required to become a TA and each school or trust has its own entry requirements and job description. Furthermore, national role profiles for school support staff in the UK – which many schools and academies use to inform job descriptions – have not been updated since 2013 (National Joint Council for Local Government Services (NJC), 2013).

To summarise, the definition, role, responsibilities, training and professional status of TAs vary widely. For the purpose of this review, we define a TA as an adult who is employed to support teachers and help pupils with learning activities, but who is not employed by the school as a class teacher. By necessity, our definition focuses on the practical function and role that TAs perform in schools, rather than their job title, qualifications or employment arrangement.

Support for pupils with SEND

TAs are often deployed to assist pupils with SEND, and this important role creates particular challenges in deploying TAs effectively. A survey of 2,715 TAs in schools in England found that 56% of TAs reported spending most of their time providing one-to-one targeted support for children with SEND (Department for Education, 2024b), and this can include considerable amounts of teaching and planning. Similarly, a survey of 1,459 classroom assistants in Northern Ireland revealed that 63% supported pupils with SEND "almost all the time" (O'Connor et al., 2024). Two studies found that in both primary and secondary settings in England, TAs were in effect assigned "primary educator status for pupils in most need", often without adequate training (Webster & Blatchford, 2013, 2017). Observational data from these studies demonstrated that TAs often have more contact and interactions with pupils with SEND than teachers do. This means that pupils with the greatest needs can miss out on vital academic support from teachers and mainstream class teaching, as well as higher quality pedagogical input (Blatchford et al., 2009; Webster & Blatchford, 2013, 2017). The potential negative impact on pupils with SEND from missing out on this academic input means it is crucial that TAs are trained and deployed effectively.

The impact of teaching assistant support

The increasing responsibilities and varied roles of TAs have led researchers to examine in detail the impact that TAs are having on pupils' attainment. Unsurprisingly, the impact of TA support varies and appears to depend on the specific way a TA is deployed. A systematic literature review concluded that TAs can have a positive impact on pupils' literacy skills when providing interventions to individual pupils or small groups, but whole-class support has a less clear impact on pupil attainment (Alborz et al., 2009; Farrell et al., 2010). The DISS project (Blatchford et al., 2009), a large and influential longitudinal research project of school support staff in the UK, found that pupils who received a higher level of support from TAs had lower attainment. The finding was replicated across multiple samples and seven year-groups for English, science and maths, and while the impact varied slightly between subjects and year groups, it was consistently in a negative direction. There was also some evidence that pupils with SEND could be more greatly affected than pupils without SEND, but overall the relationship was observed across all pupils. The project concluded that the negative impact of TA support was unlikely to be caused by the actions of TAs as such, but rather pointed to wider issues of their deployment within schools, and also their subject and pedagogical expertise. In sum, the DISS project concluded that the impact of TA support on pupil outcomes was a consequence of decisions made about TAs, not by TAs, with the researchers identifying a clear need for TAs to be deployed appropriately and effectively.

In England, this need has been addressed in several ways. The Effective Deployment of Teaching Assistants (EDTA) Project (2012), which was a direct response to the DISS project findings, worked with schools to develop and evaluate ways to improve the impact of TAs, focusing on their preparedness, deployment and practice. This work laid the foundations for the MITA programme, which is a whole-school intervention that aims to change TAs' deployment and practice by engaging school leaders, teachers and TAs. Central to MITA is the view that the most effective way to deploy TAs in the classroom is to support the development of pupils' non-academic skills, specifically independence and engagement. The MITA theory of change predicts that where schools improve TAs' practice and pupil engagement, there may be a positive knock-on effect on learning in the longer term. An independent randomised controlled trial (RCT) of the MITA programme (Dimova et al., 2021) found that it did indeed improve TA deployment and practice, as well as a moderate improvement in pupil engagement. However, it did not improve pupils' reading attainment, compared with pupils in control schools. There are three potential explanations for this lack of improvement. First, academic outcomes may take longer to visibly improve compared to other measures such as pupil engagement. Second, the 'business as usual' control group may have been compromised, with many control schools independently using MITA materials to improve their TA deployment. Last, the evaluation used only one literacy-based outcome measure rather than a broader measure of academic attainment. How to ensure that TA deployment improves pupils' educational outcomes therefore remains unclear. Given the large amount of resources currently being spent on TAs in the UK, this needs urgent clarification.

One promising way to directly improve pupils' learning outcomes is for TAs to deliver evidence-based academic interventions. The EEF recommends that TAs deliver high-quality, one-to-one and small group support using structured, evidence-based interventions as part of its 'Deployment of Teaching Assistants' guidance (EEF, 2025). Indeed, there is good evidence from two meta-analyses that small-group and one-to-one interventions can have positive effects on pupils' academic outcomes (Dietrichson et al., 2017; Pellegrini et al., 2021), regardless of who conducts them (TAs, volunteers or teachers). There is also evidence that, when conducted specifically by TAs, interventions can have a positive impact on education outcomes, particularly literacy skills (Alborz et al., 2009; Farrell et al., 2010). A meta-analysis by the EEF of interventions conducted by TAs (incorporating 65 studies) estimated that interventions can have an impact of up to five months additional progress on pupils' academic outcomes (S. Higgins et al., 2021). However, the effect of each intervention may vary: effect sizes in the report

ranged from -0.26 (the intervention is likely to have a negative impact) to 1.62 (a large positive effect). The analysis also included a variety of research designs, some of which do not provide persuasive evidence of impact, and the reliability of evidence was not explicitly reviewed.

A current challenge for schools is therefore how to easily identify evidence-based TA-led interventions appropriate for their context. Without guidance on how to select an effective intervention, schools risk spending valuable resources on something that does not improve academic outcomes. Certain types of evidence are also particularly important to consider. For example, is there causal evidence that the intervention can improve academic outcomes, in comparison to normal class activities? (That is, if a pupil takes part in an intervention, will they improve more than they would have done if they had not taken part in the intervention?) Equally, how reliable is the evidence – has the intervention been shown to work consistently in a large number of pupils, in multiple schools, and have these results been replicated? The present research aimed to help schools engage with these vital questions.

The present study: aims and scope

The main aim of the present review was to critically evaluate the effectiveness of, and quality of evidence for, TA-led interventions that are currently available in the UK. We defined interventions as educational support activities, delivered one-to-one or in small groups, which are time-limited, structured (for example, with set materials, activities or lesson plans), and designed to improve pupil learning outcomes. Our primary research question was: how effective are individual TA-led interventions in improving learning outcomes for pupils in mainstream education aged between four and 18 years old? Our goal was to complement the current EEF guidelines (EEF, 2025) that schools should use evidence-based interventions to ensure that TAs are deployed effectively, and that schools should engage with the evidence base to select an appropriate and effective intervention. To do this, we analysed evidence from RCTs for interventions currently available in the UK and produced a summary table of evidence-based interventions to help schools select an intervention that is appropriate for their pupils' learning needs. We aimed to identify interventions that are most likely to be effective based on current evidence, and in particular interventions that are likely to benefit pupils with SEND.

Although there are many TA-led interventions available internationally, we decided to focus this research only on interventions available in the UK (even if they were evaluated elsewhere)[2]. This was to provide a review and recommendations that UK schools can directly benefit from and easily implement. The particularly high reliance on TAs in the UK makes guidance on effective TA-led interventions a high priority for schools and funders. We also wanted to ensure, as far as possible, that the interventions we reviewed would be suitable for the UK education system and curriculum.

2 Note that we did not set out to only review commercial interventions, but the majority that are currently available to all UK schools, and that have been rigorously evaluated, are only available commercially.

Methods

This study was pre-registered as a systematic review on the Open Science Framework: https://osf.io/cn6qf. Ethical approval was provided by the UCL IOE Research Ethics Committee. Although the pre-registration describes a review of interventions from any OECD member country, this report evaluates only those interventions currently available in the UK. Full results of the literature search are available at https://osf.io/7zjkh, and we will publish the results of the wider review in a separate document in due course.

Literature search

We developed our literature search based on our PICO[3] statement (see Appendix A) and following PRISMA[4] (Page et al., 2021) and Cochrane (J. P. T. Higgins, Thomas, et al., 2024) best practice. As a first step, we compiled a list of job titles used in OECD member countries that broadly fit our definition of a TA. Search terms were then tested and refined using the ERIC database via EBSCO. Our final search terms (as pre-registered) are provided in Appendix B. We searched the following databases between 5-9 August 2025 using those criteria: EconLit (via EBSCO), Education Resources Information Center (ERIC) (via EBSCO), Education Abstracts (via EBSCO), PsycINFO (OVID), and PsycARTICLES. In addition, we screened all references from 13 prior meta-analyses and reports relating to TAs (Alborz et al., 2009; Cajkler & Tennant, 2009; de Boer et al., 2014; Dietrichson et al., 2017; EEF, 2018; Farrell et al., 2010; Garrote et al., 2017; Hall & Burns, 2018; S. Higgins et al., 2021; Lewis, 2023; Paul & Clarke, 2016; Sharma & Salend, 2016; Snowling & Hulme, 2012). To capture the relevant grey (non-peer-reviewed) literature, we searched for trials conducted by the EEF (including those from Higgins et al., 2021) published on their website. We also searched for TA-led interventions listed in the sixth version of What Works for Literacy Difficulties (Lavan & Talcott, 2020), Literacy and Numeracy Catch-up Strategies (Department for Education, 2018), Effective Interventions for Struggling Readers (Nugent et al., 2019), and the MetaSense database (Van Herwegen et al., 2024). To ensure that we reviewed all existing evidence for each intervention, we conducted backwards- and forwards-searches of references and citations of the most recent trial of each intervention. We also searched for any trials mentioned or published on the developers' websites. These searches were carried out between January – February 2025. All results from these searches were screened based on the inclusion criteria listed in Table 1.

3 Population, Intervention, Comparison, and Outcome. These criteria form the framework for a systematic review.

4 Preferred Reporting Items for Systematic reviews and Meta-Analyses

Screening

Articles were screened using EPPI-Reviewer software (Thomas et al., 2022) using the priority screening tool (Tsou et al., 2020). We first screened all articles based on their title and abstract. Included papers were then screened based on the full text. Three researchers conducted the full article screening and we triple-screened two batches of included articles (20% overall) to ensure consistency, resolving discrepancies through discussion. For the purposes of this report, we then screened all included articles to identify interventions that are currently available to all schools in the UK. These articles then formed the meta-analyses and review described in this report.

Table 1. Summary of screening criteria

Criteria	Definition for inclusion
Year	Study was carried out or published between 1994 and 2024.
Language	Study is fully reported in English.
Country	Intervention was conducted in an OECD member country.
Type of article	Article is a peer-reviewed research article, a doctoral (PhD) thesis, a working paper or an institutional report.
Type of research	Article reports primary research.
Population	Study is conducted on school pupils aged between four and 18 years old.
Teaching assistant (TA) role	Intervention is conducted (partially or fully) by a TA, based on the following definition of a TA: An adult who is employed to support teachers and help pupils with learning activities, but who is not employed by the school as a class teacher.
Intervention type	Intervention involves educational support activities, delivered in one-to-one or small group settings. Interventions should be time-limited, structured, and designed to improve pupil learning outcomes or abilities.
Outcomes	Study includes at least one measure of pupil attainment or progress in any subject area, measured using standardised or diagnostic tests, researcher-designed tests, or academic assessment.
Research setting	Intervention is carried out in a mainstream primary or secondary school or further education college (or non-UK equivalent).
Research design	Study is a randomised controlled trial (RCT).
Comparator	Intervention is compared to 'business as usual' (BAU) or no intervention.

Data extraction and analysis

To estimate the average effect of each intervention, we used the primary outcome that each intervention aimed to improve, for example reading attainment or spoken language. An exception to this was the Reading and Language Intervention for Down syndrome, which aimed to improve both reading and spoken language. For this intervention, the primary outcome comprised measures of both reading and spoken language.

We extracted effect sizes for the primary measure(s) analysed in each trial, as stated in the report. Most trials reported covariate-adjusted effect sizes as their primary result (for example, from analyses including pre-test scores as a covariate). We therefore extracted covariate-adjusted effect sizes whenever available for consistency and to obtain a more precise estimate of intervention effects. We calculated the standard error *(SE)* using 95% confidence intervals (CIs) of the effect size and the equation in Appendix C (J. P. T. Higgins, Li, et al., 2024). If only the effect size and *p*-value were reported, we used a function from the dmetar package in R (Harrer et al., 2019) to calculate the *SE* and 95% CIs (Altman & Bland, 2011). This ensured that our estimates were based on the covariate-adjusted effect sizes reported in the studies. However, seven trials did not report sufficient statistics for this. For one of those trials, we calculated the effect size, *SE* and 95% CIs from raw gain score data using the esc package in R (Lüdecke, 2019). For all other trials, raw data was not available, so we calculated the *SE* and 95% CIs based on the effect size and sample size, using the equation in Appendix C. For consistency, all Cohen's *d* effect sizes were converted to Hedges' *g* using the esc package in R (Lüdecke, 2019). We also converted one η_p^2 effect size using the equation listed in Appendix C. If a trial used more than one measure for the primary outcome, we calculated the simple average of all effect sizes and its *SE*[5] to obtain a single effect size. This method avoided issues from analysing dependent outcome measures but may result in a conservative estimate of effect precision (Borenstein et al., 2021; Moeyaert et al., 2017)[6]. All conversions and calculations were conducted in R version 4.4.2 (R Core Team, 2024) and are available at https://osf.io/7zjkh.

Where an intervention had multiple trials, we pooled effect sizes using meta-analysis. We pooled effect sizes this way if the primary outcome was the same in each trial, even if different measures were used to capture the outcome. For example, we pooled effect sizes of reading attainment (the primary outcome) for both trials of Switch-on Reading, even though it was measured using different reading tests. For interventions that had only one trial, we simply report that effect size.

Meta-analyses were conducted in R using the meta package (Balduzzi et al., 2019). We used random effects models to account for the likely heterogeneity between trials. Models used a restricted maximum likelihood (REML) estimator, and were weighted by the inverse variance of effect sizes. This meant that studies with more participants were given more weight. Heterogeneity between studies was examined using the I^2 statistic. If $I^2 > 50\%$ (J. P. T. Higgins & Thompson, 2002), suggesting substantial between-study heterogeneity, this was explored further. As many trials also examined secondary outcomes (for example, reading comprehension or vocabulary knowledge), we conducted subgroup analyses of outcomes for a more fine-grained evaluation of how interventions affected different types of knowledge and skills. Full statistical output and forest plots for all meta analyses are available at https://osf.io/7zjkh.

5 Calculating the *SE* for an average effect size normally requires the correlations between variables to be known or estimated (Borenstein et al., 2021). As correlations were not available in any of the trial reports, and we could not obtain reliable estimates for all variables, *SE*s were calculated as the square root of the average sampling variance for each outcome variable. It should be noted that this method is likely to result in a more conservative estimate of effect precision compared to using precise correlations (Borenstein et al., 2021; Moeyaert et al., 2017).

6 Note that we considered other options for analysing dependent measures such as robust variance estimation (Pustejovsky & Tipton, 2022) and multi-level modelling (Konstantopoulos, 2011; Van den Noortgate et al., 2013, 2015), but due to the very small number of trials/effect sizes available for each intervention, these methods were not appropriate.

