Tutoring with Alphie

Evaluation report and Executive summary
June 2015

Independent evaluators:
Centre for Effective Education Queen's University Belfast

The Education Endowment Foundation (EEF)

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- Identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;

- Evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale;

- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust, as lead charity in partnership with Impetus Trust (now part of Impetus-The Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Robbie Coleman
Research and Communications Manager
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
SW1P 4QP

p: 020 7802 1679
e: robbie.coleman@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The evaluation was conducted by the Centre for Effective Education (CEE), Queen's University Belfast. Support was provided by the Institute for Effective Education (IEE), University of York. As the IEE have established links with the programme developer, the CEE was appointed to independently oversee the evaluation, conduct the random allocation, analyse the data and produce an independent report. The main fieldwork was organised and collected by the Institute for Effective Education.

The lead evaluator was Dr Andy Biggart.

Contact details:

Dr Andy Biggart
Centre for Effective Education
School of Education
Queen's University Belfast
69/71 University Street
Belfast
BT7 1HL

p: 02890 975946
e: a.biggart@qub.ac.uk

Contents

Executive summary

The project

Tutoring with Alphie is a computer-assisted programme that aims to improve the literacy skills of struggling readers. The programme combines elements of collaborative (or 'cooperative') learning, computer-assisted instruction and small group support. Participating pupils are grouped in pairs and follow a series of activities that seek to improve reading comprehension and fluency of expression. Pupils are supervised by a tutor, most commonly a teaching assistant.

The programme was designed by the charity Success for All and was adapted from a programme which was available in the United States for younger struggling readers (6–7 years).

This project sought to assess the impact of the programme on Year 6 pupils who were at risk of not achieving Level 4a and above in reading at the end of Key Stage 2. 72 pupils from six schools across England completed the trial. Pupils followed the programme on a daily basis for 30 minutes each day over a period of six weeks. Typically four to eight pupils, split into pairs with similar levels of prior attainment, would participate at the same time.

The study was funded by the Education Endowment Foundation as one of 24 projects in a themed round on literacy catch-up at the primary–secondary transition.

Key conclusions
1. The evaluation was unable to provide a secure estimate of the programme's impact on literacy, primarily due to the large number of schools that dropped out of the trial.
2. It would have been beneficial to undertake further programme development, including software testing, prior to trialling the programme in English schools.
3. A number of schools perceived the programme as having some promise and worked with the developers and Success for All to overcome initial difficulties with the software.
4. The programme had been adapted as a catch-up literacy programme for Year 6, but the process evaluation suggested that it may be better targeted at a younger age group.
5. To assess the efficacy of approaches that combine collaborative learning with computer-assisted technology to improve reading comprehension, further evaluation would be required.

Security rating

> *Security rating awarded as part of the EEF peer review process*

Findings from this trial have **very low security**. The trial was set up as a randomised controlled trial, which aimed to compare the progress of pupils who received the programme to similar pupils who did not. The trial was classified as an efficacy trial, meaning that it sought to test whether the intervention can work under ideal or developer-led conditions in ten or more schools. However, a very large number of participating schools dropped out of the project, which substantially reduced the security of the resulting impact estimates.

In total, results from 72 pupils were assessed, compared to 248 pupils who began the project. Fifteen of 21 schools dropped out of the project. The high drop-out rate makes it hard to attribute any improvement to the programme, rather than to chance. It is also possible that the level of drop-out may have reduced the accuracy of the estimate in other ways. For example, if those who dropped out of the programme were on average less engaged, the estimate could overstate the impact of the approach.

Results

- On average, pupils who received Tutoring with Alphie showed a small positive improvement in reading comprehension, compared to pupils in the comparison group. However, because many

schools failed to provide data at the end of the programme, it is not possible it attribute the observed positive impact to the programme itself.

