Can teaching assistants improve attainment and attitudes of low performing pupils in numeracy? Evidence from a large-scale randomised controlled trial

Jeremy Hodgen, Michael Adkins & Shaaron Elizabeth Ainsworth

Published online: 20 Oct 2022.

Submit your article to this journal ⌷

Article views: 39

View related articles ⌷

View Crossmark data ⌷

Can teaching assistants improve attainment and attitudes of low performing pupils in numeracy? Evidence from a large-scale randomised controlled trial

Jeremy Hodgen [a], Michael Adkins [b] and Shaaron Elizabeth Ainsworth [b]

aDepartment of Curriculum, Pedagogy and Assessment, IOE, UCL Institute of Education, University College London, London, UK; bSchool of Education, University of Nottingham, Nottingham, UK

ABSTRACT

The use of teaching assistants (TAs) is widespread in many education systems, but the ways that TAs can support learning effectively are poorly understood. Much evidence indicates that most TA support has no, or negative, effects on pupil attainment. A small but growing body of evidence shows that structured TA intervention programmes can be effective. This paper reports the results of a large-scale randomised controlled trial of Catch Up® Numeracy, an intervention delivered one-to-one by TAs, compared to a control in which TAs provided matched-time numeracy support. The trial involved 1794 low-attaining pupils (aged 7–10) and 300 TAs from 150 English primary schools. Pupils in the intervention group showed no gains in attainment compared to the matched-time control, but there was some evidence to suggest a positive impact in pupils' attitudes. The implications for future TA interventions to address low attainment in numeracy are discussed.

Introduction

Over the past 20 years, the use of teaching assistants (TAs), or educational paraprofessionals, has dramatically increased in the UK and in many other education systems across the developed world. In England, the numbers increased more than fourfold from 60,600 in 1997 to 265,600 in 2016 (Blatchford, Bassett, Brown, & Webster, 2009; Department for Education, 2017), and schools now spend more than £4.4 billion on TAs (Sharples, Webster, & Blatchford, 2015). In the US, the numbers increased from 357,000 in 2006 to 1,308,000 in 2016 (Bureau of Labor Statistics, 2018; Butt, 2016). Similar, but less dramatic, increases have been reported elsewhere, including Australia (Butt, 2016) and Finland (Takala, 2007).

Whilst TAs now play a very major role in education, ways of deploying and using TAs to support learning effectively are poorly understood. TAs are frequently asked to support low-attaining pupils or pupils with special educational needs, often in one-to-one settings, and appear to be increasingly taking a teaching role (Warhurst, Nickson, Commander, & Gilbert, 2013). This 'conventional wisdom' has been

questioned by Giangreco (2010), who suggests that this results in reduced qualified teacher contact time for pupils with the greatest educational need. However, two recent meta-analyses indicate that some structured TA-led interventions can be effective (Dietrichson, Bøg, Filges, & Klint Jørgensen, 2017; Pellegrini, Lake, Neitzel, & Slavin, 2021), so it is important to establish what differentiates successful from less successful approaches.

In this paper, we report the results of a randomised controlled trial (RCT) of Catch Up® Numeracy, an intervention programme in which teaching assistants (TAs) provide regular numeracy support to pupils in Years 3, 4 and 5 (ages 7–10).[1] The trial, conducted in England, compared the effects of the intervention on pupil attainment and attitudes and was compared to a matched-time control in which TAs provided numeracy support to pupils for an equivalent amount of time. The focus of Catch Up® Numeracy is on improving numeracy, understood as numbers and arithmetic. Although the terms are often used interchangeably in the literature, mathematics is broader.

Background

A great deal of evidence indicates that TAs commonly work with lower-attaining pupils, often in one-to-one or small grouping tutoring settings (e.g. Webster, Blatchford, & Russell, 2013). Broadly, this evidence indicates that TAs are not deployed in effective ways and, in general, the receipt of TA support has no, or even negative, effects on pupil attainment. Gerber, Finn, Achilles, and Boyd-Zaharias (2001) report an analysis of the effects of TAs from Tennessee's Project STAR, a longitudinal experiment in which over 6000 pupils were assigned at random to small classes, regular-size classes without a TA or regular-size classes with a full-time TA. The results indicate that TAs have little, if any, positive effect on pupils' academic attainment.

The largest study to date in the UK was the Deployment and Impact of Support Staff (DISS) project, which was carried out in England and Wales between 2003 and 2008 and involved over 8000 pupils from 76 primary and secondary schools (Blatchford et al., 2009). This study examined the impact on all students receiving support, but the sample was heavily weighted towards low-attaining and disadvantaged pupils. The research found that pupils in receipt of TA support received less teacher contact time in comparison to other pupils (Blatchford et al., 2009). Although TA support was found to have some benefits for pupil engagement, pupils in receipt of TA support made less academic progress than other similar pupils and, moreover, the trend was for those with more TA support to make less progress than pupils with less support of similar prior attainment or level of special educational need (Blatchford et al., 2011). Drawing on this and other studies, '' Higgins et al.'s (2021) meta-analysis concludes that the general 'unstructured' deployment of TAs in classroom settings has no impact on children's learning.

However, there is evidence indicating that targeted TA interventions and support can positively benefit pupils' attainment, although the evidence is stronger for reading than numeracy. Farrell et al.'s (2010) systematic review of nine intervention studies targeted at pupils with difficulties found that tutoring by TAs, if trained and supported to deliver a targeted intervention, can help primary pupils to make statistically significant gains in attainment, either working on a one-to-one basis or with a small

group. However, only two of the studies involved numeracy, and one of these, Muijs and Reynolds (2003), found no academic benefits for low-attaining pupils in receipt of TA support.

In the US, a number of structured interventions involving scripted sessions by TAs have shown positive effects on low-attaining children's attainment, including Building Blocks (Clements et al., 2011) and Number Rockets (Gersten et al., 2015). In England, several recent trials of TA-led interventions in maths have shown promise. For example, ''See et al.'s (2015) study found Number Counts, a personalised one-to-one programme targeted at low-attaining primary children, to have a positive effect on children's mathematics skills (ES = 0.12), based on an efficacy trial involving 35 schools. Nunes et al.'s'' (2018) evaluation of 1stClass@Number, a structured intervention delivered by TAs to small groups of low attainers, found a positive effect on children's quantitative reasoning (ES = 0.18). However, the effect was not statistically significant.