Risk of bias and limitations analysis

To assess the reliability of effects, we analysed the limitations and risk of bias of each trial, based on the approach used by Cochrane reviews. We defined bias as "…a systematic error, or deviation from the truth, in results or inferences… [which] can lead to underestimation or overestimation of the true intervention effect" (section 7:1, Boutron et al., 2024). Risk of bias in this context does not mean that the study is of poor quality, rather that there were factors (often unavoidable) that need to be considered when interpreting the effect size of an intervention. For example, some of the trials were disrupted by the Covid pandemic, meaning that the intervention was not implemented as intended, and the effect size may therefore be lower than expected if the trial had not been disrupted.

We developed an assessment tool (available at https://osf.io/7zjkh) based on the Cochrane risk of bias template (Sterne et al., 2019), and using What Works Clearinghouse guidance for reviewing RCTs in education (What Works Clearinghouse, 2024). The tool required responses to a series of questions across eight main sources of bias and limitations likely to be present in the trial:

1) Developer involvement (to what extent was the intervention developer involved in the trial? Was it an efficacy or an effectiveness trial?[7]).
2) Pupil selection for participation in the trial (for example, were pupils selected in a standardised way and according to intervention guidelines?).
3) Randomisation (for example, was randomisation compromised?).
4) Outcome measurement (for example, was the primary outcome appropriate for the intervention, and were assessors blind to pupil allocation?).
5) Missing data (for example, was missing data assessed for randomness?).
6) Presence of confounds (for example, was there a risk of spillover effects from intervention to control pupils, or did the control group also have access to the intervention?).
7) Analytical bias (for example, was the study likely to be underpowered? Did the analysis take into account school clustering and pupils' pre-test scores?).
8) Intervention fidelity and compliance (for example, was the intervention conducted as intended?).

The reviewer (first author) evaluated the risk of bias from each of these categories as high, medium or low based on responses to each question and the gravity of the bias/limitations. They also recorded the predicted direction of bias where possible – that is, if the effect size is likely to be an overestimate (bias likely to favour the intervention), or an underestimate (bias likely to favour a null finding). An overall risk of bias (high, medium or low) was then allocated to each trial based on the proportion and spread of bias/limitations across all categories, with more weight given to any categories marked as 'high risk'[8]. This analysis was used to form the risk of bias and limitations judgement for each intervention (see Table 2).

7 Efficacy trials seek to determine whether an intervention produces the expected result under ideal circumstances, often with more involvement and monitoring from the developers. In contrast, effectiveness trials measure the impact of an intervention under 'real world' settings, and typically have less involvement from the developers.

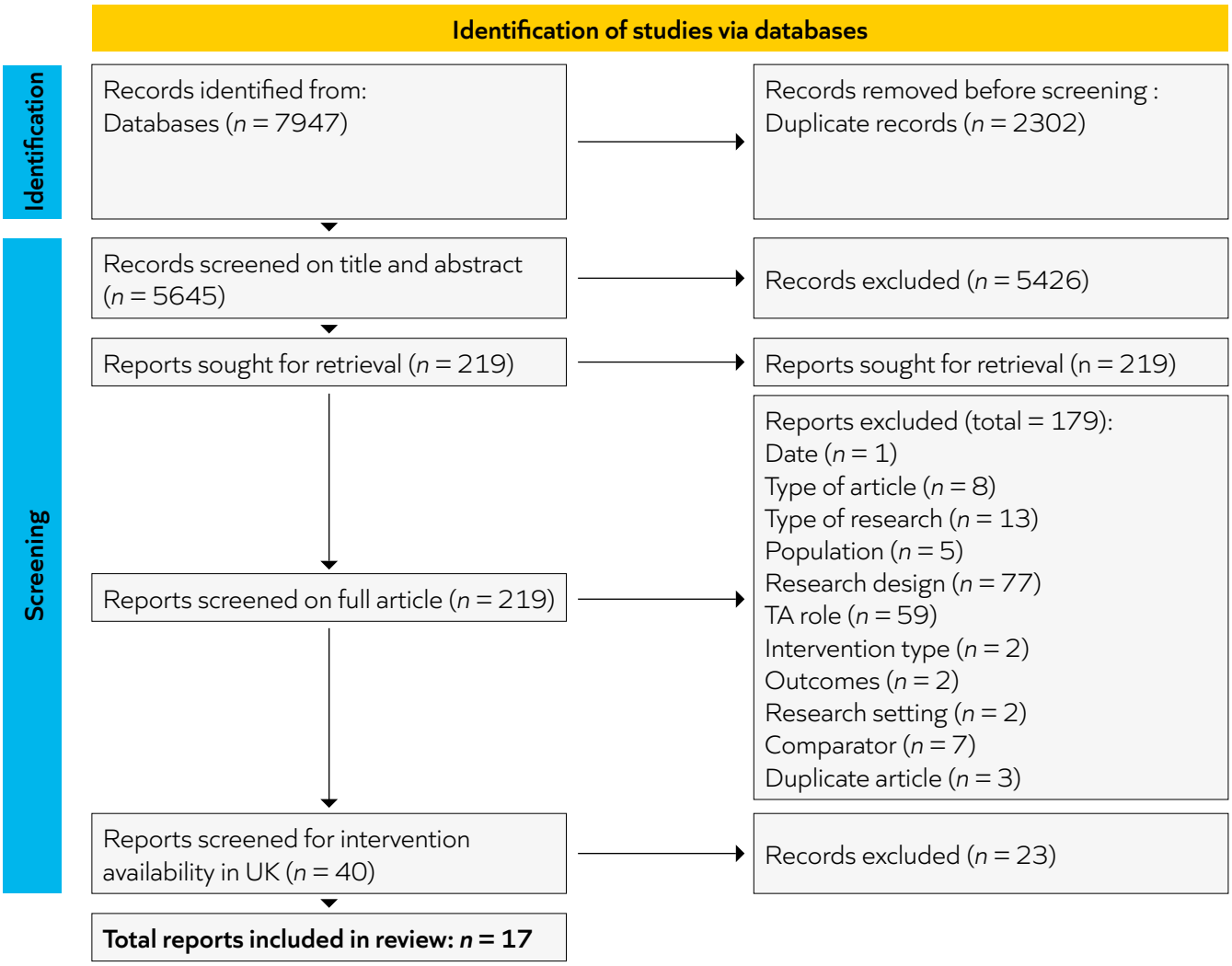
8 Overall risk for each trial was assigned based on whichever level had the highest count across all eight categories – for example, a trial with two categories marked as low risk and six as medium risk was given an overall rating of medium risk. However, categories rated as high risk were given greater weight. A trial with mostly low risk categories but one high risk category would be assigned medium risk overall. A trial with mostly medium risk categories but two high risk would be assigned medium-high risk. Any trial with more than three high risk categories was assigned high risk overall.

Results

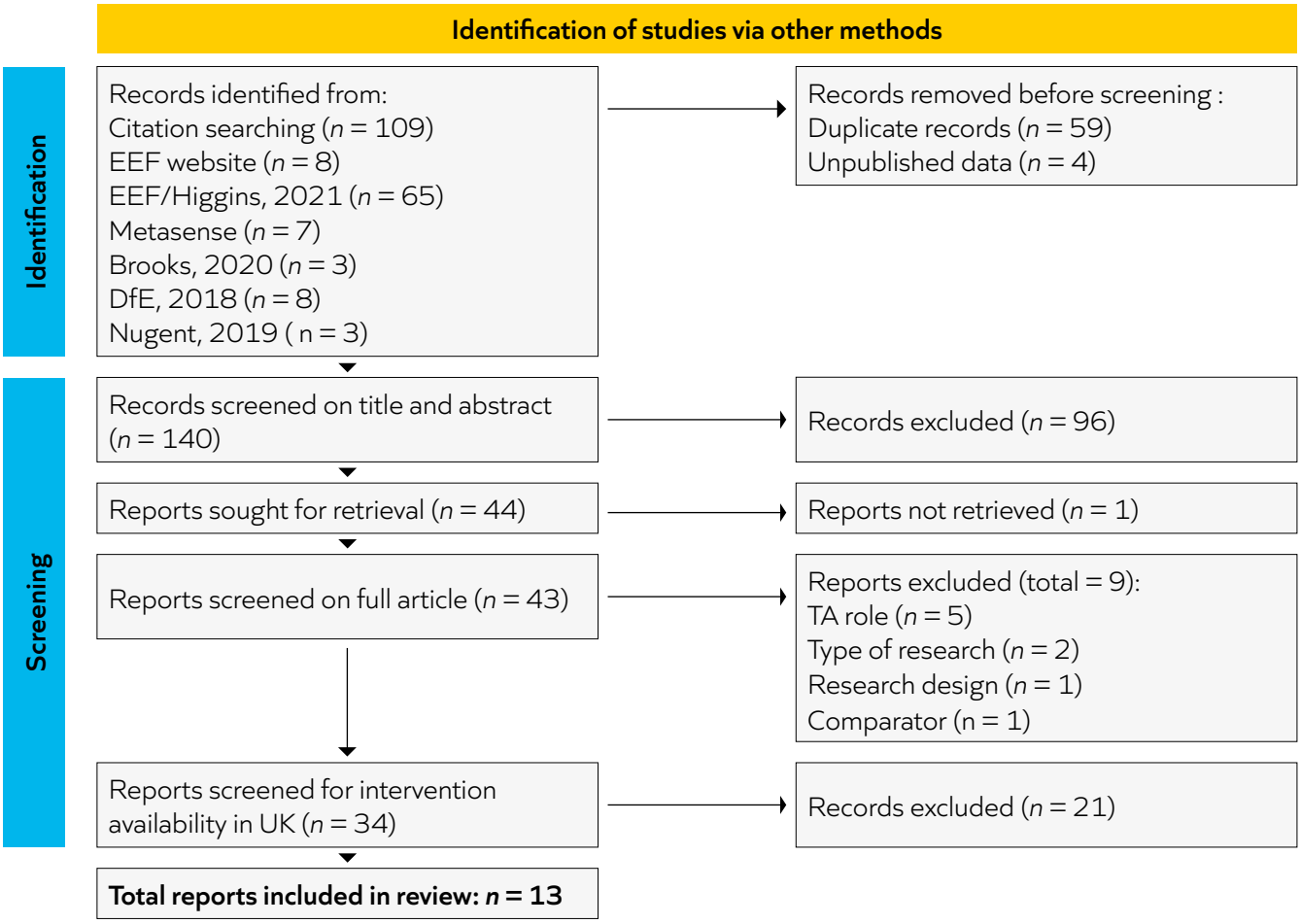
Interventions overview

Our literature searches returned a total of 5,784 articles to screen. Following screening, we included 30 research articles reporting 26 RCTs for analysis and review. These trials evaluated 17 academic interventions currently available to all schools in the UK[9] (21 including different intervention versions). Full details of all included articles and trials are available here: https://osf.io/7zjkh. Table 2 lists all 21 interventions, and Figure 1 shows the results of each stage of the screening process. Where trials included more than one version of an intervention (for example, Abracadabra (ABRA) has an online and a paper-based version) we have reviewed both versions separately.

Figure 1: PRISMA flow diagram of article searches and screening



9 By 'available to all schools in the UK', we mean that the materials can be bought or accessed by any UK school, and so can easily be adopted. In total (including interventions not available in the UK), our searches retrieved a total of 75 articles and 38 named interventions (22 of which are not currently available in the UK).



The majority of interventions that we retrieved focus on literacy (including spoken language, reading and writing), with only three targeting maths (see Figure 2); this is despite searching for interventions relating to any academic subject. The majority are aimed at primary education, although many are suitable for both primary and secondary (Figure 2). Two interventions are aimed at secondary age pupils, and two for early years (nursery to reception). Group sizes varied with a mixture of small group and individual (one-to-one support) interventions, and interventions varied in their overall duration and intensity (some requiring short daily sessions and others longer, weekly sessions; see Figure 3). Importantly, all interventions included training for the TAs leading the sessions, which is highlighted as an important element of TA deployment (EEF, 2025), and may be key to their successful implementation[10].