- The software program was an adapted version of a previous tutoring program but substantial technical problems were experienced with the implementation of the new program, particularly in the early stages.
- The technical problems had a major impact on the overall evaluation. No schools managed to implement the programme fully as intended, and a number of schools withdrew from the programme completely due to frustration with the problems experienced with the software. Some schools also experienced problems with the computerised assessments used to measure pupil progress. Only 6 out of the original 21 schools completed the planned tests.
- Many of the difficulties with the software were resolved over time and a number of schools who persisted with the programme perceived it as beneficial to the children's reading. However, they also felt that the current content of Tutoring with Alphie may be better targeted at younger readers than pupils in Year 6 and that this would also provide an opportunity for the school to see potential benefits through improved SATs results.
- Previous research on digital technologies has suggested that overall digital technology can lead to a positive effect on learning outcomes. However, there are also indications that the effectiveness of digital learning technologies depends less on the technology itself, but rather the way in which it is used and how it is combined with different pedagogical approaches.
- There is evidence from a previous trial in the United States to suggest that the approach underpinning Tutoring with Alphie which combines digital technology with cooperative learning can lead to improvements in learning. However, this was based on a younger group of struggling readers than those considered in the current evaluation.
- There remain relatively few rigorous evaluations that look at the combined effect of cooperative learning and computer-assisted instruction on reading comprehension. This makes it difficult to establish the efficacy of this approach for pupils of different ages and further research in this area would be beneficial to the existing knowledge base.

Cost

The cost of the approach as delivered in the trial is estimated at £582 per pupil. This estimate is based on a teaching assistant supporting six students per school. The estimate includes 3.5 days of externally provided training and support from Success for All (UK) delivered on-site (£3,030), website licence fee (£300) and an initial start-up pack, including a programme manual and four electronic key pads (£160) to cater for six pupils and an adult tutor in a group. It does not include direct salary costs, supply cover for training or costs associated with the provision of computers. Given that a large proportion of the cost incurred is related to staff training, the per pupil cost to schools could be substantially reduced if a greater number of pupils followed the programme and were supported by the same teaching assistant, or if a larger number of teaching assistants attended training at the same time.

Group	Effect size	Estimated months' progress	Security rating	Cost
Tutoring with Alphie vs. comparison	+0.11	+2 months[1]	🔒🔒🔒🔒🔒	£££

---

[1] In an earlier version of the report, this was incorrectly reports as 1 month.

Introduction

1.1 Intervention

Tutoring with Alphie is a computer literacy tutoring programme, which pupils use in pairs, monitored by a tutor, for example a teaching assistant. The programme has been designed to combine the strengths of collaborative (or 'cooperative') learning, computer-assisted instruction, embedded multimedia, and tutoring programmes. Pupils are paired according to current literacy attainment and the programme is typically delivered to a group of four to eight pupils, whereby each pair works together at the computer and takes turns to alternate as either a reader or coach, monitored by an adult tutor. Pupils are paired according to an initial assessment of their current reading level that is inbuilt within the software, combined with any tutor assessment of compatibility issues in relation to the pairings.

A range of animated activities are provided through the software focusing on fluency and smoothness of expression. For example, Tutoring with Alphie displays a story and the pupil acting in the role of reader reads aloud while the pupil acting as coach reads along silently. After a few minutes of practice the reader undertakes a timed fluency passage, while the coach uses the mouse to click on any problem words. When the reader gives a response to the computer program the coach with the aid of the software evaluates whether or not it is correct. If the response is incorrect the program provides instructional support and prompts the student to try again. The program also prompts the pupils when it is time to switch roles between reader and coach.

There is some existing evidence of the effectiveness of the US programme upon which Tutoring with Alphie is based, but this represents the first pilot in English schools. The programme in the current evaluation was delivered through pull-out sessions on a daily basis for 30 minutes over a period of six weeks.

The Tutoring with Alphie programme was introduced as a six-week intervention programme for eligible Year 6 pupils individually randomly assigned within schools to receive the intervention across a total of 21 schools.