This body of evidence showing promise for structured TA interventions is synthesised in three recent meta-analyses, which include the majority of the recent studies conducted in England, alongside other international studies. All three indicate positive benefits of tutoring programmes. '' Higgins et al.'s (2021) meta-analysis found a positive effect for structured TA programmes (ES = 0.35), based on 65 studies. However, they found the effect to be smaller and less robust for mathematics compared to literacy. Pellegrini et al.'s (2021) best evidence synthesis of effective programmes in primary mathematics found positive benefits for targeted tutoring interventions based on 23 studies. Somewhat counter-intuitively, this meta-analysis found that tutoring by TAs was as beneficial as tutoring by qualified teachers. Although this difference was not statistically significant, small group interventions had slightly better results (Cohen's d = 0.30) than one-to-one tutoring (d = 0.19). Dietrichson et al.'s (2017) meta-analysis of interventions in reading and mathematics aimed at raising attainment for pupils with low socio-economic status found that tutoring by adults had the largest effect size of all 15 intervention types examined (d = 0.36). This finding was based on 36 studies of mainly structured and well-described, or manualised, tutoring interventions delivered by TAs, volunteers or trained teachers. Dietrichson et al. (2017) also found a greater frequency of intervention to be associated with a small positive effect (d = 0.02 per additional 10 sessions), but a longer overall duration to be associated with a small negative effect (d = −0.09 per additional 10 weeks).

These meta-analyses highlight several limitations in the current body of evidence. The studies included in Higgins et al.'s (2021), Pellegrini et al.'s (2021) and Dietrichson et al.'s (2017) reviews are mainly of manualised interventions that provide structured support for pupils. In most cases, the level of support is much greater than that generally provided in schools. Most of these studies also include training for TAs delivering the intervention to pupils and, as a result, the studies evaluate the entire package of each intervention: TA support and the professional development (PD) for TAs. Because these different elements are not disaggregated, it is not known how these different elements impact on pupils' mathematics learning. Moreover, although the studies included in these meta-analyses were judged to be of high quality, the majority of the original studies were either small-scale or efficacy studies (i.e. evaluations under ideal or favourable conditions), involved

an immediate (rather than a delayed) post-test, and were compared to a 'business-as-usual' control group in which no additional support is provided to pupils. Typically, they were evaluated over 10–12 weeks.

The effect of TA support on pupils' attitudes towards numeracy is less well understood, although TA support appears to be widely believed to improve attitudes by both teachers (e.g. Farrell et al., 2010) and parents (e.g. Woolfson & Truswell, 2005). Blatchford et al.'s (2009) study in England found evidence of increased engagement based on systematic observations in 49 schools, whilst ' See et al.'s (2019) evaluation of Number Counts reported positive effects on children's attitudes, based solely on post-test data. However, both Farrell et al.'s (2010) and Giangreco et al.'s (2010) reviews reported mixed effects on pupils' engagement and their social and emotional development. Given these mixed findings, there is a need for a robust evaluation of the impact of TA support on pupils' attitudes towards numeracy or mathematics.

Previous research evidence for Catch Up® Numeracy

Catch Up® Numeracy is a research-based intervention delivered by TAs, targeted at pupils who are low attaining in numeracy, which is designed to be relatively inexpensive and straightforward for schools to use (Holmes & Dowker, 2013). Evidence suggesting that Catch Up® Numeracy may be effective comes from an earlier independent efficacy trial (Rutt, Easton, & Stacey, 2014) and a quasi-experimental study conducted by the developers of the intervention (Ann Dowker and colleagues). Neither study investigated the impact of the intervention on pupil attitudes directly, although Rutt et al. (2014) found that some teachers reported improvements to pupil engagement and attitudes.

The most robust evidence for attainment comes from the efficacy trial, which was independently evaluated by Rutt et al. (2014). A total of 336 pupils in Years 2 to 6 (ages 6–11) from 54 schools participated in a three-arm randomised controlled trial. Each was randomly assigned within their school to one of three groups: the Catch Up® Numeracy intervention; a 'matched-time' group in which they received the same amount of one-to-one maths instruction with a TA but not using Catch Up® Numeracy; and a 'no-intervention' group where pupils received no additional TA support beyond normal classroom instruction. All three groups sat the Basic Numeracy Screening Test (BNST, Gillham & Hesse, 2001) before and after the intervention. After attrition, the analysis involved 108, 102 and 108 pupils for each of the Catch Up®, matched-time and no-intervention groups respectively. Using an intention-to-treat analysis, the independent evaluators found that Catch Up® and the matched-time group made greater progress when compared to the business-as-usual control: effect sizes (Hedges' g) of +0.21 (CI 0.42–0.01) for Catch Up® and +0.27 (CI 0.49–0.06) for the matched-time group. These differences were statistically significant, but there was no statistically significant difference between the Catch Up® and the matched-time groups. However, there are limitations to this study. First, there may have been some cross-contamination between the Catch Up® and matched-time groups because delivery was by TAs within the same schools. Second, as an efficacy trial, the intervention was evaluated under favourable, not 'real world' conditions. For example, a trainer visited each school to support implementation, and all pupils received the same intended duration of intervention. Additionally, the trial utilised immediate post-tests rather than delayed post-test.

Consequently, it did not evaluate the intervention on the basis of sustained gains that would provide better evidence that pupils had 'caught up' and so reached an age-appropriate level of numeracy.

The contribution of this study

The evaluation of the specific intervention, Catch Up® Numeracy, is important in itself, since it is used widely in the UK (by around 6000 schools) and further afield in countries such as Australia.[2] Hence, evaluating the effectiveness of the intervention has widespread importance for policy. Additionally, if shown to be effective, it would have the potential to make a practical difference. It is typical of the level of one-to-one support provided for low-attaining pupils by primary schools in England. Additionally, the Catch Up® Numeracy programme is relatively cheap for schools to purchase. Hence, the intervention is manageable and affordable for schools. The current study makes an original and important contribution by addressing the limitations of the two previous studies of Catch Up® Numeracy and including a comparison with an active control group receiving matched-time TA support.

The current study contributes in three further ways. First, given the scale of global investment in TAs highlighted earlier, there is an urgent need to better understand the use of TAs and, in particular, with well-designed experimental studies (Giangreco et al., 2010). As highlighted earlier, there is a need to strengthen the evidence base by examining the effects of TA support under less favourable conditions and with robust designs. Our study is an effectiveness trial using a clustered RCT design where Catch Up® Numeracy intervention is evaluated at scale in 'real world' conditions, compared to an active control with matched-time support and evaluated over a full academic year with delayed post-test of mathematics attainment and attitudes. In addition, in order to minimise any risks of bias, the analyses were pre-registered to avoid 'p-hacking', and the primary analysis was conducted on an 'intention to treat' basis.

Second, it is important to better understand the effect of TA support on pupils' attitudes towards numeracy. The association between attitudes and attainment in numeracy and mathematics is well documented: improvements in pupils' attitudes may contribute to subsequent improvements in pupils' attainment (Dowker, Sarkar, & Looi, 2016). As noted earlier, we were unable to locate any published studies that evaluated the impact of TA support directly on pupils' attitudes.