10 Note that the VEIP and NIP do not currently offer training, although this was provided to TAs in both trials.

Figure 2. Domain and school phase of interventions

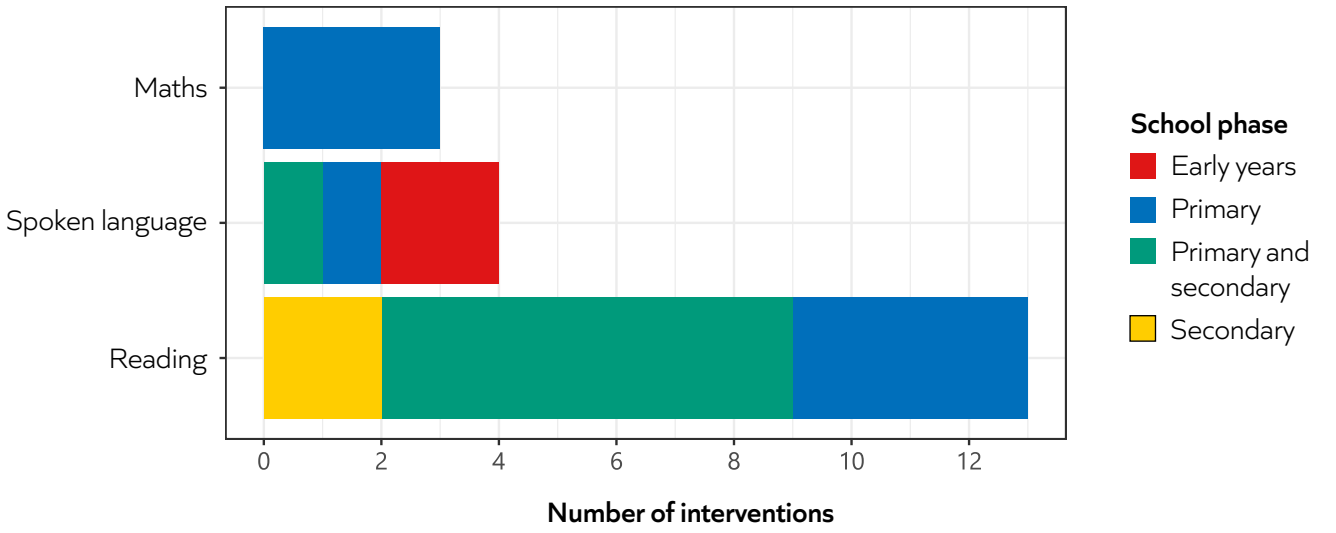
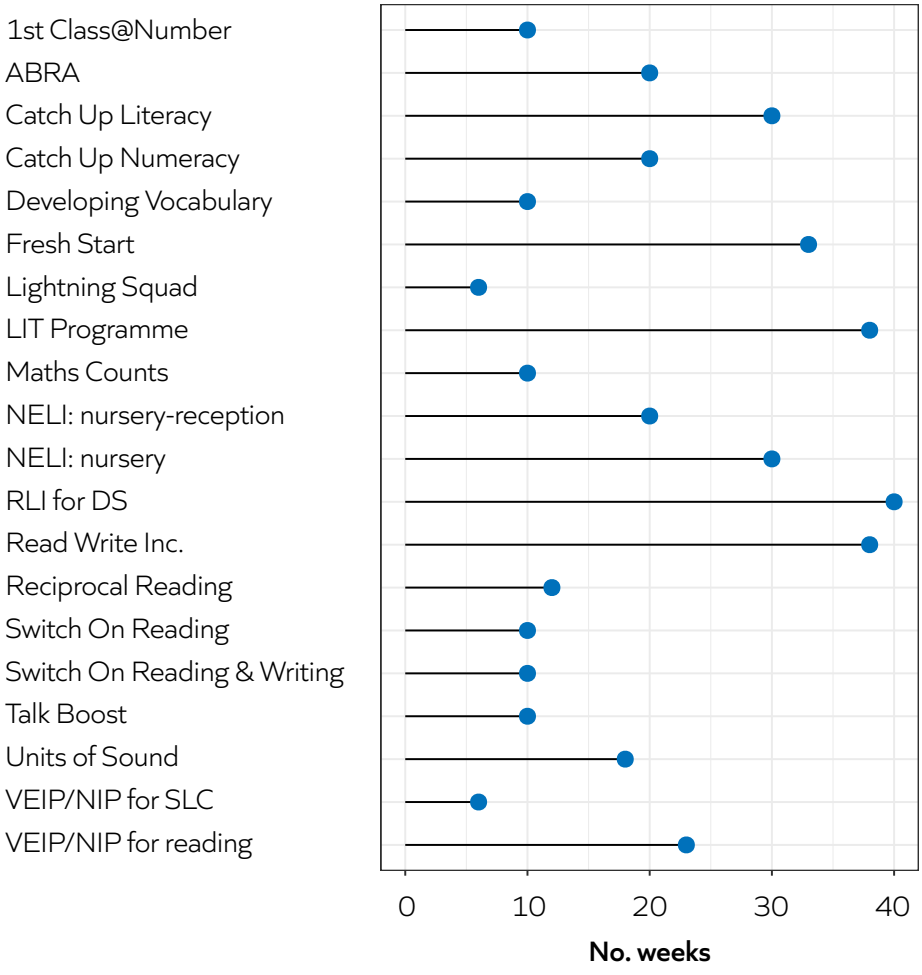
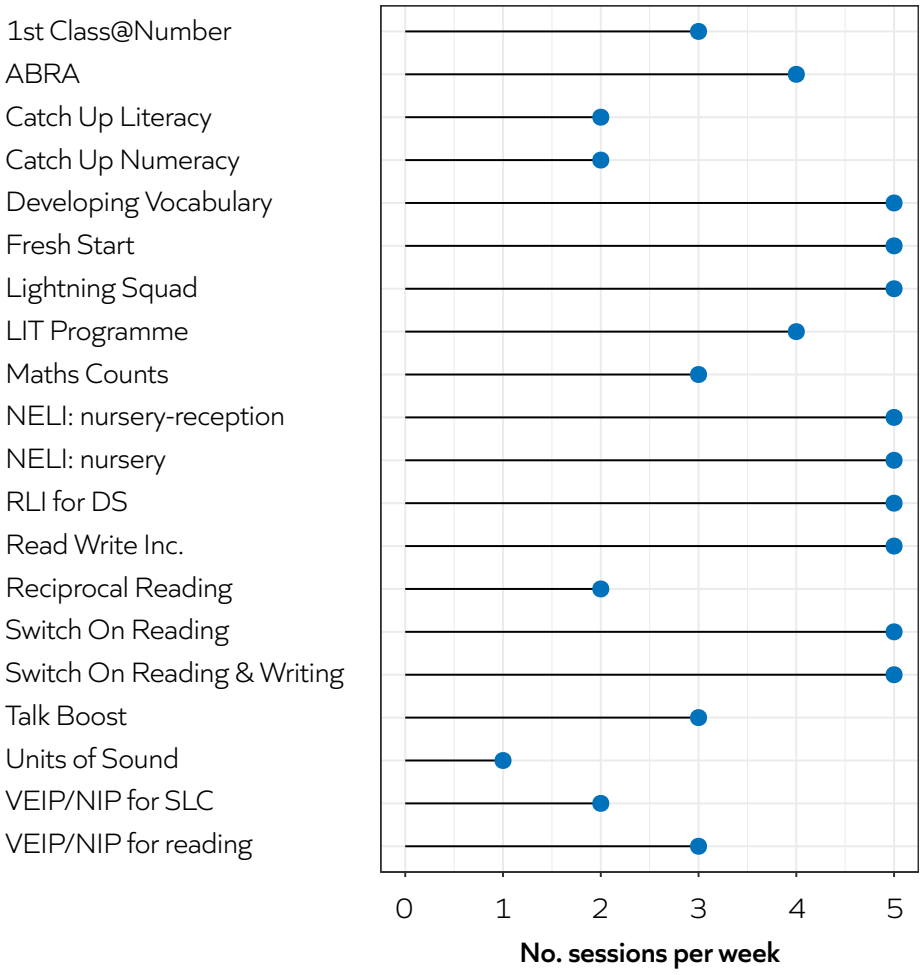
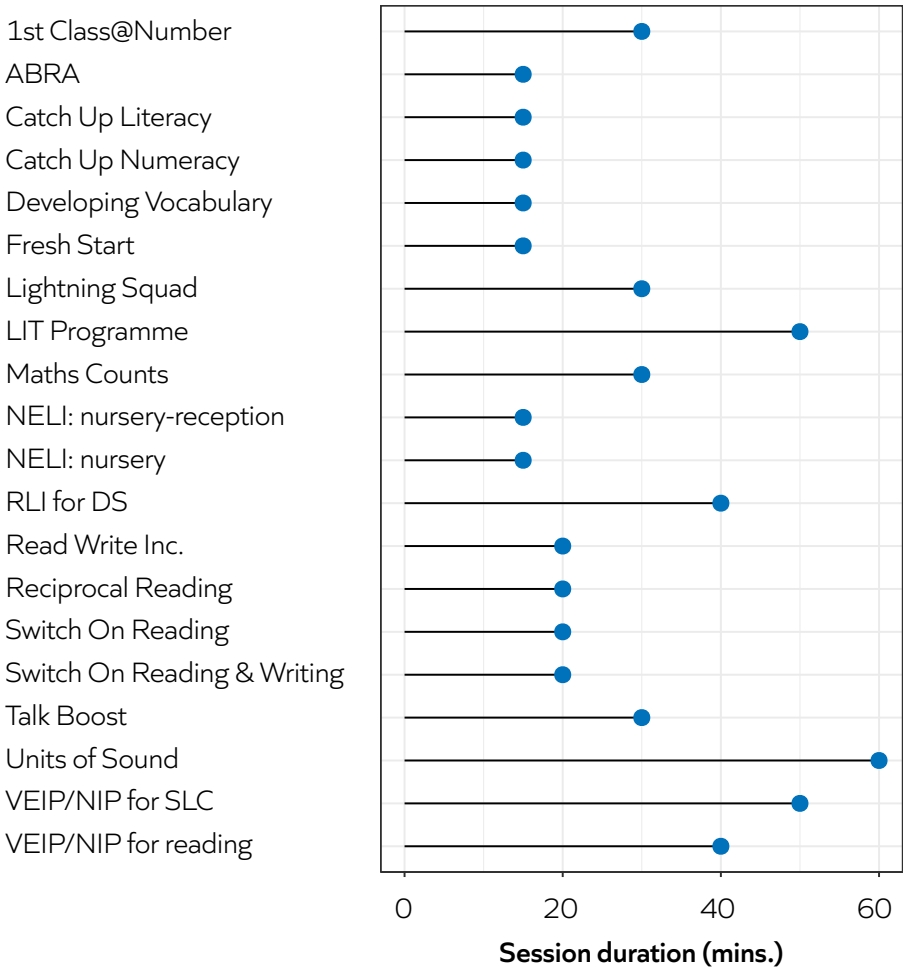


Figure 3. Intervention duration, session frequency and session duration




1st Class@Number
ABRA
Catch Up Literacy
Catch Up Numeracy
Developing Vocabulary
Fresh Start
Lightning Squad
LIT Programme
Maths Counts
NELI: nursery-reception
NELI: nursery
RLI for DS
Read Write Inc.
Reciprocal Reading
Switch On Reading
Switch On Reading & Writing
Talk Boost
Units of Sound
VEIP/NIP for SLC
VEIP/NIP for reading
0
1
2
3
4
5
No. sessions per week



Note. NELI = Nuffield Early Language Intervention; RLI for DS = Reading and Language Intervention for Down syndrome; VEIP = Vocabulary Enrichment Intervention Programme; NIP = Narrative Intervention Programme; SLC = speech, language and communication needs.

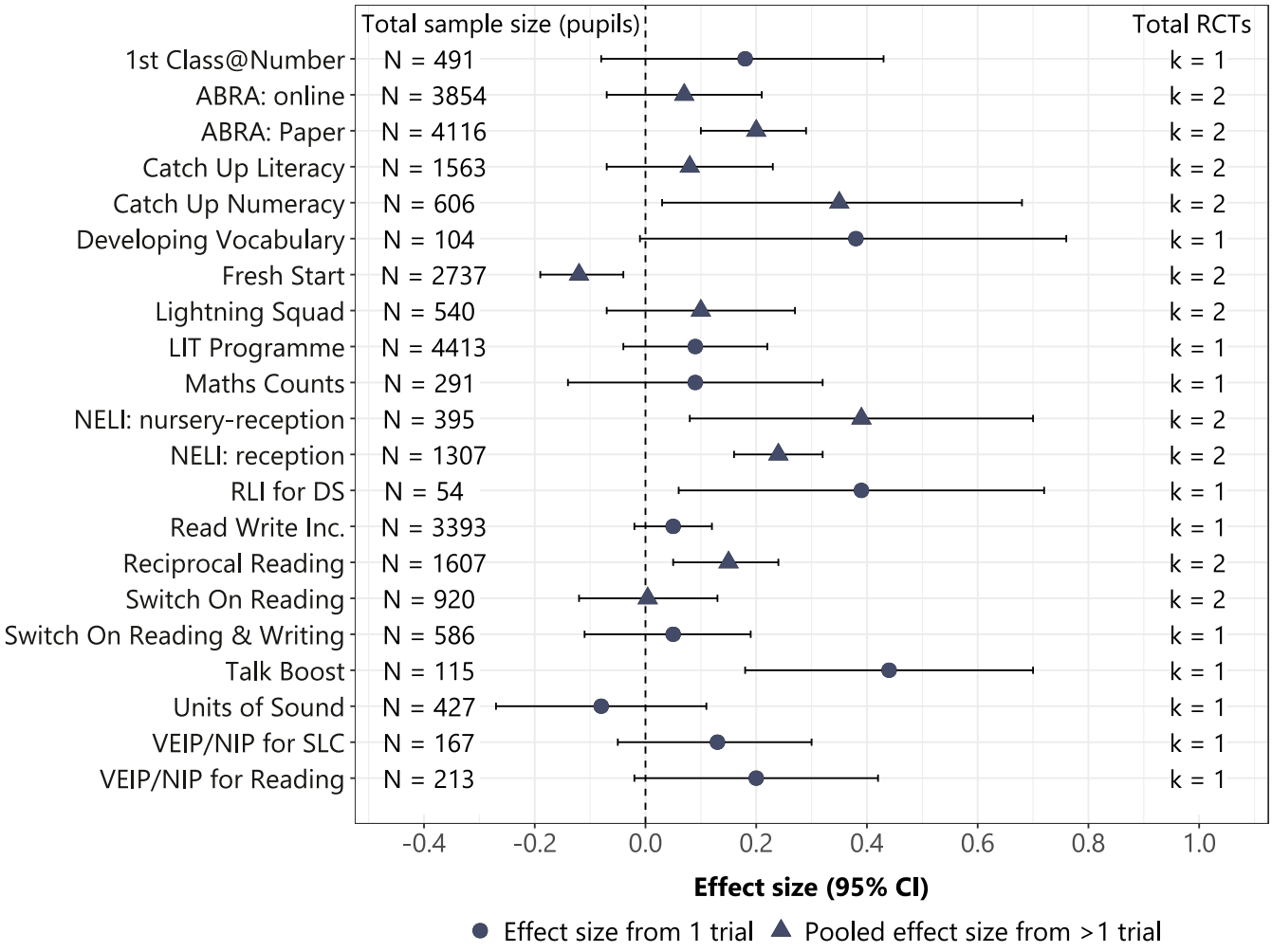
The durations and frequencies plotted here are based on the trials we reviewed but may differ from the current recommendations from intervention providers. Implementation may also have varied slightly in different schools.

Evidence of intervention effectiveness

To evaluate the evidence of each intervention and to determine its effectiveness, we considered three elements. First, we examined the estimated effect size (the pooled effect if there were multiple trials, or the average/reported effect size if only one trial was available). Second, the precision of the estimated effect size; we examined statistical significance of the pooled effect (or for the single effect size if reported) using a threshold of $p < .05$, as well as 95% CIs. Third, the outcome of the risk of bias assessment for each intervention and the predicted direction of bias. Effect sizes can vary greatly based on numerous factors, particularly sample size (for example, Brysbaert, 2019), and it is imperative to consider the reliability and precision of effects as well as the size of the effect of education interventions (Lortie-Forgues & Inglis, 2019). Precision tells us whether the same effect is likely to be observed if the study is replicated. In particular, if the lower 95% CI boundary is negative, we cannot be certain that an intervention would have a positive effect on pupil outcomes and could even have a negative impact.

Estimated effect sizes for each intervention ranged from positive to negative (min. = -0.12, max. = 0.44; see Table 2 and Figure 4). Based on empirical benchmarks proposed for education interventions (Kraft, 2020), most interventions were estimated to have medium positive effects ($g = 0.05 – 0.20$). Six interventions had large positive effects ($g > 0.20$), and one had a small effect close to zero (Switch on Reading, $g < 0.05$). Two were estimated to have negative effects (Read Write Inc. and Units of Sound), although confidence intervals for both interventions included positive values. The precision of all intervention effects varied. Based on *p*-values, only seven interventions had a positive significant effect on the primary learning outcome (see Table 2). 95% CIs varied greatly, and for several interventions they were particularly large (that is, estimates of their impact were imprecise – see Figure 4). Effect sizes for Developing Vocabulary, Catch Up Numeracy, Nuffield Early Language Intervention (nursery-reception version), RLI for Down syndrome and Talk Boost had particularly large CIs. This is potentially due to inconsistent results or the small sample sizes in some of these trials. Conversely, effect sizes for ABRA (paper version), Catch Up Literacy, Fresh Start, NELI (reception version), Reciprocal Reading and Read Write Inc. had relatively narrow CIs, indicating that we can be more confident of these estimates. Many interventions had CIs that included negative values, indicating that we can be less certain that these interventions would have a positive effect on pupil outcomes.

Figure 4. Estimated effect of each intervention on its primary outcome (with 95% confidence intervals)



Note. Where there were multiple trials for a given intervention, we report pooled effect sizes from random effects meta-analysis. Where there was only one trial for a given intervention, we report a single effect size from that trial.

NELI = Nuffield Early Language Intervention. RLI for DS = Reading and Language Intervention for Down syndrome. VEIP = Vocabulary Enrichment Intervention Programme. NIP = Narrative Intervention Programme. SLC = speech, language and communication.