The delivery and evaluation of this project was funded by the Education Endowment Foundation as one of 23 projects focused on literacy catch-up at the transition from primary to secondary school.

1.2 Background evidence

Tutoring with Alphie is a newly developed programme aimed at Year 6 pupils reading below Level 4a at the end of Key Stage 2. There have been no prior evaluations of the Tutoring with Alphie programme, although the general approach underpinning Tutoring with Alphie has been evaluated through a randomised controlled trial in the United States (Chambers *et al.*, 2011). This trial, however, focused upon a younger group of struggling readers than the Year 6 pupils included in the current evaluation. Chambers *et al.*'s (2011) previous evaluation involved a cluster randomised controlled trial involving 33 high poverty schools and found statistically significant positive effects among US First Grade pupils (6-year-olds) who received the programme when compared to a control group that received traditional one to one tutoring with effect sizes varying between Cohen's d=0.15 and d=0.21 depending on which of the three literacy outcome measures were assessed. This suggested that the approach may be more effective than one to one tutoring alone, although statistically significant effects were not found among Second Grade children on any of the three measures. An additional benefit highlighted by the research was that intervention schools through adopting a group approach were able to tutor more children compared to control schools who were using one to one tutoring (Chambers *et al.*, 2011).

The use of computer-aided Interventions or digital learning technologies has expanded rapidly over the past few decades and different technologies are increasingly being used in schools and other learning contexts as interventions designed to improve learning. However, there has been a limited number of

randomised controlled trials (RCT) examining their effectiveness as an approach. A number of meta-analyses have reviewed the overall effects of computer-aided interventions and have concluded that while digital technologies have an overall positive effect, it is one that is slightly less than other research-based interventions (Higgins et al., 2012; Chan and Slavin, 2012). Slavin et al. (2009) in their Best Evidence review of literacy interventions suggest that traditional computer-assisted software programs on their own have little impact on reading outcomes. One of the few recent UK-based studies to evaluate the impact of computer-assisted software on literacy outcomes found that it had a small positive impact on spelling, but this finding was not statistically significant, whereas it had an overall negative impact on reading scores (Brooks et al., 2006). In reviewing previous evidence of the effectiveness of digital learning interventions syntheses of the literature are, however, faced with a number of challenges due to the limited number of rigorous evaluations conducted to date, the wide variety of different digital technologies available, the pace of technological change, and the different pedagogical practices that can be applied within the technology.

Higgins et al.'s (2012) synthesis of the effectiveness of digital technology, while acknowledging the inconsistency in the evidence to date, makes a number of salient conclusions that relate to the approach adopted with Tutoring with Alphie. These suggest that the effectiveness of digital learning technologies depends less on the technology itself, but rather on the way in which they are used. Approaches that combine digital technology with collaborative learning in groups or pairs appear more effective than individual use. The evidence also suggests it may be particularly effective for low attaining pupils as a catch-up and appears more effective as a short (5–10 weeks) regular intervention, rather than extended over long periods of time. Cooperative learning approaches, based on social interdependence theory, have become one of the dominant instructional practices throughout the world (Johnson and Johnson, 2009), with strong evidence of effectiveness compared to competitive and individualistic approaches across a range of domains including attainment.

In conclusion, various forms of digital technology have become increasingly popular both within the regular school classroom and as part of specific interventions designed to raise academic outcomes among disadvantaged groups. The current evidence derived from research syntheses has suggested it may lead to positive outcomes in raising academic outcomes, but a key determinant is how it is used as an intervention rather than the simple adoption of the technology itself. The Tutoring with Alphie programme contains elements which the prior syntheses of the evidence suggest could represent a promising approach, in particular with its focus on cooperative learning. In addition, a previous trial in the US found it had benefits over one to one tutoring, at least, among First Grade pupils, and as a group approach extended the number of pupils that benefited from an intervention. This prior research was, however, based on a younger group of pupils than the Year 6 pupils included in the current evaluation. The evidence to date on computer-based interventions is largely drawn from the US and there has been a limited number of rigorous evaluations on computer-based interventions in the UK.