Third, there is a need to understand how best to enable TAs to deliver effective numeracy support. As noted earlier, there are no studies that attempt to disentangle the effects of the different elements of TA interventions. In particular, it is not known how crucial such training and materials are, or, alternatively, whether structured support on its own is sufficient. Giangreco et al.'s (2010) review raises serious questions about the use of TAs as replacements for teachers, as TAs' subject knowledge and pedagogical expertise are generally more limited than that of teachers. However, providing training and teaching materials could overcome these limitations. On the other hand, Blatchford et al. (2011) found that most TA support was unstructured. Hence, it may be sufficient to provide schools access to evidence-based guidance to encourage more thoughtful and structured use of TA support. If some forms of TA intervention are effective, as Pellegrini et al.'s (2021) meta-analysis suggests, this may be a cost-effective approach to supporting

low attainers, even if this is less effective than a teacher-led intervention. Our study addresses these questions through comparing Catch Up® Numeracy, a relatively light-touch intervention, in which the TAs were trained to use a componential approach to identifying and addressing pupils' difficulties in numeracy with a matched-time intervention, in which schools were given published guidance on the best use of TAs, but TAs were not trained in the approach.

The research questions this study was designed to address and reported in this paper were:

(1) Does a numeracy intervention delivered by TAs involving training and session materials have a significant effect on pupil attainment in mathematics when compared to an active control group receiving matched-time TA support (but no training or session materials)?

(2) Does a numeracy intervention delivered by TAs involving training and session materials have a significant effect on pupil attitudes towards mathematics when compared to an active control group receiving matched-time TA support (but no training or session materials)?

The interventions: Catch Up® Numeracy and the matched-time active control

The Catch Up® Numeracy intervention

The following summary is based on a longer description in the evaluation report to the funders (Hodgen, Adkins, Ainsworth, & Evans, 2019) in which the Template for Intervention Description and Replication (TIDieR) checklist format was used (Hoffmann et al., 2014). The intervention is described in detail in the Catch Up® Numeracy (2017) manual.

Catch Up® Numeracy is a manualised intervention targeted at the lowest attaining 15–20% of children for their age who are identified by their school as struggling with numeracy. It involves three elements: TA support adapted to the needs of pupils, training for the TAs and structured record-keeping. The Catch Up® Numeracy intervention was organised independently of the evaluation team by the Caxton Trust operating under the name of Catch Up®, a not-for-profit charity that markets the intervention to schools.[3] Catch Up® ('the developer') were responsible for training, providing materials, monitoring the recording-keeping by intervention schools and the recruitment of schools to the trial prior to randomisation.

The intervention is guided by a componential approach to numeracy (Dowker, 2009). The componential approach is based on research indicating that numeracy is not a single 'big' skill, but a compound of several 'smaller' component skills that appear to be relatively discrete (Dowker, 2005). The Catch Up® Numeracy intervention breaks numeracy down into 10 components, including counting verbally, counting objects, remembered facts, estimation and derived facts (Holmes & Dowker, 2013).

At the start of the intervention, TAs assess pupils' ability on each component. By using a checklist approach based on this assessment, the TA's subsequent instruction is intended to address the pupil's exact areas of weakness (Dowker & Sigley, 2010). All pupils in the intervention were assessed on a termly basis as to their eligibility for the

intervention. Pupils who were judged to have reached an age-appropriate level of numeracy were then rolled off the intervention. Hence, depending on these assessments, pupils receive either one, two or three terms of numeracy support as in the recommended implementation.

TAs deliver one-to-one support to pupils in twice-weekly short (15- minute) structured sessions. Catch Up® Numeracy supplies guidance and web-based materials detailing how each session should be structured, suggestions on its delivery and appropriate tasks. To prepare them for delivering the intervention, TAs receive three half-day training sessions, each involving a follow-up task in school, together with an optional follow-up session partway through the year. The training sessions address the componential approach to numeracy, the assessment of pupils' strengths and weaknesses in numeracy, and the structure and regularity of sessions together with ways of addressing pupil engagement and attitudes. Unlike the previous trial (Rutt et al., 2014), schools did not receive a support visit from a Catch Up® Numeracy trainer. Schools are also required to appoint a member of staff, usually a teacher, as the school's Catch Up® Numeracy Coordinator to monitor and review the delivery of the intervention, ensuring delivery records are up to date, and providing ongoing support for the TAs. To prepare them for their role, coordinators received training alongside the TAs.

The matched-time control group

The matched-time control group were advised to follow the key structural features of the intervention group. Thus, they were asked to provide one-to-one TA numeracy support to pupils. They were told support should be provided in one or more sessions each week totalling 30 minutes support per week for each pupil. Schools were also asked to review progress each term, rolling off approximately one third of the participating pupils at each time point – the end of autumn, spring and summer terms.

In contrast to the intervention, the choice of content and pedagogy of the session was determined by the schools. They were asked to follow 'best practice' principles based on the funders' guidance on making best use of TAs (Sharples et al., 2015). Schools were provided with both paper and electronic copies of this guidance and were advised to organise an information session on it. Some record keeping was required for the trial as schools were asked by the evaluation team (the authors) to complete half-termly records of the frequency and duration of weekly support provided to pupils and any pupils that were no longer receiving support. However, this did not follow the specific format of Catch Up® pupil progress records. TAs were not offered specific training and there was no additional co-ordinator required.

Design and methods

The data are from a two-arm clustered randomised controlled trial on the basis of 'intention to treat', which includes all pupils and schools as originally allocated at randomisation. The interventions were delivered over one academic year between September 2016 and July 2017. The trial was registered in the ISRCTN Registry with registration number ISRCTN15428227 and approved by the University of Nottingham School of Education research ethics committee (Ref: 2016/989/CD). The evaluation data

and design were pre-specified in an evaluation protocol and a statistical analysis plan (Adkins & Hodgen, 2017). Further details are available in the supplementary materials, together with an account of the departures that were made to the pre-specified plan. An extended account and discussion of all the analyses, together with additional sensitivity analyses and robustness checks, is also available (Adkins & Hodgen, 2017; Hodgen, Adkins, & Ainsworth, 2016).

Outcome measures

The pre-registered primary outcome measure was pupil test scores on the Progress Test in Mathematics (PTM) published by GL Assessment (2015), a standardised test of mathematics with a validated age-adjusted score, thus enabling scores for different aged pupils to be equated on a common scale. In order to evaluate whether any gains were sustained, post-tests were administered on a delayed rather than immediate basis between October and December 2017, in the following academic year and at least three months after the end of the intervention. Pupils, who were then in Years 4, 5 and 6, took the appropriate paper and pencil versions of the test: PTM 8, 9 and 10, respectively. These were administered by trained independent invigilators employed by the evaluation team and were then marked by GL Assessment. The invigilators and GL Assessment markers were blind to the allocation of schools to intervention or control.

The pre-registered secondary outcome measure was the pupil score on an attitudes to mathematics survey using an amalgam of four single-scale items that were combined by summing, drawn from Dowker's instruments designed for assessing young children's attitudes to mathematics (e.g. Krinzinger et al., 2007). These were judged to be quick and cost-effective to administer. Similar short attitude scales using similar items also drawn from Dowker's work have been shown to be valid and reliable (Núñez-Peña, Guilera, & Suárez-Pellicioni, 2014).

Pre-tests and other data

Prior attainment of the pupils was assessed using two measures: specially administered PTM tests, and Key Stage 1 (KS1) national test scores.[4] The age-appropriate pencil and paper version of the pre-test, PTM 7, 8 or 9, was administered by schools prior to randomisation, with delivery and marking by GL Assessment. KS1 scores were obtained directly from the National Pupil Database (NPD). These data enable us to assess the extent to which the groups were balanced on attainment as well as the extent to which the schools identified appropriate pupils for the intervention. In addition, given the predictive power of prior attainment, these data were used to improve the precision of the estimates.