Table 2. Interventions: evidence of impact, risk of bias and limitations

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
1st Class @Number					
Maths attainment	0.18† [-0.08, 0.43] .09	1 effectiveness trial in England (*N* = 491) independently evaluated by EEF. Results from second EEF trial in England due spring 2025.	Medium (may favour null)	Pupils were not selected for the trial using the recommended test and alternative guidance was unclear; risk of pupils being selected incorrectly/ inconsistently (for example, pupils with better maths abilities).	1 RCT: Nunes (2018)
ABRA					
Online version					
Reading attainment	0.07 [-0.07, 0.21] .307	Inconsistent effects from 1 efficacy trial (*N* = 1,731) and 1 effectiveness trial (*N* = 2,123) in England, both independently evaluated by EEF.	Medium-high (may favour null)	Developers provided laptops/ software to schools in efficacy trial only. Major problems with software in effectiveness trial may explain null finding in contrast to efficacy trial.	2 RCTs: Bell (2022) McNally (2016)
Paper version					
Reading attainment	0.20 [0.10, 0.29] <.001***	Consistent effects from 1 efficacy trial (*N* = 1,715) and 1 effectiveness trial (*N* = 2,401) in England, both independently evaluated by EEF.	Low (may slightly favour null)	Some evidence of spillover effects in efficacy trial (control pupils within same schools also benefitted from the intervention).	2 RCTs: Bell (2022) McNally (2016)

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
Catch Up					
Catch Up Literacy					
Reading attainment	0.08 [-0.07, 0.23] .274	Consistent negative effects from 1 efficacy trial (*N* = 557) and 1 effectiveness trial (*N* = 1006) in England, both independently evaluated by EEF.	Medium (may favour null)	Unclear how pupils were selected in either trial – no clear guidance or threshold for selection. Different pre- and post-test measures in both trials may have increased measurement error of pupil progress.	2 RCTs: Roy (2019) Rutt (2015)
Catch Up Numeracy					
Maths attainment	0.35 [0.03, 0.68] .033*	Consistent results from 1 trial in England conducted by the intervention developers/ University of Oxford (*N* = 390), and 1 effectiveness trial in England and Wales (*N* = 216) independently evaluated by EEF.	Medium (may favour intervention)	Possible risk from developer involvement and underreporting of methods in efficacy trial.	2 RCTs: Holmes (2013) Rutt (2014)
Developing Vocabulary [11]					
Reading attainment	0.38[†] [-0.01, 0.76][a] >.05[b]	1 trial conducted by Queen's University Belfast/Fisher Family Trust in England (*N* = 104). Upcoming Phase 3 trial in England by Queen's University Belfast/ Fisher Family Trust (see protocol: Cockerill, 2024).	Medium (may favour intervention)	Possible risk from developer involvement and underpowered analysis. Methods underreported.	1 RCT: Cockerill (2021)

11 Note that the effect size, CIs and *p*-value reported here are based on mean difference in gain scores and indicate a non-significant effect. Cockerill et al (2021) reported that an ANCOVA showed a significant effect of the intervention, but no statistics from this analysis were reported.

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
Fresh Start[12]					
Reading attainment	-0.12 [-0.19, -0.04] .002**	Results from 1 efficacy trial (*N* = 419) and 1 effectiveness trial (*N* = 2318), both conducted in England and independently evaluated by EEF. Effectiveness trial also included Read Write Inc. intervention.	Medium (direction unclear)	Major implementation issues in effectiveness trial, possibly due to high staffing requirements. Non-blind assessors in efficacy trial, and treatment groups unbalanced (intervention had no effect when pupil characteristics and prior attainment were accounted for).	2 RCTs: Gorard (2015b) Molotsky (2022)
Lightning Squad					
Reading attainment	0.10 [-0.07, 0.27] .269	2 independent efficacy trials (*N* = 150 and 380) conducted in USA by John Hopkins University/ US Department of Education. Upcoming trial in England to be independently evaluated by EEF.	Medium-high (direction unclear)	Major implementation/ fidelity issues in one trial due to Covid disruption. Evidence of non-random missing data in both trials.	2 RCTs: Neitzel (2024) Ross (2017)

12 Note that Gorard (2015b) report effects from two analytical models: one using gain scores as the outcome variable with no covariates, and one using post-test scores as the outcome variable, with prior attainment and pupil age as covariates. The former analysis produced a positive, medium-sized effect (*d* = 0.24). In the latter model, the intervention effect size was negative (*d* = -0.03), but there were large, positive effects of pupil age and prior attainment. This indicates that differences between the treatment and control groups at post-test were largely due to differences in pupil characteristics, rather than the effects of the intervention. We report the latter (adjusted) effect size as a more precise estimate.

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
LIT Programme					
Reading attainment	0.09[†] [-0.04, 0.22] >.05[b]	1 efficacy trial in England (*N* = 4413) independently evaluated by EEF.	High (may favour null)	Randomisation compromised and estimate based on all Year 7 pupils rather than those eligible. Large variability in implementation. Ceiling effects in outcome measure.	1 RCT: Crawford (2014)
Maths Counts [13]					
Maths attainment	0.09[†] [-0.14, 0.32][a] >.05[b]	1 efficacy trial in England (*N* = 291) independently evaluated by EEF.	Medium (direction unclear)	Different pre- and post-test measures may have increased measurement error of pupil progress. Analysis and statistics underreported.	1 RCT: See (2018)
Nuffield Early Language Intervention (NELI)					
Nursery-reception version					
Spoken language	0.39 [0.08, 0.70] .014*	Consistent effects from 2 efficacy trials in England: 1 developer-led (*N* = 166), 1 independently evaluated by EEF (*N* = 229).	Low-medium (may favour intervention)	Risk of bias from developer involvement and some underreporting of methods in developer-led trial.	Fricke (2013; 2017) Sibieta (2016)
Reception version					
Spoken language	0.24 [0.16, 0.32] <.001***	Consistent effects from 1 efficacy trial (*N* = 236) and 1 effectiveness trial (*N* = 1071) in England, both independently evaluated by EEF.	Low (may favour null)	Minor risk from fidelity issues in effectiveness trial (schools with greater fidelity had better outcomes).	2 RCTs: Dimova (2020) Fricke (2017) Sibieta (2016) West (2021)

13 Note that See et al (2018) report two effect sizes: one using gain scores as the outcome variable with no covariates, and one using post-test scores as the outcome variable, with prior attainment as a covariate. We report the latter (adjusted) effect size for consistency with other trials, and to provide a more precise estimate of the intervention effect.

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
Read Write Inc.					
Reading attainment	0.05[†] [-0.02, 0.12] .528	1 effectiveness trial in England (*N* = 3,393) independently evaluated by EEF (trial also included the Fresh Start intervention).	High (direction unclear)	Likely contamination of control group (20% control schools attended training/ purchased materials for Read Write Inc.). Low fidelity and high attrition (31%). Outcome measure (NGRT online) too difficult leading to teachers withdrawing pupils from trial.	1 RCT: Molotsky (2022)
Reading and Language Intervention (RLI) for Down Syndrome					
Reading attainment and spoken language	0.39[‡] [0.06, 0.72] <.05[b]	1 efficacy trial in England (*N* = 54) led by Down Syndrome Education International and intervention developers.	Medium (direction unclear)	Study likely underpowered (small sample) and some methods underreported.	1 RCT: Burgoyne (2012)
Reciprocal Reading					
Reading attainment	0.15 [0.05, 0.24] .002**	Consistent effects from two efficacy trials in England: 1 independently evaluated by EEF (*N* = 1296), 1 conducted by Queen's University Belfast/Fisher Family Trust (*N* = 311). Upcoming effectiveness trial in England to be independently evaluated by EEF (due summer 2025).	Low-medium (direction unclear)	Some bias possible in developer-led trial: no independent evaluation and significant effects only observed in unplanned analyses.	2 RCTs: O'Hare (2019) O'Hare (2023) Thurston (2020)

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
Switch-on					
Switch-on Reading					
Reading attainment	0.004 [-0.12, 0.13] .947	Inconsistent effects from 1 efficacy trial (*N* = 308) and 1 effectiveness trial (*N* = 296) in England, both independently evaluated by EEF.	Medium-high (may favour null)	Some fidelity issues in both trials. Evidence of non-random data, and 75% control schools implemented additional reading interventions in effectiveness trial.	Gorard (2014; 2015a)
Switch-on Reading and Writing					
Reading attainment	0.05[†] [-0.11, 0.19][a] >.05[b]	1 effectiveness trial (*N* = 586) in England, independently evaluated by EEF.	Medium-high (may favour null)	Fidelity issues and evidence of non-random data. 75% control schools implemented additional reading interventions.	1 RCT: Patel (2017)
Talk Boost					
Spoken language	0.44[†] [0.18, 0.70][a] <.001***	1 efficacy trial in England (*N* = 115) conducted by the developers (the Communication Trust and City University London). Second trial in England (independently evaluated by EEF) abandoned due to recruitment issues; no report available.	High (may favour intervention)	Randomisation compromised and pupil selection not as intended (pupils with high pre-test scores excluded after randomisation). Likely underpowered (small sample).	1 RCT: Lee (2016)

Primary outcome	Estimated effect size and significance: *g* [95% CIs] *p*-value	Evidence summary	Estimated risk of bias (direction of bias)	Risk of bias and limitations	RCTs / articles
Units of Sound					
Reading attainment	-0.08[†] [-0.27, 0.11] .210	1 efficacy trial in England (*N* = 427) independently evaluated by EEF.	Medium-high (unclear direction)	Very high attrition (46%), unclear if biased. Implementation and fidelity issues (possibly due to lengthy training and technology issues).	1 RCT: Sheard (2015)
Vocabulary Enrichment Intervention Programme (VEIP) and Narrative Intervention Programme (NIP) combined					
For language and communication difficulties					
Spoken language	0.13[‡] [-0.05, 0.30] >.05[b]	1 developer-led trial in England (*N* = 167) for pupils with language and communication difficulties.	Medium (may favour intervention)	Possible developer bias (no independent evaluation) but otherwise low risk.	1 RCT: Joffe (2019)
For reading difficulties (Talk for Literacy)					
Reading attainment	0.20[†] [-0.02, 0.42] >.05[b]	1 efficacy trial in England (*N* = 107) for pupils with reading difficulties, independently evaluated by EEF.	Medium (direction unclear)	Some issues with implementation due to pacing, complexity of materials and mixed ability groups. Different pupil eligibility thresholds used in different schools.	1 RCT: Styles (2015)

Note. Effect sizes are pooled using random effects meta-analysis, except:

[†] an un-pooled effect size from 1 trial and 1 outcome measure;

[‡] an average (unweighted) effect size of >1 primary outcome measure, from 1 trial.

[a] 95% CIs calculated using equation $g \pm 1.96*SE$

[b] Significance inferred from 95% CIs (*p*-value not reported in article)

Significance: *$p<.05$; **$p<.01$; ***$p<.001$

Maths Counts effect size based on post-test scores, adjusted for pre-test scores and prior attainment.
Sample sizes (*N*) reported above reflect the analytical sample used to obtain effect sizes for primary outcomes.

Our risk of bias analysis revealed that the reliability of effects varied greatly, and that overall, the evidence base for interventions is rather uncertain. Most interventions had medium risk of bias. Four interventions were classed as having either low or low-medium risk, another five interventions had medium-high risk, and three were classed as high risk due to multiple major problems occurring in the trials. However, the possible direction of bias varied. Sometimes a null or negative effect was more likely, meaning that the effectiveness of the intervention may be underestimated. Sometimes a positive effect was more likely, meaning that the effect may be an overestimate. For example, 1stClass@Number was estimated to have a medium positive effect (g = 0.18) on pupil outcomes. However, pupils in the trial were often selected inaccurately, and pupils with higher abilities were less likely to benefit from the intervention. This means that the effect size may be an underestimate: if pupils had been selected accurately, the effect may have been greater. In contrast, Talk Boost was estimated to have a large positive effect (g = 0.44), but randomisation was compromised in the trial. Pupils with higher scores (who were likely to benefit less from the intervention) were excluded from the analysis after randomisation meaning that the effect size may be an overestimate of the true effect.

Our risk of bias analysis highlighted some recurrent problems that reduced the reliability of many of the trials. The most common of these were: inaccurate or subjective (and therefore variable) pupil selection; lack of fidelity (for example, the intervention was not carried out as intended due to technical problems or resourcing issues); non-random missing data; inappropriate baseline and outcome measures (for example, incomparable pre- and post-test measures, or measures that were too difficult for pupils); compromised control groups (that is, control schools were running the same or very similar interventions to the one being tested); studies being underpowered (sample sizes too small), and possible bias from developer involvement. We return to these issues in the discussion.

Taken together, our analyses of primary outcomes revealed positive, significant effects with a good level of precision (95% CIs that did not cross zero), and low-medium bias (no major issues were identified with the research) for only five interventions: ABRA (paper version), Catch Up Numeracy, Nuffield Early Language Intervention (both versions), RLI for Down syndrome, and Reciprocal Reading. These interventions can be considered to have the most reliable evidence of effectiveness.

What knowledge and skills are interventions likely to improve?

Analysing primary outcomes is useful as an estimate of how effective each intervention is overall. However, interventions may have different impacts depending on the exact knowledge and skills that are measured. This is particularly important for literacy interventions where pupils can commonly have difficulties in one specific area of literacy but not in others (for example, they may have good decoding skills but have difficulty comprehending what they read). Interventions may also have different impacts depending on the time at which outcomes were measured (for example, immediately after the intervention or at a later follow-up assessment). We therefore analysed the effects of each intervention at a more fine-grained level, examining specific learning outcome measures including secondary outcomes. If the same learning outcome (for example, vocabulary knowledge) was measured in more than one trial and using more than one outcome measure, effect sizes were again pooled by conducting a subgroup meta-analysis based on the outcomes measured. Otherwise, we examined the individual effect sizes that were reported in a trial for each outcome measure. We did not analyse the three interventions found to have a high risk of bias as estimates are likely to be too unreliable. To evaluate whether an intervention is likely to address a certain learning need, we again examined effect sizes and precision. As *p*-values were not available for all effect sizes, we examined 95% CIs to estimate effect precision; if the lower boundary was positive, this was taken to be a significant, positive improvement in outcomes. Effects and corresponding statistics are provided in full in Table S2 (see Appendix D), but results are summarised here.

Overall, Reciprocal Reading is the most likely literacy intervention to improve reading overall, as it significantly improved decoding skills and reading comprehension. RLI for Down Syndrome is also likely to improve pupils' reading, and particularly decoding and vocabulary. ABRA and Lightning Squad are likely to improve decoding skills, while Developing Vocabulary and the VEIP/NIP combined are likely to improve reading comprehension; however, it is unclear how likely these four interventions are to improve other reading skills. Both versions of NELI are very likely to improve pupils' spoken language, and the reception version may also improve pupils' reading. The VEIP/NIP combined may improve spoken language in pupils with language and communication difficulties. Catch Up Numeracy is likely to improve pupils' overall maths and numeracy. Results for each of these interventions are discussed in more detail in the following paragraphs. The evidence for all other interventions was inconclusive (that is, their impact was either negative or not significant – see Table S2, Appendix D).

ABRA improved decoding skills by +2 months

In one efficacy and one effectiveness trial (both independently evaluated by the EEF), both versions of the intervention (online and paper) showed consistent improvements in pupils' letter-sound knowledge and phonological decoding. In contrast, only the paper version of ABRA consistently improved reading overall. The discrepancy between the two versions of ABRA may be due to the technical issues that schools experienced with the online version. It could also be due to its North American pronunciation and cultural focus, meaning that the paper version was overall more effective at improving reading. It should be noted that only the online version of ABRA is currently available to schools.

Catch Up Numeracy improved maths attainment by +4 months

One efficacy trial and one effectiveness trial (independently evaluated by the EEF) showed that pupils' maths attainment improves following this intervention. The standardised test used in both trials (Basic Number Screening Test) measures pupils' numeracy and mathematical understanding, however different aspects of pupils' mathematical knowledge and skills were not investigated separately.

Developing Vocabulary improved reading comprehension by +5 months

A single efficacy trial has shown improvements in reading comprehension as measured by the New Group Reading Test (NGRT). Large effects were also found for overall reading and for decoding, but these estimates were very imprecise (confidence intervals were large and crossed zero), making it uncertain whether the intervention would have a positive effect. The intervention shows promise for improving overall reading, but as only one trial has been conducted it is currently unclear whether this will be true if rolled out on a larger scale. Results from an upcoming Phase 3 trial (Cockerill et al., 2024) should be reviewed when available.