1.3 Evaluation objectives

The evaluation was designed as an efficacy trial designed to rigorously test through an individually randomised controlled trial whether the adapted version of the programme could lead to demonstrable improvements in children's reading comprehension among struggling Year 6 pupils in English schools. Pupils were selected on the basis that they were unlikely to achieve a Level 4a in reading by the end of Key Stage 2. It was funded as part of a series of evaluations through funding provided to the EEF by the Department of Education focused upon literacy catch-up programmes for pupils at the end of primary school.

In addition to the outcome evaluation a process evaluation was conducted to assess issues in relation to the tutors' and pupils' experience of the programme and to ascertain whether it could be effectively implemented in primary schools in England.

1.4 Project team

The intervention was led by Success for All (UK), who recruited schools to the intervention and provided the initial training for schools as well as ongoing support. The fieldwork was led by the Institute for Effective Education, with day to day management overseen by Dr Mary Sheard (IEE) under the direction of Professor Bob Slavin. Dr Andy Biggart acted as the Principal Investigator for the independent evaluation supported by Dr Liam O'Hare, Dr Sarah Miller and Professor Paul Connolly.

It is with sadness that the independent evaluation team acknowledges the passing of Dr Mary Sheard. She died after a short illness towards the end of the evaluation. Mary was a delightful colleague to collaborate with during the evaluation and always maintained a positive and professional approach to the research despite some of the particular challenges the current evaluation raised.

1.5 Ethical review

As the IEE were responsible for the main data collection, ethical review was obtained primarily through the Education Ethics Committee at the University of York. Consent was initially sought from headteachers and opt-out consent sought from parents or guardians. Additional ethical review was obtained through the Ethics Committee from the School of Education Queen's University Belfast for the aspects of the fieldwork conducted by the CEE.

Trial registration

The trial was registered with the ISRCTN registry (ref: ISRCTN84996648).

Methodology

2.1 Trial design

The Tutoring with Alphie programme was evaluated through an individually randomised controlled trial, whereby Year 6 pupils were recruited and randomised within schools. Ideally, 12 pupils who were performing below the expected level in reading were to be recruited within each school with half randomly allocated to receive Tutoring with Alphie and the other six pupils to act as a control. As a small-scale trial involving a target of a relatively small number of schools (target n=20) an individualised trial was deemed most appropriate. As pupils received the intervention outside of normal class and were required to login to the software monitored by an adult tutor it was considered that there would be no serious threats of contamination across the control and intervention groups.

Most schools provided 10–12 pupils who met the inclusion criteria (see Section 2.2). The minimum number of pupils recruited at a school was 7 and the maximum number was 33. Where a larger number of pupils were recruited these schools operated multiple sessions with around six pupils in each. In cases where recruitment of eligible pupils led to an uneven number in the intervention group, one group worked as a triad rather than working in pairs.

2.2 Eligibility

Success for All (UK) led the recruitment of schools and in total 21 primary schools agreed to take part in the trial. Schools were recruited throughout England, but although a majority were based in the North (North East and Yorkshire and Humber) other schools were recruited from the South East, South West and East regions. Schools were targeted on the basis of both a higher than average proportion of pupils receiving free school meals and the numbers achieving a Level 3 or borderline Level 4 in their end of Key Stage 2 reading tests.

The pupils were selected on the basis of teacher assessments of their Key Stage 2 reading level. Eligible pupils were those who were either not expected to be reading at Level 4 at the end of Year 6 or those who would potentially reach a Level 4b or 4c.