To enable analysis of compliance and implementation, data were collected from both intervention and control schools on the frequency of support provided to pupils together with surveys of teaching assistants and other school staff and case studies of both intervention and control schools. A mixed-methods approach was used to investigate the process of implementation using the dimensions and factors identified by Humphrey et al. (2016). The quantitative survey data was analysed descriptively and with inferential statistics. The qualitative data were analysed using thematic analysis (Braun & Clarke, 2006).

Recruitment and randomisation

Schools were recruited to the trial by the developer in four regions in England, the North East, Yorkshire, Peterborough and Havering. All state primary schools were eligible as long as they had not previously purchased Catch Up® Numeracy and could identify 12 pupils from Years 4 and 5 (or alternatively 4 pupils each in Years 3, 4 and 5). Each school identified the pupils, TAs and coordinator prior to randomisation. Schools were advised to identify pupils in the bottom 15–20% of mathematics attainment for their age group based on two criteria: KS1 scores and the professional judgment of teachers.

We present the trial's CONSORT diagram in Figure 1, which shows the flow of schools and pupils through the trial. It can be seen that a total of 151 schools were randomised into treatment and active control after the pre-test had been administered to all identified pupils. Schools were randomised within regional blocks in order to facilitate the delivery of training by the developer.

Figure 1 highlights some attrition, particularly at the pupil level (18%). This is likely to be a result of the 'real-world' nature of this efficacy trial coupled with the relative disadvantage of the target group of low-attaining pupils. These pupils tend to change schools more often and have higher levels of absence compared with their more advantaged peers (Hodgen et al., 2020). Unlike the majority of previous studies, this trial was conducted over an entire school year rather than a shorter 10–12 week period and involved a delayed post-test. It is notable, therefore, that a substantial proportion of the attrition (10%) was due to pupils either having left the school or being absent when the delayed post-test was conducted during the academic year following the trial.

Nevertheless, whilst the level of attrition can be explained, the question remains as to whether the level of attrition should be considered a potential source of bias. In order to address this, we calculated overall and differential attrition across the two groups. From Table 1, it can be seen that attrition is relatively balanced with a differential attrition rate of 0.2 percentage points at pupil level. This is regarded as a tolerable level of attrition at the most conservative 'cautious' boundary in the What Works Clearinghouse (2020, p. 12) standards for RCTs (maximum 5.7 percentage points differential for 18% overall attrition) and, thus, we consider the missing-at-random assumption to be acceptable.

Modelling

The primary analysis estimates the effect of Catch Up® Numeracy against the active control matched-time condition on the basis of intention-to-treat using a four-level linear multi-level model estimated by Bayesian inference. This offers several advantages over classical, or frequentist, methods (Kruschke & Liddell, 2018). First, classical inference relies on an assumption of repeated sampling from an imaginary distribution, whereas Bayesian inference is based on the accumulation of knowledge from actual data together with our prior knowledge. In the case of the trial reported here, where we have limited prior knowledge, the expectation is that point estimates and intervals will be broadly similar between classical and Bayesian approaches. However, Bayesian inference is more straightforward to interpret. Classical inference relies on p-values and confidence intervals to report results, although these tools are often misunderstood even by some experienced researchers (Hoekstra, Morey, Rouder, & Wagenmakers,
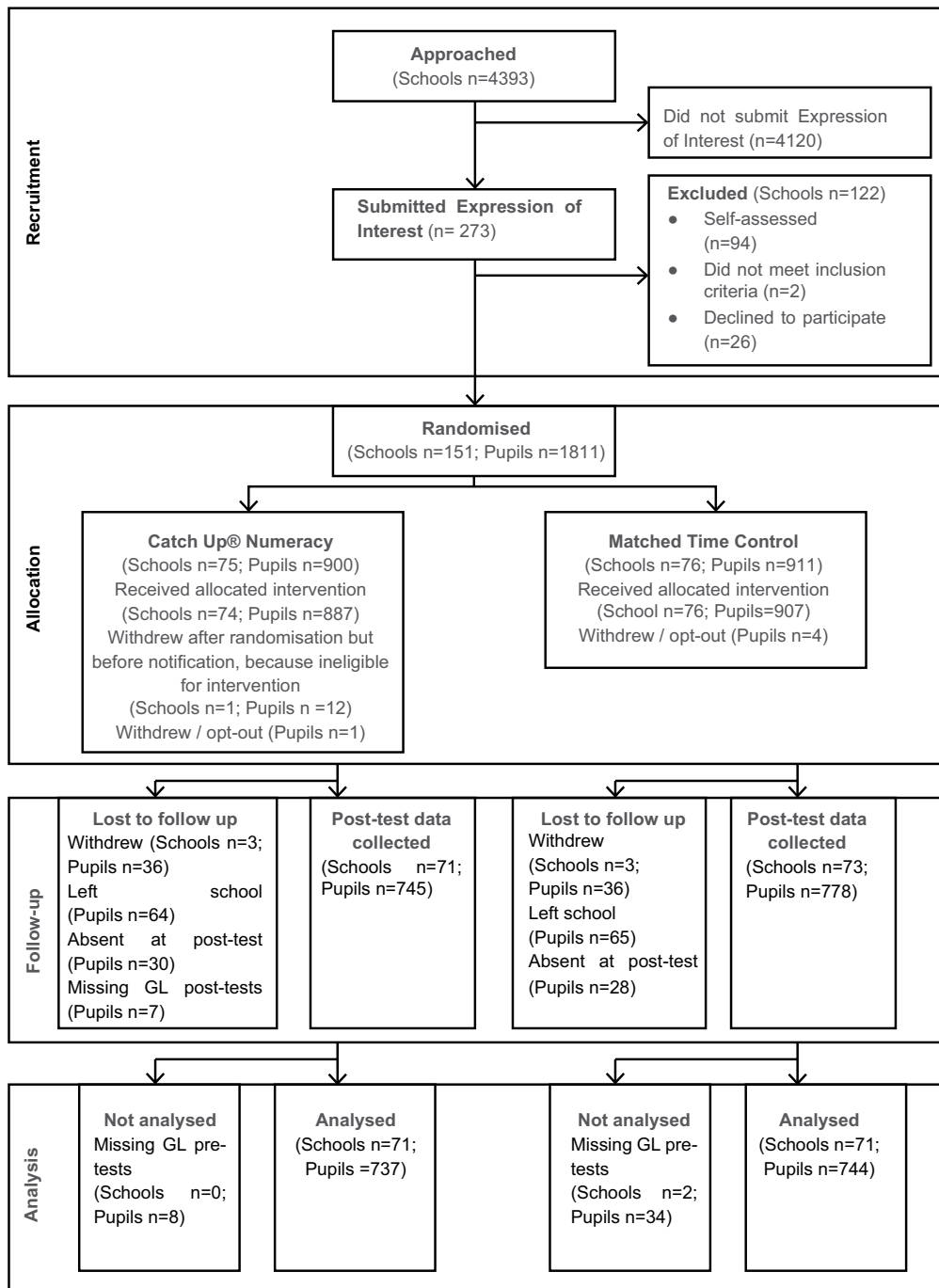
Figure 1. Consort flow diagram showing participation and attrition during the trial.