Nuffield Early Language Intervention (NELI) improved pupils' spoken language by +3-4 months

Three trials (one effectiveness, two independently evaluated by the EEF), have consistently shown that both versions of NELI led to significant improvements in pupils' spoken language. This intervention therefore has a very strong evidence base. The shorter reception-only version also improved pupils' reading by +2 months, but the longer nursery-reception version did not. The evidence for longer-term effects of NELI is less clear. Six months after the intervention, improvements in pupils' spoken language were still evident in the nursery-reception version, but not in the reception-only version (although this was only measured in one of the trials). A follow up of the reception-only version (M. Groom et al., 2023) suggests that improvements in spoken language and reading were maintained two years after the intervention. However, this should be interpreted with caution as it was an exploratory analysis of around half of all pupils included in the original trial (Dimova et al., 2020) and was severely disrupted by the Covid pandemic.

Reading and Language Intervention for Down syndrome improved pupils' reading by +5 months

One relatively small-scale trial has shown significant improvements in reading in children with Down syndrome, based on measures of word reading, letter-sound knowledge and phoneme blending, and vocabulary taught in the intervention. However, improvements did not extend to vocabulary not taught in the intervention, or to pupils' spoken language.

Reciprocal Reading improved reading (decoding and comprehension) by +2 months

Two efficacy trials (one independently evaluated by the EEF) showed that this intervention consistently improves reading (as measured by the NGRT), including reading comprehension (a subtest of the NGRT). These consistent results suggest that the intervention reliably improves pupils' decoding and comprehension skills. However, as these were efficacy trials the intervention needs to be evaluated in 'real world' school conditions. An upcoming effectiveness trial in England by the EEF (due summer 2025) should be reviewed when available to confirm the intervention's effectiveness.

Tutoring with the Lightning Squad improved decoding skills by +3 months

Two efficacy trials have shown improvements in decoding skills and letter-sound knowledge (as measured by the Woodcock Johnson III Word Attack test). However, the trials did not reveal improvements in reading comprehension or word and letter reading. Furthermore, both trials were conducted in the USA. An upcoming EEF efficacy trial in England may clarify whether similar results are likely in UK schools.

VEIP/NIP combined improved spoken language by +4 months in pupils with spoken language needs, and reading comprehension by +3 months in pupils with reading difficulties

The VEIP and NIP have been trialled as a combined intervention twice (Joffe et al., 2019; Styles & Bradshaw, 2015), but in different pupil populations and using different outcomes. When the intervention was given to pupils with reading difficulties (Styles & Bradshaw, 2015), their reading comprehension significantly improved (based on the comprehension subtest of the NGRT), but evidence was inconclusive for overall reading (based on the full NGRT) and spoken language (based on sentence recall). When the intervention was given to pupils with language and communication difficulties (Joffe et al., 2019), their verbal narrative (storytelling) skills improved but not their vocabulary. Combined[14], the VEIP and NIP interventions may therefore improve reading comprehension and spoken language, but this depends on which pupils are selected and their specific learning needs. More evidence would help ascertain which abilities are likely to improve when these interventions are combined, and how to match them to the correct pupil needs.

Which interventions are effective for pupils with SEND?

A secondary aim of this review was to evaluate which TA-led interventions are most effective for pupils with SEND. However, most trials we reviewed did not specifically target pupils with SEND or report whether pupils had additional needs or diagnoses beyond those targeted in the trial. For example, the Units of Sound intervention has been used as part of Dyslexia Action's Partnership for Literacy projects (Lavan & Talcott, 2020), but the RCT we reviewed did not target pupils with dyslexia, nor did it report whether any pupils in the trial had a diagnosis of dyslexia or additional needs beyond general reading difficulties. Most trials did not report the number of pupils with SEND included in the trial, and some trials even excluded pupils with SEND altogether. In one trial this was because certain disabilities could have prevented pupils from being screened effectively using the intervention app (NELI Language Screen; Dimova et al., 2020)[15], while in two other trials no clear reason was given for excluding pupils with SEND (Talk Boost, Lee & Pring, 2016; Switch-on Reading, Patel et al., 2017)[16]. Due to this overall lack of data, it is difficult to evaluate whether the interventions we reviewed are likely to be effective for pupils with SEND, or pupils with more complex needs than those targeted by the interventions. The three interventions discussed in the following section are the only interventions which provide informative analyses with regards to pupils with SEND.

14 Joffe (2019) also analysed outcomes for the VEIP and NIP separately, but as their statistical model fixed the weighting of each intervention, it is unclear whether the individual versions were more or less effective than the combined version.

15 Teachers were given the option to exclude children "with deafness, vision problems, or severe emotional and behavioural difficulties" from screening. It was not reported whether any children were excluded on this basis.

16 Lee (2016) reports that "children known to have language impairment or special educational needs were not included". Patel (2017) states that pupils with a "high level of special needs" were excluded. No reasons were provided in either report.

Reading and Language Intervention for Down syndrome

One trial has tested the Reading and Language Intervention in a small sample of 54 pupils with Down syndrome in 50 mainstream schools. (Burgoyne et al., 2012). The intervention aimed to improve reading and language skills in any pupils with Down syndrome, and TAs provided one-to-one daily support. After 20 weeks, pupils' word reading, decoding skills and vocabulary (words taught in the intervention) improved compared to a wait list control group (g = 0.39, 95% CIs [0.06, 0.72]). Although this effect was large and significant, it might be an overestimate due to the small sample size of the study. The study observed little transfer to non-taught vocabulary or other measures of pupils' spoken language. The specific learning needs of pupils were not reported, but importantly, younger pupils with better receptive language skills made better progress.

VEIP/NIP for speech, language and communication difficulties

Pupils with language and communication difficulties are frequently supported in the classroom by TAs (Armstrong, 2016; B. Groom, 2006). With this in mind, Joffe et al. (2019) tested a language intervention in 357 Year 7 pupils from 19 schools with language disorder (defined as language and communication difficulties). A combination of the VEIP and NIP interventions improved verbal narrative (storytelling) skills (g = 0.30, 95% CIs [0.12, 0.47]), but not vocabulary (g = -0.04, 95% CIs [-0.21, 0.14]). When used together, these two interventions may therefore be effective in improving spoken language and communication in secondary pupils with this specific need. It was unclear if pupils in this trial had any additional needs or diagnoses beyond the language and communication abilities assessed as part of the trial.

Switch-on Reading

An initial efficacy trial of this intervention (Gorard et al., 2014, 2015a) included a very large proportion of pupils with SEND (72% in the intervention group and 74% in the control group). It is unclear why this was (the authors state that it may be an overestimate due to variation in terminology and reporting between schools), and no further breakdown of pupil diagnoses or needs were reported. In a subgroup analysis of the primary reading outcome, the authors reported a much larger effect size for pupils with SEND (g = 0.33, 95% CIs [0.04, 0.57], n = 225) than the equivalent effect for all pupils (g = 0.24, 95% CIs [0.02, 0.47], N = 308). However, these effect sizes were based on gain scores. The effect for all pupils was smaller and less precise when the authors analysed post-test scores and controlled for other contributing variables, including baseline scores: g = 0.12, 95% CIs [-0.10, 0.34]. This finding suggests that Switch-on Reading may be beneficial for pupils with SEND, but without further data on the diagnoses or needs of the pupils in the trial this interpretation is uncertain. Moreover, a follow-up effectiveness trial of Switch-on Reading (Patel et al., 2017) excluded pupils with "high levels of special needs" and did not report SEND data, so cannot confirm the original finding from Gorard (2014, 2015a).

Table 3. Summary of effective, evidence-based interventions

	ABRACADABRA (ABRA) software	Tutoring with the Lightning Squad	Developing Vocabulary	Reciprocal Reading	Vocabulary Enrichment and Narrative Intervention Programmes (VEIP/NIP)
Learning domain	Literacy (reading)	Literacy (reading)	Literacy (reading)	Literacy (reading)	Literacy (reading)
Which pupils?	Ages 5-7	Ages 5-11 with reading difficulties	Ages 5-13 with reading difficulties	Ages 7-14 with difficulty understanding texts	Ages 8-18 with reading difficulties (only trialled in year 7 pupils)
How long and how often?	20 weeks 4 x 15-minute sessions/week	6 weeks+ 30 minutes daily	10-12 weeks 15-20 minutes daily	12 weeks 20-30 minutes twice a week	23 weeks 40-60 minutes twice a week
How many pupils?	Small groups (4-5)	Small groups (2-6 in pairs)	Small groups (approx. 4)	Small groups (4-8)	Small groups (3-8)
Is TA training available?	Online training course	Yes (contact developers for details)	1-day training for 1 teacher and 2 TAs	2 days online or face to face	No training currently available
What further support is available?	Support from regional trainers	Regular and on-demand training	Contact developers for information	1 to 1 online support meeting and Facebook group	No further support is currently provided
What resources are provided?	• Online platform • Printable activities • Magnetic letter cards	• Printed resources • Matched texts • Online home access for pupils • Online reading assessment, pupil progress monitor	• Tools to diagnose children with less developed word-knowledge • Teaching templates and pupil support materials	• Guides • Matched texts • Plans, videos, text suggestions, book journals • Selection of stories • Reading assessment programme	• Lesson plans • DVD • Templates/ worksheets • Picture cards
Where can I find out more?	https://literacy.concordia.ca/resources/abra/teacher/en/	https://fft.org.uk/literacy/	https://fft.org.uk/literacy/	https://fft.org.uk/literacy/	Available from publishers (see Joffe, 2017)

	Nuffield Early Language Intervention (NELI)	Catch Up Numeracy	For pupils with SEND	
			Reading and Language Intervention for Down syndrome (RLI)	Vocabulary Enrichment and Narrative Intervention Programmes (VEIP/NIP)
Learning domain	Literacy (spoken language)	Maths	Literacy (reading)	Literacy (spoken language)
Which pupils?	Ages 4-5 with spoken language needs	Ages 6 to 12 with maths needs	Ages 5-11 with Down syndrome and reading age under 8 years	Ages 8-18 with speech, language and communication needs
How long and how often?	20 weeks Per week: three 30-min. group and two 15-min. individual sessions	5 months 15 minutes twice a week	40 weeks 40 minutes daily	12 weeks 50-60 minutes twice a week
How many pupils?	Small groups (3-6) and individual	Individual	Individual	Small groups (3-8)
Is TA training available?	Online training course (10-12 hours, flexible and self-paced).	x3 ½ day sessions	Online training course (12 hours over 6 weeks)	No training currently available
What further support is available?	• Support hub with extra resources and peer support • Support team • Free webinars	Lifetime support package including online resources and training videos	Tutor support during training	No further support is currently provided
What resources are provided?	• Language Screen assessment app • Handbook • Teaching materials including flashcards, story cards and 'Ted the puppet'	• Coordinator resources • Activities, templates and digital games • Progress tracking tools	• Handbook • Online assessment, record forms, example teaching materials, video demonstrations and instruction	• Lesson plans for all 24 sessions • DVD • Templates/ worksheets • Picture cards
Where can I find out more?	https://www.teachneli.org/	https://www.catchup.org/	https://www.down-syndrome.org/en-gb/resources/reading-language-intervention/	Available from publishers (see Joffe, 2017)

Selecting an evidence-based intervention

The results from our research indicate which interventions have good, reliable evidence of their effectiveness. We present an overview of these interventions in Table 3. This summary will help school leaders to engage with the evidence base of interventions. It will also help leaders to compare evidence-based interventions on important criteria, and to select one that is likely to be effective and appropriate for their pupils' learning needs. To this end, effect sizes from our meta-analyses have been converted to estimated months progress using the standard EEF conversions, which indicate the average number of additional months of progress made by pupils who received the intervention, compared to similar pupils who received 'business as usual' (EEF, 2016). The interventions included in this tool are the most likely to benefit pupils based on current evidence. However, it should be noted that the versions of the VEIP/NIP that were trialled included training for TAs, which is not currently available. Upcoming RCTs in England of Lightning Squad (by Fisher Family Trust, evaluated by the EEF) and Developing Vocabulary (by Fisher Family Trust and Queen's University Belfast) will also provide further evidence of their effectiveness.

Principles for implementing an intervention effectively

In its current guidance (EEF, 2025), the EEF makes several recommendations to schools when implementing interventions led by TAs, and the present research broadly supports these recommendations. Although we cannot identify the factors that led to certain interventions being more effective than others (which is likely to be greatly influenced by the bias/reliability of studies), we identified three key areas that are likely important for achieving the greatest benefits from an intervention: implementation fidelity, training for TAs, and selecting the correct pupils.

Implementation fidelity

Interventions have been designed and tested to be effective when implemented in a particular way. This normally involves providing a minimum number of sessions at a certain frequency (for example, two 15-minute sessions a week for 20 weeks). Evidence suggests that pupils will get the most impact from an intervention if they attend the recommended number of sessions. For example, two studies in our review reported that the intervention was more effective when pupils had better attendance at their sessions (Dimova et al., 2020; Nunes et al., 2018). This may mean prioritising the intervention to ensure that TAs are available to implement all the required sessions.

Training for TAs

All of the RCTs reviewed here involved training for TAs to conduct the intervention correctly, and often to provide them with background knowledge and understanding of the subject. In one trial of NELI, attendance of TAs at training led to better language outcomes for pupils (Dimova et al., 2020). Many studies also reported that interventions had positive effects on TAs, such as improving their confidence and job satisfaction (Roy et al., 2019). EEF guidance recommends that TAs are adequately trained to carry out their role, and when delivering interventions (EEF, 2025). Intervention training is likely to improve both subject and pedagogical knowledge of TAs, which is vital given their increasing teaching responsibilities (Webster & Blatchford, 2013, 2017). To get the full benefits of an intervention, TAs need to have time to attend the required training and receive ongoing support from teachers and, if available, the intervention providers.

Selecting the correct pupils

Selecting the right pupils for an intervention is important to ensure that it is effective and does not negatively affect pupil outcomes. This firstly involves matching pupils' learning needs to an intervention that is likely to address those needs. The summary of effective interventions in Table 3 provides guidance on which learning needs can be addressed by each intervention. It is important to carefully follow the guidance for an intervention and, where possible, use a reliable measure to consistently select pupils. For example, Catch Up Numeracy recommend using their assessment tools to identify pupils who would benefit from the intervention and to more fully assess pupils' learning needs before taking part. Similarly, NELI

has an assessment app to guide pupil selection. This should ensure as far as possible that the intervention helps the right pupils to improve.

Finally, many of the trials we reviewed included qualitative evaluations of observations, surveys and interviews with TAs and other school staff. Although it is beyond the scope of this review to formally analyse these responses, some common issues identified across all interventions were a lack of appropriate space to conduct the intervention, insufficient planning/preparation time, and a lack of support or buy-in from teachers and senior leaders. To get the full benefits of an intervention, schools should pay careful attention to these logistical and managerial considerations. Sufficient buy-in and liaison with teachers in particular will ensure that pupils' learning from the intervention is integrated with learning in the classroom (see recommendation 3 in EEF, 2025).

Implications for schools and school leaders

We provide here a comprehensive list of effective, evidence-based small group and one-to-one interventions that TAs can conduct in schools in the UK. We recommend that schools consider using one of these interventions when selecting additional literacy or numeracy support for pupils. Using an evidence-based intervention will help ensure that TAs add value, provide the right kind of support, and help pupils improve academically. Using any of these interventions may also help TAs to be better trained, more knowledgeable and more confident in the subject area they are teaching.