Initial consent to participate in the programme and evaluation was sought from headteachers and an agreement signed prior to randomisation. This included agreeing to make a teaching assistant available for training and the delivery of the programme, to provide post-test data for all of the pupils recruited to the trial whenever possible, and a commitment that the school would not participate in any other literacy evaluation for the duration of the programme. Opt-out parental consent was obtained from the parents or guardians of pupils participating in the evaluation.

2.3 Intervention

Tutoring with Alphie is a computer literacy tutoring programme that involves pupils working in pairs, monitored by a tutor. It has been designed to combine the strengths of cooperative learning, computer-assisted instruction, and tutoring. Pupils are paired according to current attainment in literacy and the programme is typically delivered simultaneously to a group of around four to eight pupils, whereby each pair works together at the computer and takes turns to alternate as either reader or coach. The group is monitored by an adult tutor who is either a teaching assistant or a qualified teacher who monitors and guides the tutoring process. In the current evaluation teaching assistants acted as programme tutors in all schools. The programme was originally designed for younger struggling readers, but for the current evaluation it was adapted to be delivered to Year 6 pupils reading below the expected Key Stage level.

Prior to beginning the actual tutoring process the staff tutor administers a computer-based reading assessment for each of the pupils who have been selected for the intervention. Once the individual *assessments have been completed for all pupils, Tutoring with Alphie then ranks pupils on the basis* of

these assessments. The teaching assistant then creates a number of similarly attaining pairs based on this ranking and identified skill level, while taking account of any potential incompatibility issues between pupils. In other words, pupils are largely matched on their current level of attainment, but consideration is given as to whether the individual pupils would work well together as a pair. The programme then creates the first of a series of two-week tutoring plans for each pupil pair comprising a range of progressive computer-based literacy activities based on fluency and expression.

Following the initial assessment, pairing and development of a work plan, pupils then begin working on the programme on a daily basis for 30 minutes. It is intended to be delivered in paired groups, of four to eight pupils, with oversight by an adult tutor through a pull-out session during the normal school day. In the case of the current evaluation it was intended to be delivered over a period of six weeks. Once the tutor reviews and accepts the tutoring plan created by the programme for each pair, pupils begin working on the computer alternating in their respective roles as peer reader or coach. The software allows each pair to create a team icon, and once the pupils login at each session the programme is designed to record daily activities and progress. During each daily session one pupil starts as reader and begins an activity and at the end of the activity will be requested for a response. When the reader gives a response the peer coach evaluates whether or not it is correct. If the response is incorrect the programme provides instructional support and prompts the student to try again. The programme also prompts the pupils when it is time to switch roles between reader and coach.

Cooperative learning between the pairs is actively encouraged through the respective roles of peer reader and coach. Peer coaches are encouraged not to provide the answer and the team cannot move forward until both pupils have mastered the skill level they are working on. When a team masters an instructional level the teaching assistant is alerted on their computer and is then required to undertake a quick assessment as to whether mastery has been achieved before advancing the programme on to the next level of more difficult activities.

The Tutoring with Alphie programme in the current evaluation was accessed online through the Google Chrome web browser with web hosting and file storage provided by the developers in the US. The initial training and ongoing support to schools was provided locally by Success for All (UK).

Control group children attended the same schools, were not exposed to the intervention and received their normal classes as per usual.

The process evaluation highlighted that when the programme was first introduced into the schools in the current evaluation the adaptations to the programme designed to adjust the content for an older age group in the English context had not been fully developed or sufficiently piloted prior to implementation. While considerable efforts were made to resolve these issues as they emerged by Success for All staff in the UK together with the programme developers in the US, these technical issues had a considerable impact on both the implementation of the programme and the response rates achieved for the impact evaluation. This is discussed in detail in Section 4.

2.4 Outcomes

There was only one primary outcome considered in the impact evaluation. In common with all projects funded through the EEF Literacy Catch-up funding round, the outcome was pre-specified as the GL Assessment New Group Reading Test (NGRT) digital edition. This provided a post-test measure of reading comprehension and is based on assessments of sentence and passage completion. Schools administered the digital testing online under exam conditions and these were automatically scored by GL Assessment and therefore blind to treatment allocation.