Table 1. Comparison of overall attrition rates for schools and pupils.

	Combined sample		Catch Up© Numeracy		Matched-time control	
	Schools	Pupils	Schools	Pupils	Schools	Pupils
Randomised	151	1811	75	900	76	911
Total Attrition	9*	310	4*	163	5	167
% Attrition	6.0%	17.1%	5.3%	18.1%	6.6%	18.3%

Note: * Includes one school randomised to the Catch Up© Numeracy intervention arm that was found to be ineligible following randomisation, because it was a special school.

2014). In contrast, the Bayesian equivalents are more intuitive. Credible intervals, the Bayesian equivalent of confidence intervals, are exactly what most people think frequentist confidence intervals are – i.e. that there is 95% probability that the interval captures the value – and, hence, are more straightforward to accurately communicate to a lay audience. Second, Bayesian inference is more informative by providing information on the uncertainty of results. This is particularly important for our current study. Given the findings of the previous trial (Rutt et al., 2014), it is possible that the effects of the intervention and the active control are similar. However, the classical approach of null hypothesis significance testing (NHST) does not provide a framework for addressing this. Within NHST failure to reject the null hypothesis would not provide evidence of no effect, but instead is inconclusive (Dienes, 2014). Bayesian procedures, such as the region of practical equivalence or ROPE (Kruschke & Liddell, 2018), which is described later, provide a framework for establishing, or at least evidence to support, a finding of no difference.

The primary model is as follows:

$$y_{ijkl} = \beta_0 + \beta_1 \text{Treatment}_k + \beta_2 \text{Pre} - \text{Test}_i + f_{0l} + v_{0kl} + u_{0jkl} + \in_{ijkl}$$

$$f_{0l} \sim N\left(0, \sigma^2_{region}\right) for\, l = 1 \ldots L \qquad (1)$$

$$v_{0k} \sim N\left(0, \sigma^2_{school}\right) for\, k = 1 \ldots K$$

$$u_{0j} \sim N\left(0, \sigma^2_{Teaching\, Assistant}\right)\, for\, j = 1 \ldots J$$

$$\in_i \sim N\left(0, \sigma^2\right) for\, i = 1 \ldots N$$

In this four-level model, y is the primary outcome, the PTM standardised score, for pupil *i* supported by TA *j* in school *k* within randomisation region *l*. This is a departure from our pre-specified plan in which we proposed to use several dummy variables for randomisation region. In this revised parameterisation, the randomisation group has been entered as a additional level in order to aid the interpretatability of the intercept. As a sensitivity analysis, the randomisation group was included as school-level variable. This made no significant change to the effect. The individual level of our model has a grand mean of the PTM post-test (represented by β_0), which we allow to vary by the Teaching Assistant, School and randomisation

region indexes (represented by the intercept adjustments u_{0jkl}, v_{0kl}, and f_{0l}); a binary treatment covariate where 0 represents those pupils in schools that were randomised to receive the matched-time intervention and 1 which represents those pupils in schools that were randomised to receive the Catch Up® Numeracy intervention; one normally distributed and mean-centred pre-test covariate (PTM), and lastly an error term (ε_{ijk}). At the group level, we assume all group-level terms are normally distributed with means of 0, and estimate four variance parameters, one for each level. In a minor change from the pre-specified three-level model, randomisation region was set as a group-level variable to keep the comparison group within the regression model as simple to interpret as possible: the average pupil receiving the active matched-time condition, with the average score on the PTM pre-test, clustered with the average TA, in the average school, in the average randomisation region (all mean centred on 0). In addition to our pre-specified model, following standard practice, we estimate four further related models to provide further context which are reported in the supplementary materials, together with details of the sensitivity analyses and robustness checks that we conducted. The secondary analysis estimating the effect on pupil attitudes is based on an equivalent model to the primary analysis using the baseline attitudes survey as a pre-test.

Effect sizes are calculated using total variance as set out in the following formula:

$$ES = \frac{\overline{Y_t} - \overline{Y_c}}{\sqrt{\sigma^2_{region} + \sigma^2_{school} + \sigma^2_{TeachingAssistant} + \sigma^2_y}} \qquad (2)$$

Effect size distributions are simulated using the values from the posterior distribution of the fitted model which allowed us to generate 95% credible intervals. Following ʻʼ Kruschke (2018), the ROPE (region of practical equivalence) was set as half of Cohen's definition of a small effect (i.e. 0.1).

Results

Balance

Although it is expected that randomisation should lead to balance across all observable and unobservable characteristics, there is a risk of differences arising either by chance or due to post-randomisation selection effects (such as non-random attrition). We report school-level and pupil-level characteristics in the treatment and active control groups and the differences between these two groups. In the case of categorical characteristics, these differences are expressed in terms of percentage point differences; in the case of continuous characteristics, these differences are expressed in terms of standardised mean differences.

Sample demographics and characteristics are presented in Table 2. It can be seen that there is an imbalance in the pre-test score for the pre-specified PTM test – a standardised mean difference of 0.21. This is particularly surprising as the KS1 scores across the two groups showed a standardised mean difference of 0.03 and 0.06 at baseline and at analysis, respectively, which suggests that the two groups were reasonably balanced in terms of prior attainment. Nevertheless, in order to address any concerns of bias due to imbalance,

Table 2. Sample characteristics.

School-level (categorical)	Catch Up© Numeracy intervention		Active matched-time control		
	n	Percentage	n	Percentage	
Number of schools in each arm	71			71	
School type:					
LEA Maintained	45	63%	43	61%	
Academy	26	37%	28	40%	
Ofsted rating:					
Outstanding	4	6%	11	15%	
Good	58	82%	49	70%	
Requires improvement	8	11%	9	12%	
Inadequate	1	1%	1	1%	
Rating unavailable	0	0%	1	1%	
Urban vs. rural:					
Urban	56	80%	60	84%	
Rural	15	20%	11	16%	
School-level (continuous)	**n**	**Mean (SD)**	**n**	**Mean (SD)**	
Percentage eligible for FSM (in past six years)	71	30.5 (16.7)	71	26.5 (16.3)	
FSM data unavailable	0		1 (1%)		
Percentage achieving level 4 in KS2 combined scores	69	78.4 (13.1)	69	76.18 (15.9)	
KS2 data unavailable	2 (3%)		2 (3%)		
Pupil-level (categorical)	**n**	**Percentage**	**n**	**n**	**Percentage**
Number of pupils in each arm	737		744		
Eligible for FSM (in past 6 years)	290	39%	261	35%	
FSM data missing	7	1%	9	1%	
Gender (% Female)	378	51%	381	51%	
Gender missing	0		0		
Pupil-level (continuous)	**n**	**Mean (SD)**	**n**	**Mean (SD)**	**Standardised mean difference**
Pre-test score (GL PTM)	737	86.4 (9.8)	744	84.3 (9.5)	0.22
KS1 average points score	693	13.6 (2.4)	713	13.8 (2.4)	0.06
KS1 data missing	45 (6%)		31 (4%)		

sensitivity analyses were conducted with KS1, and other covariates, including gender and FSM status, and including both KS1 and the PTM pre-test.[5] These analyses all indicated practically equivalent results. Hence, this is not judged to be a serious threat to validity.