We recommend schools use the summary of effective interventions (Table 3) to help them compare and select the most appropriate intervention for their learning needs. However, we recommend that schools carefully consider several things when doing so.

1. What is the pupil need you want to address? How has it been identified, and is an intervention the best way to address this need? Any intervention has a cost for pupils if it involves time away from high-quality classroom teaching, and it is important to consider the impact this may have on their progress (for example, Blatchford et al., 2009). Interventions are not the only way for TAs to support pupils with SEND, nor is using TAs always the best way to meet their learning needs, and it is important to consider other approaches. (Giangreco, 2021). Importantly, interventions delivered by TAs should supplement and add value to high-quality teaching and not act as a substitute (Blatchford et al., 2012; EEF, 2025).

2. Which intervention would best match pupils' learning needs and is it likely to improve the knowledge and skills you want to focus on? Different interventions are likely to improve different areas of knowledge and skills. It is important to correctly match an intervention to pupils' needs, and to follow intervention guidelines for selecting pupils using a reliable and objective measure where possible. If not, schools risk the intervention not being effective, or actually worsening pupil outcomes. We also recommend considering how the approach relates to wider evidence around the universal offer and adaptive teaching, as well as government guidance.

3. Carefully consider whether you can meet the requirements needed for the intervention to be effective, and whether you can give it the same priority as any other lesson. This will involve having adequate resources (such as a quiet space or appropriate IT facilities) and adequate time for TAs to be trained, carry out the intervention as intended, and to plan and prepare. It is also vital that TAs have time to liaise with teachers to ensure that the intervention is integrated with the mainstream curriculum. These requirements will likely mean that TAs' time needs to be more focused on the intervention than on other duties such as lesson cover.

Discussion

This systematic review and meta-analysis evaluated the evidence of available interventions that can be effectively conducted by TAs in UK schools. We found that three interventions in particular (Catch Up Numeracy, NELI, and Reciprocal Reading) have strong, consistent and reliable evidence overall. These are very likely to be effective if delivered by TAs in schools. A further four interventions (ABRA, Lightning Squad, Developing Vocabulary, and the VEIP/NIP combined) have good evidence of a positive impact on specific literacy abilities, while only two interventions (Reading and Language Intervention for Down syndrome, and VEIP/NIP combined) have evidence of effectiveness for pupils with SEND. We provide details of these in the summary of effective interventions (Table 3).

Overall, we found a limited amount of reliable, causal evidence for TA-led interventions. This is in contrast to the many interventions that are currently available to schools and suggests that most interventions have either been tested in non-randomised or uncontrolled trials, or have not been tested independently of the developers. In particular, the research has highlighted a substantial gap in evidence with regards to interventions for pupils with SEND. Considering the extensive use of TAs to support pupils with SEND in the UK (Webster & Blatchford, 2013, 2017), it is concerning that only two interventions we reviewed had intentionally been trialled in pupils with SEND, and several trials excluded these pupils. There is therefore limited evidence that the interventions reviewed here would be effective for pupils with needs above and beyond those targeted by the intervention. The result here is similar to findings from the Metasense project (Van Herwegen et al., 2024) which reviewed 467 studies of interventions for pupils with SEND, only 21 of which were conducted in the UK, and even fewer of which were led by TAs.

The project found that interventions overall (regardless of who leads them) can be highly effective for pupils with SEND. Nevertheless, there is clearly a need for more thorough research into how TAs can effectively support pupils with SEND – whether through interventions, or through other approaches which avoid the drawbacks of heavy reliance on TAs (see, for example, Giangreco, 2021).

There are several possible reasons for the lack of causal evidence for TA interventions overall, and particularly for pupils with SEND. First, RCTs in education are challenging and costly to conduct. Second, the relatively small size of effects that can be expected means that sample sizes need to be very large to obtain sufficient statistical power (to avoid a false positive or false negative finding), posing an additional challenge to conducting an effective RCT. This is a particular challenge for interventions aimed at pupils with SEND, as the available population will be limited and trials would potentially require a greater number of schools. Another challenge for evaluating interventions for pupils with SEND is that standardised tests are often not appropriate for them. If they are already receiving support from a TA, this may also unintentionally influence their scores on any outcome test if the TA helps them to complete it. Last, as most interventions are developed commercially, developers may have less incentive to conduct thorough and independent evaluations. Nevertheless, RCTs do provide the strongest form of evidence and, when conducted well, can be highly informative. A focus on fewer, higher quality (and non-commercial) interventions and trials would greatly benefit the sector and ensure that schools are making a worthwhile investment in evidence-based interventions.

The quality of the evidence for the 26 RCTs we reviewed varied greatly. Based on our criteria, almost all RCTs had a medium risk of bias, and three interventions were deemed high risk due to major issues during the trial, largely due to randomisation being compromised (for example, non-random missing data, pupil exclusions after randomisation, and a compromised control group). This means that the effectiveness of most interventions carries a level of uncertainty, and most interventions would benefit from more evidence to ensure that results can be replicated. The risks we identified fell broadly into two categories: 1) issues with the implementation of the intervention, and 2) issues with the research and evaluation methods.

Considering the first category, one of the most common sources of bias was inaccurate or subjective/variable pupil selection. Sometimes this stemmed from imprecise guidance from the intervention, and sometimes from schools not following the guidance. Both are equally and ethically important as selecting the wrong pupils for an intervention will be ineffective at best, and detrimental to pupil outcomes at worst. Similarly, interventions were often not carried out as intended, most commonly due to pupils missing sessions, meaning that they did not experience the full 'dosage'. Other reasons for poor fidelity were a lack of resources (such as issues with TA scheduling/availability or physical space) and technical issues where an intervention involved computer software. These issues highlight the need for schools to carefully consider whether they can meet the requirements of an intervention before implementing it, otherwise they risk the intervention not being effective. This may mean prioritising interventions in the same way as any other lesson. Indeed, in a survey of 6,000 TAs, 58% said that providing lesson cover disrupts the routine delivery of interventions (Webster, 2024).

We also encountered issues with the research methods employed in trials, some of which are common challenges for RCTs in education. Many studies were underpowered to detect the expected effect sizes from interventions (Kraft, 2020; Lortie-Forgues & Inglis, 2019) despite researchers' best efforts, highlighting the challenges of recruiting very large numbers of schools and pupils for the duration of a trial. Another common issue was use of a business as usual comparison; sometimes a high number of control schools used interventions very similar to, or the same as, the one being trialled. This means that the intervention was essentially being compared to similar interventions which would likely lead to smaller effect sizes and increase the likelihood of a false negative (that is, concluding that an intervention is not effective when it is). These challenges suggest that in some cases, RCTs may not be the best method to assess the effectiveness of an intervention (see Lortie-Forgues & Inglis, 2019 for a related discussion). Other common problems we encountered related to outcome measures, which were sometimes too easy or too difficult for pupils, or differed at pre- and post-test (thus increasing measurement error around how much pupils improved). Some trials also had compromised randomisation which meant that confounds were likely to be present. Lastly, underreporting of methods and statistics (for example, not indicating the precision of effect sizes) was common and meant that the reliability of any conclusions could not be assessed.

One of the key issues was the involvement of developers in the intervention evaluation, which varied greatly across the reports we reviewed. Many trials were independently evaluated by the EEF, and this body of research has gone a long way to improve the evidence base for interventions. However, many trials were fully led and conducted by the developers with no independent evaluation, increasing the likelihood of bias. Related to this is the extent to which interventions were trialled in everyday conditions. Even with some independent evaluations, the developers had a higher degree of involvement than they would ordinarily, for example, providing schools with additional resources or closely monitoring how the intervention was implemented. While this level of involvement is not wrong as such, and is often necessary to successfully conduct the trial, it does not help us to understand whether an intervention will be effective in normal, everyday school contexts. It is therefore vital to trial interventions in everyday conditions (for example, in a large-scale effectiveness trial), and to understand the challenges schools might face when implementing the intervention.

Many of the research limitations we encountered could, and should, be mitigated by aiming for higher research standards when evaluating interventions – for example, ensuring that all trials are independently evaluated and follow the guidelines set out by What Works Clearinghouse (US Department of Education, 2022). Consistently following an agreed set of standards for evaluating education interventions could greatly improve the reliability of the evidence base, and ultimately help school leaders to make better decisions.

Limitations

There are several important limitations to the present work. First, any systematic review is limited by its inclusion criteria, and applying different criteria may result in different findings. For example, we excluded quasi-experimental research designs and only included RCTs. This was to ensure that the studies we reviewed had high internal validity, thereby maximising the comparability of effect sizes for meta-analysis. It should be noted that other research designs are of course valid, and similar reviews (for example, S. Higgins et al., 2021; Van Herwegen et al., 2024) have included quasi-experimental studies. We also chose (for this report) to focus only on interventions available to all schools in the UK, which resulted in a bias towards commercially available interventions. We acknowledge that many other interventions with valid RCTs exist (for example, REACH, RALI and Reading Intervention; see full results from our literature searches: https://osf.io/7zjkh), although at present the materials for these interventions are not available to all schools in the UK. In sum, the effective interventions identified here are based on transparent criteria that indicate reliable evidence, but it should be noted that they may not match other recommendations using different methods or criteria.

Second, our searches were limited to specific research databases and grey literature sources to focus as much as possible on reliable, peer-reviewed research. During the review, it became apparent that much intervention research is not available through research databases and often not publicly available, particularly research conducted by intervention developers. Although we tried to search as comprehensively as possible, the diversity of sources meant that an exhaustive search was not practically feasible. This means that some interventions that claim to be evidence-based may not have been included in this review.

Third, most of the research we identified was conducted in England. This is likely due to the large number of RCTs conducted by the EEF (which aims to improve attainment in English schools), but also because we focused on interventions available in the UK. This means that the results may be less applicable to the other UK nations (Scotland, Wales and Northern Ireland) and to other countries beyond the UK.

Last, it is important to note that the evidence-based interventions identified in this review do not cover all learning needs or areas of knowledge and skills that schools may want to address. This is because we chose to focus on the highest-quality evidence so we could identify interventions that are most likely to be effective. We also focused solely on academic outcomes, while TAs can have an important pastoral role and conduct interventions that aim to improve emotional wellbeing, for example. Our summary of effective interventions may therefore not have a suitable intervention for every school. Particularly for pupils with SEND, we recommend also consulting the Metasense review and database (https://www.educationalneuroscience.org.uk/metasense/; Van Herwegen et al., 2024). In general, we recommend following the EEF 'Deployment of Teaching Assistants' guidance (EEF, 2025). If schools do not wish to adopt the interventions reviewed here, they can still follow the implementation recommendations along with EEF guidance to ensure that any intervention they are using is as effective as possible.

Implications for researchers

Our research has implications for the education sector more widely. The quality of the evidence base for interventions is very mixed, meaning that it is difficult for schools to know how effective an intervention is, and whether it is value for money. Research from the EEF has addressed this issue for many interventions, but the sector could be greatly improved by consistently following widely agreed upon research and evidence standards for trials and evaluations (such as the What Works Clearinghouse guidance). Particularly, an intervention should be required to meet certain standards of evidence before it is recommended for use by schools. Any claims that an intervention is evidence-based should be evaluated transparently and independently from the developers and should relate to whether the intervention works in everyday conditions. There is also an urgent need for more evidence around interventions conducted by TAs for pupils with SEND. Evidence is needed to establish whether existing interventions (such as the ones recommended here) are effective in pupils with additional or more complex needs and if not, how these pupils can be best supported. It is important to note that TA-led interventions are just one approach to supporting pupils with SEND, and schools should always consider whether a different (and more inclusive) approach would be more effective. Finally, the evidence base around effective deployment of TAs would benefit from research comparing interventions delivered by TAs and qualified teachers, to better understand where TAs can have the most impact.

References

Alborz, A., Pearson, D., Farrell, P., & Howes, A. (2009). *The Impact of Adult Support Staff on Pupils and Mainstream Schools* (Technical Report No. 1702T; Research Evidence in Education Library). EPPI Centre, Social Science Research Unit, Institute of Education, University of London. https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2438

Altman, D. G., & Bland, J. M. (2011). How to obtain the confidence interval from a P value. *BMJ (Clinical Research Ed.), 343*(d2090). https://doi.org/10.1136/bmj.d2090

Armstrong, F. (2016). Inclusive education: The key role of Teaching Assistants. In *Key Issues for Teaching Assistants* (2nd ed.). Routledge.

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *BMJ Mental Health, 22*(4). https://doi.org/10.1136/ebmental-2019-300117

Bell, K., Ainsworth, H., Elliott, L., Fairhurst, C., Gascoine, L., Hanley, P., Haynes, A., Hewitt, C., Mitchell, A., Torgerson, C. J., Torgerson, D. J., & Whiteside, K. (2022). *ABRACADABRA (ABRA) and Reading and Understanding in Key Stage 1 (RUKS) Evaluation Report*. Education Endowment Foundation.

Blatchford, P., Bassett, P., Brown, P., Koutsoubou, M., Martin, C., & Russell, A. (2009). *Deployment and Impact of Support Staff in Schools. The Impact of Support Staff in Schools (Results from Strand 2, Wave 2)* (No. DCSF-RR148). Institute of Education, University of London.

Blatchford, P., Bassett, P., Brown, P., Martin, C., Russell, A., & Webster, R. (2007). *Deployment and Impact of Support Staff in Schools. Report on findings from the Second National Questionnaire Survey of Schools, Support Staff and Teachers (Strand 1, Wave 2—2006)* (No. DCSF-RR005). Institute of Education, University of London.

Blatchford, P., Webster, R., & Russell, A. (2012). *Challenging the Role and Deployment of Teaching Assistants in Mainstream Schools: The Impact on Schools. Final Report on the Effective Deployment of Teaching Assistants (EDTA) project.* Institute of Education, University of London.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis.* John Wiley & Sons.

Boutron, I., Page, M. J., Higgins, J. P. T., Altman, D. G., Lundh, A., & Hróbjartsson, A. (2024). Chapter 7: Considering bias and conflicts of interest among the included studies [last updated August 2022]. In *Cochrane Handbook for Systematic Reviews of Interventions version 6.5.* Cochrane. Available from cochrane.org/handbook

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition, 2*(1), 16. https://doi.org/10.5334/joc.72

Burgoyne, K., Duff, F. J., Clarke, P. J., Buckley, S., Snowling, M. J., & Hulme, C. (2012). Efficacy of a reading and language intervention for children with Down syndrome: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 53*(10), 1044–1053. https://doi.org/10.1111/j.1469-7610.2012.02557.x

Cajkler, W., & Tennant, G. (2009). Teaching assistants and pupils' academic and social engagement in mainstream schools: Insights from systematic literature reviews. *The International Journal of Emotional Education, 1(2), 71–90.*

Cockerill, M., Roseth, C., & O'Keeffe, J. (2024). A Phase 3 randomized controlled trial of a vocabulary program in elementary schools in England: Protocol. *International Journal of Educational Research, 128,* 102448. https://doi.org/10.1016/j.ijer.2024.102448

Cockerill, M., Thurston, A., Taylor, A., O'Keeffe, J., & Chiang, T.-H. (2021). A Phase 2 Exploratory Trial of a Vocabulary Intervention in High Poverty Elementary Education Settings. *Social Inclusion, 9*(4), 12–25. https://doi.org/10.17645/si.v9i4.4553

Crawford, C., & Skipp, A. (2014). *LIT Programme Evaluation Report and Executive Summary.* Education Endowment Foundation (EEF). *https://educationendowme ntfoundation.org.uk/projects-and-evaluation/projects/lit-programme*

de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the Attributes of Educational Interventions on Students' Academic Performance: A Meta-Analysis. *Review of Educational Research, 84*(4), 509–545. https://doi.org/10.3102/0034654314540006

Department for Education. (2018). *Literacy and numeracy catch-up strategies.* Department for Education, UK. https://www.gov.uk/government/publications/literacy-and-numeracy-catch-up-strategies

Department for Education. (2024a). *School workforce in England, Reporting year 2023.* Department for Education, UK. https://explore-education-statistics.service.gov.uk/find-statistics/school-workforce-in-england

Department for Education. (2024b). *Use of teaching assistants in schools—Research report.* Department for Education, UK. https://www.gov.uk/government/publications/use-of-teaching-assistants-in-schools.