Primary analysis: the effect on pupil attainment

The results of the main model are presented in Table 3 along with comparisons of the pre- and post-test scores for each group in Table 4. This shows that the effect on pupil attainment of Catch Up® Numeracy over the matched-time control is 0.0 (−0.2, 0.1), with approximately three-quarters of the effect size 95% highest posterior density (HDI) within the ROPE indicating that there is no difference in the impact on pupil attainment between the Catch Up® Numeracy intervention and the matched-time control.

Table 3. Summary of the primary model of the effect on pupil attainment.

Independent variables in the model to predict post-test attainment scores (Progress Test in Mathematics)	Estimate (SD)
Treatment: Allocation to Catch Up® Numeracy intervention (β_1)	−0.3 (0.7)
Pre-test score: Progress Test in Mathematics (β_2)	0.7 (0.0)
Constant (β_0)	84.9 (1.2)
Effect Size	**Estimate (95% credible interval)**
Effect size	−0.04 (−0.21, 0.13)
Region of Practical Equivalence (ROPE): % within ± 0.1 around 0	74%
Number of observations	**N**
Pupils	1481
Teaching Assistants (TAs)	279
Schools	142
Region (randomisation blocks)	4
Variance (σ^2)	**Estimate (SD)**
Region-level (randomisation blocks)	6.0 (17.2)
School-level	6.6 (2.3)
TA-level	10.7 (2.3)
Pupil-level	38.4 (1.6)

Table 4. Comparison of pre- and post-tests of attainment for Catch Up© Numeracy and matched-time control.

	Catch Up© Numeracy	Matched -time control	Absolute mean difference	Standardised mean difference
N (Pupils)	737	744		
	Mean (SD)	Mean (SD)		
Pre-test attainment score (GL PTM)	86.38 (9.80)	84.28 (9.46)	2.10	0.22
Post-test attainment score (GL PTM)	85.70 (10.48)	84.49 (10.10)	1.21	0.12

Table 5. Summary of the model of the effect on pupil attitude.

Independent variables in the model to predict post-test attitude survey scores	Estimate (SD)
Treatment: Allocation to Catch Up® Numeracy intervention (β_1)	0.2 (0.1)
Pre-test attitudes (β_2)	0.3 (0.0)
Constant (β_0)	7.5 (0.2)
Effect size	**Estimate (95% credible interval)**
Effect Size	0.1 (−0.04, 0.29)
Region of Practical Equivalence (ROPE): % within ± 0.1 around 0	38%
Number of observations	**N**
Pupils	1093
Teaching Assistants (TAs)	257
Schools	132
Region (randomisation blocks)	4
Variance (σ^2)	**Estimate (SD)**
Region-level (randomisation blocks)	0.1 (0.3)
School-level	0.2 (0.1)
TA-level	0.2 (0.1)
Pupil-level	1.9 (0.1)

Secondary analysis: the effect on pupil attitudes

The results of the secondary model are presented in Table 5 along with comparisons of the pre- and post-test scores for each group in Table 6. This shows an effect of 0.1 (−0.1, 0.3). As the uncertainty intervals highlight, this effect is not bounded away from 0. Nevertheless, given the substantial proportion of the posterior effect size distribution being greater than 0, this result is worthy of further examination. The ROPE analysis suggests that more than 60% of the 95% Bayesian highest posterior density (HDI) falls outside of ± 0.1 around 0. This provides some weak evidence of a potential positive effect on pupil attitudes for those in receipt of the Catch Up® Numeracy intervention over the matched-time control, although, as can be seen from the comparison of scores in Table 6, this may be partly due to a decrease in attitudes in the matched-time control group.

Implementation and compliance

The analysis of implementation data indicated a high level of compliance from the treatment group in terms of attendance at training, staffing, session delivery and termly review. The use of termly review to 'roll off' pupils assessed as having 'caught up' with their peers was similar across both the treatment and active control groups. However, data from the TA surveys indicated that, whilst almost all (98%) of the Catch Up® Numeracy support was delivered individually, 55% of the support in the matched-time control was delivered in groups. Indeed, surveys indicated that many schools and teachers in both trial arms had a strong preference for group delivery. Exploratory modelling suggested that pupils did not appear to be disadvantaged by group rather than individual delivery.

The surveys indicated that, whilst TAs in both arms reported that they needed time to plan their support, the TAs in active control school TAs were more inclined to feel they had enough time to prepare for sessions: $t(215) = 2.2$, $p = 0.028$, $d = 0.30$. Additionally, survey results revealed that Catch Up® TAs talked less to teachers: only 45% talked to teachers weekly compared to 78% of their control group equivalents. Scarcity of time was the overriding reason for this situation. Underlying these results, there appeared to be two interrelated factors. First, the developers recommended that Catch Up® should be delivered in a quiet space away from the pupil's class. This not only took TAs time to locate, but, when TAs did not physically work in the teacher's class, there was limited communication between teachers and TAs. Second, many participants perceived there was no need to connect the learning in the intervention to pupils' classwork. Thus, the incentive to discuss the pupils' learning was reduced. The compliance data also suggested that the dosage, or total support time, may have been greater in the active control, although this should be treated with caution due to limitations in the data.

The analysis of pupil prior-attainment data indicated that around 40% of pupils had a KS1 score above 15 and, hence, were not low attaining. These low-attaining pupils were spread evenly across the two groups. The 'per-protocol' analysis indicated that there was no practical significant difference in the effect for this group of 'eligible' low-attaining pupils (ES = −0.08; 95% credible interval: −0.27, 0.12). The interaction analysis indicated that there was no relationship between the treatment effect and prior attainment. Hence, these analyses indicate that the inclusion of higher-attaining pupils did not affect the

results. We also note that strictly speaking the Catch Up® Numeracy eligibility criteria covers 'struggling' pupils with poor attitudes to mathematics and, hence, the figure of 40% overstates any non-compliance.

Discussion and conclusion

The findings of the impact evaluation suggest that there is no evidence of the impact of the Catch Up® Numeracy intervention on pupil attainment compared to the active control. Whilst this is disappointing given the earlier signs of promise for the intervention, this finding is in line with the findings of the previous three-arm trial in which Catch Up® Numeracy was compared to two controls: one receiving matched-time support and one with no support (Rutt et al., 2014). Both the Catch Up® Numeracy and the matched-time control group were found to have a statistically significant positive effect compared to the no support group, with effects of a similar size. The impact of the intervention on pupil attitudes was more positive, with weak evidence for a potentially small effect of approximately 0.1, although a longer trial would be needed to evaluate whether this led to subsequent improvements to attainment.

Our study addresses the limitations in previous studies on Catch Up® Numeracy identified earlier. This was a large-scale, robustly designed and pre-registered effectiveness trial of the intervention under 'real-world' conditions, with high levels of compliance, and including delayed post-tests rather than immediate tests, and which avoided the potential problem of cross-contamination between groups in the previous trial by using school-level randomisation. We are furthermore unaware of robust studies like ours that have considered the impact of TA support on pupil attitudes.

Whilst our results are in line with those of the previous trial of Catch Up® Numeracy, they are at odds with a growing body of evidence from meta-analyses showing that structured and targeted tutoring interventions can be effective. It is, however, important to emphasise that we compared the Catch Up® Numeracy intervention to an active matched-time control rather than a passive business-as-usual control. It may be that our results are due to the 'success' of the matched-time control rather than the lack of success of the intervention itself. Although the active control was 'light-touch' and did not provide training or specific materials for pupil support, it was nevertheless structured. Schools were provided with 'best practice' guidance on the use of TAs (Sharples et al., 2015). The implementation analysis indicated that this guidance appeared to have influence practice. For example, 80% of senior leaders in the active control had read this guidance. In addition, typically there is only limited communication between TAs and class teachers (Sharples et al., 2015), but some 78% of the TAs in the active control group reported that they discussed numeracy support with class teachers (compared to 45% of the TAs in the Catch Up® Numeracy group). In addition, unlike the typical use of TAs in schools, this support was targeted through initial assessment, was delivered on a regular basis and involved regular termly review. It is important to note, however, that we believe that some aspects of the research data collection are likely to have helped schools to implement this. In particular, we provided schools with logs to record the support provided to pupils and the results of the termly reviews. In addition, we contacted schools every six weeks asking for these logs. Whilst this 'prompting' to schools was relatively light touch, it was nevertheless a regular reminder to schools of the importance

of maintaining regular TA support for the pupils. Hence, our results suggest that providing training and teaching materials for TAs may be less important in improving attainment in numeracy than simply ensuring that TA support is structured and regular. Further research is need to fully understand whether a light touch intervention of this nature supported by 'best practice' guidance can be an effective way of addressing low attainment in mathematics.

Our study has a number of limitations. First, the Catch Up® Numeracy intervention was compared to an active control. A three-arm trial was originally proposed with an additional 'no support' passive control. This option was judged not feasible due to the large number of schools that would be required to adequately power the trial. As a result, it is not possible to robustly estimate the effects of Catch Up® Numeracy compared to a 'no support' option, although there is some evidence of an effect from the previous trial (Rutt et al., 2014). Second, there was some indication that the matched-time control group may have received a higher dosage of TA support. The case studies of active control schools were relatively limited compared to the intervention schools. More intensive case studies would be required to fully understand how TA support was delivered across this group.

We consider that it is somewhat surprising that the trial did not show an effect of Catch Up® Numeracy on attainment. The intervention is in our view well-designed and strongly informed by the research evidence on children's mathematical development and on approaches to addressing low attainment in numeracy with young children. A central feature of the intervention is that the numeracy support is adapted to the needs of pupils and it may be that, as Pellegrini et al. (2021) suggest, such adaptation is less important than previously thought. The implementation analysis suggested that numeracy support in the matched-time group was better aligned with learning in class and this better alignment may be more effective than tailoring the support to individual needs. It is also possible that the result may relate to the professional development provided to TAs. The intervention depends on TAs implementing the assessment and pedagogy associated with the componential approach to numeracy. Changing educational practice is not straightforward, even when supported by a well-designed programme. Each TA received three half days of training and, whilst this is relatively limited, it is in line with other TA interventions shown to be effective, such as the US-based ROOTS intervention (Clarke et al., 2016). However, in addition, ROOTS TAs receive several coaching visits, a professional development intervention that Kraft, Blazar, & Hogan's (2018) recent meta-analysis has shown to be effective with teachers. Alternatively, the more positive results on pupil attitudes may indicate that it may be more straightforward to positively influence TAs' approaches to attitudes in comparison to TA pedagogy or subject knowledge.

Another explanation concerns the level of support provided. Although the level provided is typical for low attainers in English primary schools, the intensity (two 15-minute sessions per week) is relatively modest. Reflecting on Dietrichson et al.'s (2017) findings regarding frequency, intensity and duration, it may be that a more frequent intervention of greater intensity, but perhaps of shorter overall duration, would be more effective. One attraction of the Catch Up® Numeracy intervention is that it is cheap to implement and a more frequent intervention would be more expensive to implement. However, the evidence from the active control suggested that group delivery did not

appear to disadvantage pupils. This interpretation is supported by Pellegrini et al.'s (2021) meta-analysis and by Clarke et al.'s (2017) study which found no difference between an intervention delivered to groups of two and groups of five pupils. Hence, a more frequent intervention could be delivered in groups (as in the ROOTS intervention, for example, which is delivered by a TA to groups in five 30-minute sessions every day for 12 weeks).

We believe that there is an increasing knowledge base about strategies that could support the design of interventions that support TAs' teaching of numeracy from around the world (Hodgen et al., 2020). Given the size of the contribution of TAs to education and the likely use of TAs and adult tutoring to mitigate the effects of the pandemic on children's education, developing robust trials like ours that evaluate this work to assess what is more or less effective is of considerable importance.

Notes

1. In England, upper primary education is referred to as Key Stage 2 (KS2) and consists of Years 3, 4, 5 and 6, each with pupils aged 7–8, 8–9, 9–10 and 10–11, respectively.
2. See www.catchup.org.
3. Catch Up® is the working name of The Caxton Trust, a charity registered in England and Wales (1072425) and Scotland (SC047557), as well as a company limited by guarantee (03476510). Catch Up® is a registered trademark.
4. These cohorts of pupils took a national test at the end of lower primary, or Key Stage 1 (KS1), at age 7.
5. Since the trial was designed, Allen et al.'s (2018) analysis has raised some concerns about the properties of commercial tests such as PTM. In particular, the age standardisation process involves all raw scores below a certain level being given the same age standardised score, thus creating an artificial floor effect. See supplementary materials for brief details of the additional analysis.

ORCID

Jeremy Hodgen http://orcid.org/0000-0002-9196-4088
Michael Adkins http://orcid.org/0000-0002-5289-7302
Shaaron Elizabeth Ainsworth http://orcid.org/0000-0002-9453-7196

Data availability statement

Data analysed as part of this project have been archived and are available on application to the Education Endowment Foundation Data Archive https://educationendowmentfoundation.org.uk/.

References

Adkins, M., & Hodgen, J. (2017). *Statistical analysis plan for Catch Up® NumeracySAP, Education Endowment Foundation*. Retrieved from https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-numeracy-2015/

Allen, R., Jerrim, J., Parameshwaran, M., & Thompson, D. (2018). *Properties of commercial tests in the EEF database*. London: Education Endowment Foundation.