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research, 87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Dimova, S., Culora, A., Brown, E. R., Sutherland, A., & Curran, S. (2021). *Maximising the Impact of Teaching Assistants—Evaluation Report.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maximising-the-impact-of-teaching-assistants

Dimova, S., Ilie, S., Brown, E. R., Broeks, M., Culora, A., & Sutherland, A. (2020). *The Nuffield Early Language Intervention Evaluation Report.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/nuffield-early-language-intervention1

Education Scotland. (2023). *Pupil Support Staff Engagement Programme Evaluation Report.* https://education.gov.scot/resources/pupil-support-staff-engagement-programme-2022-2023/

Education Workforce Council. (2024). *Annual Education Workforce Statistics for Wales 2024.* https://www.ewc.wales/site/index.php/en/about-us/policy-and-research/workforce-statistics#annualeducation-workforce-statistics

EEF. (2016). *The Sutton Trust-Education Endowment Foundation Teaching & Learning Toolkit.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/resources/teaching-learning-toolkit

EEF. (2018). *Making best use of teaching assistants*. Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/tools/guidance-reports/making-best-use-of-teaching-assistants

EEF. (2025). *Deployment of Teaching Assistants Guidance Report*. Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/teaching-assistants

Farrell, P., Alborz, A., Howes, A., & Pearson, D. (2010). The impact of teaching assistants on improving pupils' academic achievement in mainstream schools: A review of the literature. *Educational Review, 62*(4), 435–448. https://doi.org/10.1080/00131911.2010.486476

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280–290. https://doi.org/10.1111/jcpp.12010

Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., Lervåg, A., Snowling, M. J., & Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 58*(10), 1141–1151. https://doi.org/10.1111/jcpp.12737

Garrote, A., Sermier Dessemontet, R., & Moser Opitz, E. (2017). Facilitating the social participation of pupils with special educational needs in mainstream schools: A review of school-based interventions. *Educational Research Review, 20*, 12–23. https://doi.org/10.1016/j.edurev.2016.11.001

Giangreco, M. F. (2021). Maslow's Hammer: Teacher assistant research and inclusive practices at a crossroads. *European Journal of Special Needs Education, 36*(2), 278–293. https://doi.org/10.1080/08856257.2021.1901377

Gorard, S., See, B. H., & Siddiqui, N. (2014). *Switch-on Reading Evaluation Report and Executive Summary*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/switch-on-reading

Gorard, S., Siddiqui, N., & See, B. H. (2015a). An evaluation of the 'Switch-on Reading' literacy catch-up programme. *British Educational Research Journal, 41*(4), 596–612. https://doi.org/10.1002/berj.3157

Gorard, S., Siddiqui, N., & See, B. H. (2015b). *Fresh Start Evaluation Report and Executive Summary*. Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/fresh-start

Groom, B. (2006). Building relationships for learning: The developing role of the teaching assistant. *Support for Learning, 21*(4), 199–203. https://doi.org/10.1111/j.1467-9604.2006.00432.x

Groom, M., Brown, E. R., & Lymperis, L. (2023). *The Nuffield Early Language Intervention Addendum Report*. Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/nuffield-early-language-intervention1

Hall, M. S., & Burns, M. K. (2018). Meta-analysis of targeted small-group reading interventions. *Journal of School Psychology, 66*, 54–66. https://doi.org/10.1016/j.jsp.2017.11.002

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2019). *dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'* (version 0.1.0.) [Computer software]. http://dmetar.protectlab.org/

Higgins, J. P. T., Li, T., & Deeks, J. J. (Eds.). (2024). Chapter 6: Choosing effect measures and computing estimates of effect [last updated August 2023]. In *Cochrane Handbook for Systematic Reviews of Interventions version 6.5*. Cochrane. Available from www.training.cochrane.org/handbook

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5* (updated August 2024). Cochrane. Available from www.training.cochrane.org/handbook

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, S., Villanueva Aguilera, B., Dobson, E., Gascoine, L., Kalambouka, A., Katsipataki, M., Reardon, J., Rovida, M., & Uwimpuhwe, G. (2021). *Teaching Assistant Interventions.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/teaching-assistant-interventions

Holmes, W., & Dowker, A. (2013). Catch Up Numeracy: A targeted intervention for children who are low-attaining in mathematics. *Research in Mathematics Education, 15*(3), 249–265. https://doi.org/10.1080/14794802.2013.803779

Joffe, V. 2017. Vocabulary Enrichment Intervention Programme. Routledge.

Joffe, V. 2017. Narrative Intervention Programme. Routledge.

Joffe, V. L., Rixon, L., & Hulme, C. (2019). Improving storytelling and vocabulary in secondary school students with language disorder: A randomized controlled trial. *International Journal of Language & Communication Disorders, 54*(4), 656–672. https://doi.org/10.1111/1460-6984.12471

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*(1), 61–76. https://doi.org/10.1002/jrsm.35

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lavan, G., & Talcott, J. (2020). *Brooks's What Works for Literacy Difficulties? 6th Edition.* The School Psychology Service. https://www.theschoolpsychologyservice.com/what-works-for-literacy-difficulties-6th-edition/

Lee, W., & Pring, T. (2016). Supporting language in schools: Evaluating an intervention for children with delayed language in the early school years. *Child Language Teaching and Therapy, 32*(2), 135–146. https://doi.org/10.1177/0265659015590426

Lewis, G. (2023). The classroom deployment of teaching assistants in England: A critical review of literature from 2010 to 2020. *Educational Review, 77*(1), 303–319. https://doi.org/10.1080/00131911.2023.2184773

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned? *Educational Researcher, 48*(3), 158–166. https://doi.org/10.3102/0013189X19832850

Lüdecke, D. (2019). *esc: Effect Size Computation for Meta Analysis* (Version 0.5.1) [Computer software]. https://CRAN.R-project.org/package=esc

McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016). *ABRA: Online Reading Support Evaluation report and executive summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/abracadabra-abra-pilot

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology, 20*(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189

Molotsky, A., Dias, P., & Nakamura, P. (2022). *Read Write Inc. Phonics and Fresh Start Evaluation Report.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/read-write-inc-and-fresh-start

National Joint Council for Local Government Services (NJC). (2013). *NJC Job Evaluation Scheme: Development of NJC School Support Staff Model Role Profiles.* National Joint Council for Local Government Services.

Navarro, F. M. (2015). *Learning support staff: A literature review.* OECD. https://doi.org/10.1787/5jrnzm39w45l-en

Neitzel, A. J. (2024). *An Evaluation of Tutoring with the Lightning Squad on the Reading Achievement of Struggling Readers in Grades 2 and 3: Technical Report.* John Hopkins School of Education. https://doi.org/10.35542/osf.io/yp2wr

Nugent, M., Gannon, L., Mullan, Y., & O'Rourke, D. (2019). *Effective Interventions for Struggling Readers— A Good Practice Guide for Teachers (Second edition).* National Educational Psychology Service. https://www.gov.ie/en/publication/0053fd-effective-interventions-for-struggling-readers-a-good-practice-guide/

Nunes, T., Barros, R., Evangelou, M., Strand, S., Mathers, S., & Sanders-Ellis, D. (2018). *1stClass@ Number evaluation report and executive summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/1stclassnumber

O'Connor, U., Courtney, C., Montgomery, A., & Curran, C. (2024). *A Profile of Classroom Assistants in Northern Ireland.* Ulster University. https://pure.ulster.ac.uk/en/publications/a-profile-of-classroom-assistants-in-northern-ireland

O'Hare, L., Stark, P., Cockerill, M., Lloyd, K., McConnellogue, S., Gildea, A., Biggart, A., Bower, C., & Connolly, P. (2023). Comparing the effectiveness of two reciprocal reading comprehension interventions for primary school pupils in disadvantaged schools. *British Journal of Educational Psychology, 93*(4), 1123–1145. https://doi.org/10.1111/bjep.12623

O'Hare, L., Stark, P., Cockerill, M., Lloyd, K., McConnellogue, S., Gildea, A., Biggart, A., Connolly, P., & Bower, C. (2019). *Reciprocal Reading Evaluation Report.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reciprocal-reading

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71. https://doi.org/10.1136/bmj.n71

Patel, R., Jabin, N., Bussard, L., Cartagena, J., Haywood, S., & Lumpkin, M. (2017). *Switch-on Effectiveness Trial Evaluation report and executive summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/switch-on-effectiveness-trial

Paul, S.-A. S., & Clarke, P. J. (2016). A systematic review of reading interventions for secondary school students. *International Journal of Educational Research, 79,* 116–127. https://doi.org/10.1016/j.ijer.2016.05.011

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). Effective Programs in Elementary Mathematics: A Meta-Analysis. *AERA Open, 7,* 2332858420986211. https://doi.org/10.1177/2332858420986211

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. *Prevention Science, 23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.2) [Computer software].

Ross, S. M., Laurenzano, M., & Madden, N. A. (2017). *An Evaluation of the Lightning Squad Computer-Assisted Small Group Tutoring Program on the Reading Achievement of Disadvantaged Students in Grades 1-3. Technical Report*. Center for Research and Reform in Education, John Hopkins University. https://www.successforall.org/wp-content/uploads/2018/05/Ross-Laurenzano-Madden-2017.pdf

Roy, P., Rutt, S., Buchanan, E., Rennie, C., Martin, K., & Walker, F. (2019). *Catch Up® Literacy Evaluation report and executive summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-literacy-effectiveness-trial

Rutt, S., Easton, C., & Stacey, O. (2014). *Catch Up® Numeracy Evaluation Report and Executive Summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-numeracy

Rutt, S., Kettlewell, K., & Bernardinelli, D. (2015). *Catch Up® Literacy Evaluation Report and Executive Summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-literacy

See, B. H., Morris, R., Gorard, S., & Siddiqui, N. (2018). *Maths Counts evaluation report and executive summary*. Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/maths-counts

Sharma, U., & Salend, S. J. (2016). Teaching Assistants in Inclusive Classrooms: A Systematic Analysis of the International Research. *Australian Journal of Teacher Education, 41*(8), 118–134.

Sheard, M., Chambers, B., & Elliott, L. (2015). *Units of Sound Evaluation Report and Executive Summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/units-of-sound

Sibieta, L., Kotecha, M., & Skipp, A. (2016). *Nuffield Early Language Intervention Evaluation report and executive summary*. Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/nuffield-early-language-intervention

Skipp, A., & Hopwood, V. (2019). *Deployment of teaching assistants in schools.* Department of Education. https://www.gov.uk/government/publications/the-deployment-of-teaching-assistants-in-schools

Snowling, M. J., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language & Communication Disorders, 47*(1), 27–34. https://doi.org/10.1111/j.1460-6984.2011.00081.x

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, 366*, l4898. https://doi.org/10.1136/bmj.l4898

Styles, B., & Bradshaw, S. (2015). *Talk for Literacy Evaluation Report and Executive Summary.* Education Endowment Foundation (EEF). https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-literacy

Thomas, J., Graziosi, S., Brunton, J., Ghouze, Z., O'Driscoll, P., Bond, M., & Koryakina, A. (2022). *EPPI-Reviewer: Advanced software for systematic reviews, maps and evidence synthesis* (Version 4.14.2.0) [Computer software]. EPPI Centre, UCL Social Research Centre. https://eppi.ioe.ac.uk/eppireviewer-web/home

Thurston, A., Cockerill, M., Chiang, T.-H., Taylor, A., & O'Keeffe, J. (2020). An efficacy randomized controlled trial of Reciprocal Reading in secondary schools. *International Journal of Educational Research, 104*, 101626. https://doi.org/10.1016/j.ijer.2020.101626

Tsou, A. Y., Treadwell, J. R., Erinoff, E., & Schoelles, K. (2020). Machine learning for screening prioritization in systematic reviews: Comparative performance of Abstrackr and EPPI-Reviewer. *Systematic Reviews, 9*(1), 73. https://doi.org/10.1186/s13643-020-01324-7

US Department of Education. (2022). *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.* Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/handbooks

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2

Van Herwegen, J., Dockrell, J., Thomas, M. S. C., Marshall, C., Gordon, R., & Masterman, T. (2024). *Raising educational outcomes for students with Special Educational Needs and Disabilities.* Nuffield Foundation. https://www.educationalneuroscience.org.uk/metasense/

Webster, R. (2024). Teaching on the cheap? The extent and impact of teaching assistants covering classes and leading lessons. *British Educational Research Journal, 50*(6), 2599–2622. https://doi.org/10.1002/berj.4043

Webster, R., & Blatchford, P. (2013). The Making a Statement project Final Report. *Nuffield Foundation.* https://www.nuffieldfoundation.org/sites/default/files/files/mastreport.pdf

Webster, R., & Blatchford, P. (2017). The Special Educational Needs in Secondary Education (SENSE) study Final Report. *Nuffield Foundation.* https://www.nuffieldfoundation.org/project/special-educational-needs-in-secondary-education-sense

Webster, R., Bosanquet, P., Franklin, S., & Parker, M. (2021). *Maximising the Impact of Teaching Assistants in Primary Schools: A Practical Guide for School Leaders.* Routledge. https://doi.org/10.4324/9781003031420

Webster, R., & de Boer, A. A. (Eds.). (2022). *Teaching Assistants, Inclusion and Special Educational Needs: International Perspectives on the Role of Paraprofessionals in Schools.* Routledge.

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

What Works Clearinghouse. (2024). *Study Review Protocol and Supplement to the What Works Clearinghouse Procedures and Standards Handbook, Version 5.0* (No. Version 5.1). Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/

Appendix

Appendix A: PICO statement

The **population** of interest is pupils in mainstream education aged between four and 18, in any OECD member country. We are interested in pupils with and without SEND but will report detailed subgroup analyses based on SEND.

The **interventions** of interest are educational support activities led by TAs, delivered in one-to-one or small group settings. Interventions should be time-limited, structured, and designed to improve pupil learning outcomes or learning ability. A **TA** is defined as an adult who is employed to support teachers and help pupils with learning activities, but who is not primarily employed by the school as a class teacher. As terminology varies both within the UK and internationally, our definition focuses on the practical function and role that TAs perform in the classroom, rather than their job title, qualifications or employment arrangement. Responsibilities may also vary, but the role should primarily be to support teachers and pupil learning.

The primary **comparator** of interest is 'business as usual' (BAU), defined as no additional educational support beyond the standard curriculum and teaching arrangements.

The primary **outcome** of interest is pupil attainment or progress in any subject, measured using standardised or diagnostic tests (for example, assessing literacy or numeracy), researcher-designed tests, or academic assessments.

Appendix B: Literature search and screening

Literature search terms (with wildcards)

Intervention:

tutor*, small group*, academic support*, educational support*, learning support*, intervention*, structured support*, time-limited intervention*, one-to-one support, small-group intervention*

Population:

pupil*, student*, child*, children, youth*, adolescent*, grade*, school*, classroom*, education*, primary, secondary, elementary, early years

Comparison:

business as usual, BAU, no additional support, no intervention, regular instruction, standard practice, usual care, typical practice

Teaching assistant role:

teaching assistant*, TA*, support staff*, classroom assistant*, learning assistant*, paraeducator*, teacher aide*, education assistant*, instructional assistant*, educational support staff*, higher level teaching assistant*, HLTA*, learning support staff*, learning support assistant*, LSA*, non-teaching assistant*, special needs assistant*, SNA*, classroom support staff*, learning support practitioner*, teaching support staff*, behavior support teaching assistant*, minority ethnic pupil support staff*, bilingual support staff*, ethnic assistant*, student support staff*, learning mentor*, resource teacher*, instructional aide*, paraprofessional*, educational assistant*, integration aide*, aboriginal education officer*, indigenous education worker*, inclusion assistant*

Outcome:

learning*, comprehension*, understanding*, retention*, transfer, reading*, problem solving, instruction*, recall*, performance*, knowledge*, literacy*, working memory, academic achievement*, educational attainment

Research type:

RCT, randomi?ed controlled trial, random trial, random study, random controlled, intervention*, control group*

Exclusions:

“higher education” OR “adult education”

Table S1. Full inclusion and exclusion criteria for screening

Criteria	Definition	Examples
Year	Study was carried out or published between 19941 and 2024.	N/A
Language	Study is fully reported in English.	
Country	Intervention was conducted in an OECD member country.	Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Türkiye, United Kingdom, United States.
Type of article	Article is a peer-reviewed research article, doctoral (PhD) thesis, working paper or institutional report.	Include journal articles, working papers, institutional reports or doctoral (PhD) theses. Exclude conference papers/proceedings, extended abstracts, master's theses, or books.
Type of research	Article reports primary research.	Exclude reviews, meta-analysis, and discussion/opinion articles.
Population	Study is conducted on school pupils aged 4-18.	Examples to include: • **UK**: Reception to Year 13 (Ages 4-18) • **US**: Kindergarten to Grade 12 (Ages 5-18) • **Australia**: Foundation Year to Year 12 (Ages 5-18) • **Canada**: Kindergarten to Grade 12 (Ages 5-18) • **New Zealand**: Year 1 to Year 13 (Ages 5-18) • **European Union**: Primary to Upper Secondary Education (Ages 6-18)

Criteria	Definition	Examples
Teaching assistant (TA)	Intervention is conducted (partially or fully) by a TA, based on the following definition of a TA: An adult who is employed to support teachers and help pupils with learning activities, but who is not primarily employed by the school as a class teacher.	We will search for and include articles using the following terms for 'TA': teaching assistant, TA, support staff, classroom assistant, learning assistant, paraeducator, teacher aide, education assistant, instructional assistant, educational support staff, higher level teaching assistant, HLTA, learning support staff, learning support assistant, LSA, non-teaching assistant, classroom support staff, learning support practitioner, teaching support staff, behaviour support teaching assistant, minority ethnic pupil support staff, bilingual support staff, ethnic assistant, student support staff, learning mentor, resource teacher, instructional aide, paraprofessional, educational assistant, integration aide, aboriginal education officer, indigenous education worker, inclusion assistant Include interventions carried out by both TAs and other volunteers/ teachers. Include support staff employed specifically to support pupils with SEN. Include TAs who fit the definition regardless of their qualifications. Exclude support staff not directly involved in pupils' learning, for example: • Technician • Welfare officer • Counsellor • Psychologist Exclude the following support staff: • Speech and language therapist • Educational psychologist Exclude if researchers or university students carried out the intervention. Exclude if no details are given about who carried out the intervention.
Intervention type	Intervention involves educational support activities, delivered in one-to-one or small group settings. Interventions should be time-limited, structured, and designed to improve pupil learning outcomes or abilities.	Exclude whole-class interventions, unless they are delivered exclusively to small groups or individuals within the class setting. Exclude interventions not directly related to learning or learning ability (for example, targeting behaviour or pupil attitudes). Include interventions on any subject area (for example, literacy, numeracy, maths). Include interventions that aim to improve learning abilities (for example, working memory, metacognition, language).

Criteria	Definition	Examples
Outcomes	Study includes at least one measure of pupil attainment or progress in any subject area, measured using standardised or diagnostic tests, researcher-designed tests, or academic assessment.	Include formal and formative classroom tests, for example, SATs, spelling tests. Include standardised diagnostic tests measuring pupil attainment – for example, Clinical Evaluation of Language Fundamentals (CELF; literacy), British Picture Vocabulary Scales (BPVS; literacy), Wechsler Individual Achievement Test (WAIT; reading, language and numerical attainment), New Group Reading Test (literacy), Progress in Maths test. Include researcher-designed tests that measure pupil attainment or progress. Exclude studies that only report outcomes not directly measuring pupil attainment – for example, behaviour, engagement, attitudes. Exclude studies that compare intervention outcomes to previous years' results or historical data rather than to a contemporaneous control group.
Research setting	Intervention is carried out in a mainstream primary or secondary school or further education college (or non-UK equivalent).	Include kindergarten/reception. Include SEND classes or settings within mainstream environments. Exclude special schools. Exclude higher/adult education. Exclude nursery education (or equivalent).
Research design	Study is a randomised controlled trial (RCT)	Include studies randomised at the school, class or individual level. Include studies with a matched-pair design as long as the pairs are randomised. Exclude case studies and quasi-experimental designs. Exclude non-randomised studies. Exclude studies that do not have a control group (pre-post only designs/within participant designs).
Comparator	Intervention is compared to 'business as usual' (BAU; no additional educational support beyond the standard curriculum) or no intervention.	Include studies where the intervention is compared to BAU, defined as no additional educational support beyond the standard curriculum and teaching arrangements. Include studies with a waiting group design where pupils on waiting list experience BAU. Exclude studies without a BAU control group.

Appendix C. Equations

To convert η_p^2 effect size to Cohen's *d* (https://haiyangjin.github.io/2020/05/eta2d/):

$$d = \sqrt{((N-1) / N * \eta_p^2 / (1-\eta_p^2))}$$

- where *N* = total sample size

To calculate the *SE* of Hedges' *g* from 95% CIs:

$$SEg = (CI_upper - CI_lower) / 3.92$$

To calculate the standard error *(SE)* of Hedges' *g* (Borenstein et al., 2021) based on the sample size:

$$SEg = \sqrt{((n1 + n2) / (n1 * n2)) + g^2 / (n1 * n2)}$$

- where n1 = intervention group sample size, and n2 = control group

To calculate 95% CIs of *g*:

g +/- 1.96 x SE_g

To calculate the variance of Hedges' *g*:

$Varg = SEg^2$

Appendix D: Subgroup analyses of pupil knowledge and skills

Table S2. Effects of each intervention on different learning outcomes

Learning domain	Learning outcome (measure)	Primary/ secondary outcome?	Estimated effect size, *g* [95% CIs]	RCTs / articles
1st Class @Number				
Maths	Number knowledge and problem solving (Quantitative Reasoning Test)	Primary	0.18† [-0.08, 0.43]	Nunes, 2018
	Reasoning and arithmetic (KS1 maths)	Secondary	0.04† [-0.18, 0.28]	
ABRA: online				
Literacy	Reading (PiRA)	Primary	0.07 [-0.07, 0.21]	Bell, 2022 McNally, 2016
	Phonological decoding and sight word recognition (DTWRP)	Secondary	0.10 [0.01, 0.19]	
	Letter sound knowledge (LeST)	Secondary	0.19 [0.09, 0.30]	
ABRA: paper				
Literacy	Reading (PiRA)	Primary	0.20 [0.10, 0.29]	Bell, 2022 McNally, 2016
	Phonological decoding and sight word recognition (DTWRP)	Secondary	0.13 [0.05, 0.22]	
	Letter sound knowledge (LeST)	Secondary	0.15 [0.04, 0.25]	
Catch Up Literacy				
Literacy	Reading (NGRT and HGRT)	Primary	0.08 [-0.07, 0.23]	Roy, 2019 Rutt, 2015
	Sentence reading accuracy (SSRT)	Secondary	0.04† [-0.05, 0.12]	Roy, 2019
	Sentence reading comprehension (SSRT)	Secondary	0.04† [-0.10, 0.18]	
Catch Up Numeracy				
Maths	Numeracy and mathematical understanding/maths age (BNST)	Primary	0.35 [0.03, 0.68]	Holmes (2013) Rutt (2014)

Learning domain	Learning outcome (measure)	Primary/ secondary outcome?	Estimated effect size, *g* [95% CIs]	RCTs / articles
Developing Vocabulary				
Literacy	Reading (NGRT)	Primary	0.38† [-0.01, 0.76][a]	Cockerill, 2021
	Reading accuracy (NGRT sentence completion subscale)	Secondary	0.36† [-0.03, 0.74][a]	
	Reading comprehension (NGRT subscale)	Secondary	0.42† [0.03, 0.80][a]	
Fresh Start				
Literacy	Reading (NGRT, KS2 reading)	Primary	-0.12 [-0.19, -0.04]	Gorard, 2015a Molotsky, 2022
	Writing (KS2 writing)	Secondary	0.02† [-0.06, 0.10]	Molotsky, 2022
Lightning Squad				
Literacy	Reading comprehension (WJ3 Passage Comprehension)	Primary	0.12 [-0.10, 0.35]	Neitzel, 2023 Ross, 2017
	Decoding and letter-sound knowledge (WJ3 Word Attack)	Primary	0.20 [0.03, 0.37]	
	Word and letter reading (WJ3 Letter-Word Identification)	Primary	-0.02 [-0.19, 0.15]	
Maths Counts				
Maths	Maths attainment (general maths, CEM InCAS)	Primary	0.09†‡ [-0.14, 0.32][a]	See, 2018
Nuffield Early Language (nursery-reception version)				
Post-test outcomes				
Literacy	Spoken language (composite)	Primary	0.39 [0.08, 0.70]	Fricke, 2013; 2017
	Reading (composite)	Secondary	0.15 [-0.08, 0.39]	
6-month follow-up				
Literacy	Spoken language (composite)	Primary	0.40 [0.23, 0.58]	Fricke, 2013; 2017 Sibieta, 2016b
	Reading (composite)	Secondary	0.20 [-0.11, 0.3Z1]	

Learning domain	Learning outcome (measure)	Primary/ secondary outcome?	Estimated effect size, *g* [95% CIs]	RCTs / articles
Nuffield Early Language (reception version)				
Post-test outcomes				
Literacy	Spoken language (composite)	Primary	0.24 [0.16, 0.32]	Dimova, 2020 Fricke, 2017 Sibieta, 2016b West, 2021
	Reading (YARC early word reading; reading composite)	Secondary	0.14 [0.02, 0.27]	
6-month follow-up				
Literacy	Spoken language (composite)	Primary	0.21† [0.02, 0.41]	Fricke, 2017 Sibieta, 2016b
	Reading (composite)	Secondary	0.07† [-0.15, 0.28]	
Reading and Language Intervention for Down syndrome				
Literacy	Reading (YARC early word reading, letter-sound knowledge and phoneme blending)	Primary	0.39 [0.08, 0.71]	Burgoyne, 2012
	Vocabulary (taught and non-taught expressive and receptive vocabulary)	Primary/ secondary	0.20 [-0.05, 0.45]	
	Spoken language (RAPT: expressive grammar and information)	Secondary	0.12 [-1.04, 1.28]	
Reciprocal Reading				
Literacy	Reading (NGRT)	Primary	0.15 [0.05, 0.24]	O'Hare 2019; 2023 Thurston, 2020
	Reading comprehension (NGRT subscale)	Secondary	0.17 [0.07, 0.27]	
Switch-on Reading				
Literacy	Reading (HGRT and NGRT)	Primary	0.004 [-0.12, 0.13]	Gorard, 2014; 2015b Patel, 2017
Switch-on Reading and Writing				
Literacy	Reading (HGRT)	Primary	0.05† [-0.10, 0.19]	Patel, 2017

Learning domain	Learning outcome (measure)	Primary/ secondary outcome?	Estimated effect size, *g* [95% CIs]	RCTs / articles
VEIP/NIP combined				
For language and communication difficulties				
Literacy	Verbal narrative skills (composite)	Primary	0.30† [0.12, 0.47]	Joffe, 2023
	Vocabulary (composite)	Primary	-0.04† [-0.21, 0.14]	
For reading difficulties ('Talk for literacy')				
Literacy	Reading (NGRT)	Primary	0.20† [-0.02, 0.42]	Styles, 2015a
	Reading comprehension (NGRT subscale)	Secondary	0.25† [0.02, 0.48]	
	Spoken language (CELF sentence recall)	Secondary	-0.04† [-0.28, 0.20]	
Units of Sound				
Literacy	Reading (NGRT)	Primary	-0.08† [-0.27, 0.11]	Sheard, 2015
	Spelling (SWST)	Secondary	-0.07† [-0.27, 0.13]	

Note. Effect sizes are pooled using random effects meta-analysis, except:

† effect size as reported in publication (not pooled);

‡ adjusted effect size as reported in publication: Fresh Start – adjusted for pre-test scores and pupil age; Maths Counts – adjusted for pre-test scores and prior attainment;

[a] indicates 95% CIs were calculated using equation $g \pm 1.96*SE$.

KS1 = Key Stage 1; PiRA = Progress in Reading Assessment; DTWRP = Diagnostic Test of Word Reading Processes; LeST = Letter Sound Test; NGRT = New Group Reading Test, HGRT = Hodder Group Reading Test; SSRT = Salford Sentence Reading Test; BNST = Basic Number Screening Test; WJ3 = Woodcock-Johnson III; YARC = York Assessment of Reading Comprehension; RAPT = Renfrew Action Picture Test; CELF = Clinical Evaluation of Language Fundamentals; SWST = Single Word Spelling Test.

A great teacher changes the future every day. They can be the critical factor in a child's success, especially for those who have had a tough start in life. At Ambition Institute, we support teachers and school leaders at every stage of their careers, helping them to keep getting better. We are a charity providing training and professional development based on the most rigorous research and evidence about what really works. Together, we're shaping the future of education to give every child the best start in life.

Learn more

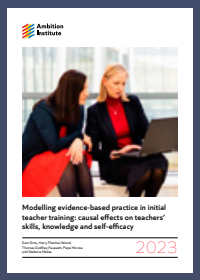
Get in touch today to find out how you can nominate your staff, or yourself, for one of our programmes.

ambition.org.uk
info@ambition.org.uk
020 3668 6865
@Ambition_Inst
@Ambition.Inst
Ambition Institute

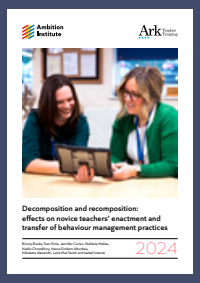
Ambition Institute
2 Bridge Wharf,
156 Caledonian Road,
London, N1 9UU

Discover more

Evidence is at the heart of what we do at Ambition Institute. Through our research we want to push forward our collective knowledge of effective professional development and share what works, to turn evidence into action. We aim to answer important questions faced by the sector, help teachers and leaders to address disadvantage, and show the impact of evidence-informed teaching.



Does adding modelling to professional development help to bridge the theory-practice gap? This research looked at whether video models helped initial teacher trainees to develop their retrieval practice skills.



This research explored whether breaking down (decomposing) and then recombining teaching practices into a different sequence of teaching (recomposing) helped initial teacher trainees apply what they learnt to new contexts.

Ambition Institute

© Ambition Institute 2025. Ambition Institute is a registered charity, number 1146924.