Blatchford, P., & Baines, P. Eds. 2009. *Sex differences in the structure and stability of children's playground social networks and their overlap with friendship relations*.

Blatchford, P., Bassett, P., Brown, P., Martin, C., Russell, A., & Webster, R. (2011). The impact of support staff on pupils' 'positive approaches to learning' and their academic progress. *British Educational Research Journal*, *37*(3), 443–464.

Blatchford, P., Bassett, P., Brown, P., & Webster, R. (2009). The effect of support staff on pupil engagement and individual attention. *British Educational Research Journal*, *35*(5), 661–686. Retrieved from http://www.informaworld.com/10.1080/01411920902878917

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Bureau of Labor Statistics. (2018). *Occupational Outlook Handbook, Teacher Assistants*. Retrieved from https://www.bls.gov/ooh/education-training-and-library/teacher-assistants.htm

Butt, R. (2016). Teacher assistant support and deployment in mainstream schools. *International Journal of Inclusive Education*, *20*(9), 995–1007.

Catch Up®. (2017). *Catch Up®Numeracy file v46, 30.11.2017*. Thetford, UK: Author.

Clarke, B., Doabler, C. T., Kosty, D., Kurtz Nelson, E., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the efficacy of a kindergarten mathematics intervention by small group size. *AERA Open*, *3*(2), 2332858417706899.

Clarke, B., Doabler, C., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*, *9*(4), 607–634.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics Learned by Young Children in an Intervention Based on Learning Trajectories: A Large-Scale Cluster Randomized Trial. JRME, 42(2), 127–166. 10.5951/jresematheduc.42.2.0127

Department for Education. (2017). *School workforce in England: November 2016. SRF 25/2017*. London: DfE.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(781). doi:10.3389/fpsyg.2014.00781

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.

Dowker, A. (2005). *Individual differences in arithmetic. Implications for psychology, neuroscience and education*. Hove: Psychology Press.

Dowker, A. (2009). *What works for children with mathematical difficulties?* Nottingham: DCSF Publications.

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*(508). doi:10.3389/fpsyg.2016.00508

Dowker, A., & Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *Understanding Number Development and Difficulties, BJEP Monograph Series II*, *7*, 65–81.

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.

Gerber, S. B., Finn, J. D., Achilles, C. M., & Boyd-Zaharias, J. (2001). Teacher aides and students' academic achievement. *Educational Evaluation and Policy Analysis*, *23*(2), 123–143.

Gersten R, Rolfhus E, Clarke B, Decker L E, Wilkins C and Dimino J. (2015). Intervention for First Graders With Limited Number Knowledge. American Educational Research Journal, 52(3), 516–546. 10.3102/0002831214565787

Giangreco, M. F. (2010). One-to-one paraprofessionals for students with disabilities in inclusive classrooms: Is conventional wisdom wrong? *Intellectual and Developmental Disabilities*, *48*(1), 1–13.

Giangreco, M. F., Suter, J. C., & Doyle, M. B. (2010). Paraprofessionals in inclusive schools: A review of recent research. *Journal of Educational and Psychological Consultation*, *20*(1), 41–57.

Gillham, B., & Hesse, K. (2001). *Basic number screening test: National numeracy strategy edition: Forms A and B, for ages 7 to 12 years* (3rd ed.). London: Hodder Education.

GL Assessment. (2015). *PT Maths: Technical Information*. Retrieved from http://www.gl-assessment.co.uk/sites/gl/files/images/PTM%20Technical%20Information_incl%2011T_FINAL.pdf

Higgins, S., Major, L. E., Coleman, R., Katsipataki, M., Kay, J., Henderson, P., Mason, D., Aguilera, A. B. V., Dobson, E., Gascoine, L., Rajab, T., Reardon, J., Stafford, J., & Uwimpuhwe, G. (2021). The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit London: Education Endowment Foundation https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit

Hodgen, J., Adkins, M., & Ainsworth, S. (2016). *Catch Up® Numeracy Evaluation protocol, Education Endowment Foundation*. Retrieved from: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/catch-up-numeracy-2015/

Hodgen, J., Adkins, M., Ainsworth, S., & Evans, S. (2019). *Catch Up® Numeracy: Evaluation report and executive summary*. London: Education Endowment Foundation.

Hodgen, J., Barclay, N., Foster, C., Gilmore, C., Marks, R., & Simms, V. (2020). *Review of evidence on early years and KS1 mathematics teaching*. London: Education Endowment Foundation.

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., . . . Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ: British Medical Journal*, *348*, g1687.

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.

Humphrey, N., Ledrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook*. London: Education Endowment Foundation.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588.

Krinzinger, H., Kaufmann, L., Dowker, A., Thomas, G., Graf, M., Nuerk, H.-C., & Willmes, K. (2007). Deutschsprachige Version des Fragebogens für Rechenangst (FRA) für 6- bis 9-jährige Kinder. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *35*(5), 341–351.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206.

Muijs, D., & Reynolds, D. (2003). The effectiveness of the use of learning support assistants in improving the mathematics achievement of low achieving pupils in primary school. *Educational Research*, *45*(3), 219–230.

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

Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The single-item math anxiety scale: An alternative way of measuring mathematical anxiety. *Journal of Psychoeducational Assessment*, *32*(4), 306–317.

Pellegrini, M., Lake, C., Neitzel, A., & Slavin, R. E. (2021). *Effective programs in elementary mathematics: A meta-analysis* (p. 7). AERA Open. doi:10.1177/2332858420986211

Rutt, S., Easton, C., & Stacey, O. (2014). *Catch Up Numeracy: Evaluation report and executive summary*. London: Education Endowment Foundation.

See, B. H., Gorard, S., & Siddiqui, N. (2015). Best practice in conducting RCTs: Lessons learnt from an independent evaluation of the Response-to-Intervention programme. *Studies in Educational Evaluation*, *47*, 83–92. doi:10.1016/j.stueduc.2015.09.003

See B Huat, Gorard S and Siddiqui N. (2015). Best practice in conducting RCTs: Lessons learnt from an independent evaluation of the Response-to-Intervention programme. Studies in Educational Evaluation, 47 83–92. 10.1016/j.stueduc.2015.09.003

Sharples, J., Webster, R., & Blatchford, P. (2015). *Making best use of teaching assistants: Guidance report*. London: Education Endowment Foundation.

Takala, M. (2007). The work of classroom assistants in special and mainstream education in Finland. *British Journal of Special Education*, *34*(1), 50–57.

Warhurst, C., Nickson, D., Commander, J., & Gilbert, K. (2013). 'Role stretch': Assessing the blurring of teaching and non-teaching in the classroom assistant role in Scotland. *British Educational Research Journal*, *40*(1), 170–186.

Webster, R., Blatchford, P., & Russell, A. (2013). Challenging and changing how schools use teaching assistants: Findings from the effective deployment of teaching assistants project. *School Leadership & Management*, *33*(1), 78–96.

Woolfson, R. C., & Truswell, E. (2005). Do classroom assistants work? *Educational Research*, *47*(1), 63–75.