2.5 Sample size

The original estimates in the protocol suggested that for individually randomised trial, and a minimum detectable effect size of 0.2, a sample size in the region of 290 pupils would be required. This would be split evenly between Tutoring with Alphie and the control group. Approximately 12 to 16 pupils were to be recruited at each school with half randomised to receive the programme and the other half to act as a control. Prior to randomisation the pupils in each school were paired according to prior literacy ability.

The following parameters were used in the power calculation:

- Significance level (α) = 0.05
- Power (P) = 80%
- Pre post-test R^2 = 0.64

The achieved sample recruited to the study was 248 pupils from a total of 21 schools. Although the final sample size recruited to the study was slightly below the desired sample size this should still have been adequate to detect significant differences between the two groups (control and intervention) of an effect size of 0.2. However, due to the number of schools that did not complete the NGRT tests at post-test the final achieved sample of 72 pupils for analysis, drawn from only six schools, did not have sufficient power to detect statistically significant effects within the expected range.

While the final achieved sample remained balanced according to baseline characteristics it was underpowered and likely to be influenced by attrition bias as the schools who returned NGRT data were unlikely to reflect the characteristics of the original sample.

2.6 Randomisation

Schools were recruited by Success for All (UK) and following recruitment IEE supplied the CEE with an Excel database containing pupil identifiers and background data. Random allocation was conducted independently by the CEE using the following procedure outlined below.

A total of 248 pupils from 21 schools were recruited to the study and pupil identifiers were supplied with data on teacher-predicted reading grades according to Key Stage level. Pupils were block randomised within schools whereby as a first step they were rank ordered on predicted KS reading levels. Following ranking, pupils were randomly allocated a number between 0 and 1. Pupils were then randomly allocated in pairs, whereby the higher random number in each pair was allocated to Tutoring with Alphie, and the other pupil allocated to control. Allocating pupils in pairs based on their predicted reading ability was intended to help ensure that both samples were evenly matched according to the main outcome of concern.

In the event of uneven numbers in a school the remaining pupil was allocated to Tutoring with Alphie if their random number was greater than 0.5, or if less than 0.5 they were allocated to the control group. This resulted in 125 pupils being allocated to Tutoring with Alphie and 123 to the business as usual control condition.

2.7 Analysis

The impact analysis was conducted using multiple linear regression which was planned on an intention-to-treat basis, meaning that the results of all pupils who were registered to begin the trial would be used in the final analysis, regardless of whether, or how many times, they used the programme. To take account of stratification of the randomisation within schools the analysis was conducted using robust standard errors, with school identifier as the grouping variable. The regression models included a number of covariates which were standardised prior to analysis. These included a proxy pre-test score

based on prior reading attainment through Key Stage 2 assessments obtained from the National Pupil Database (NPD), as well as on gender, age in months and free school meal eligibility.

Further pre-specified analyses were outlined in the protocol to assess the programme's impact on various subgroups, including boys, girls, those eligible for free school meals and those with low prior reading attainment. However, due to the extent of attrition at post-test through school dropout, it was agreed in advance of analysis that only headline findings should be presented in the main body of the report. With the diminished sample size, and only a small number of schools completing the NGRT assessments, the risk of spurious and misleading findings from subgroup analyses was potentially high. These results are, however, presented in Appendix 1.

2.8 Process evaluation methodology

The main process evaluation was conducted by the IEE and involved an online survey of both the teaching assistants delivering the programme, and the pupils that received Tutoring with Alphie. All intervention schools were contacted. The survey collected both quantitative measures and qualitative responses in relation to both perceived positive and negative aspects of the programme. Email correspondence from Success for All trainers was also collected and analysed. The original data was made available to the independent evaluation team.

While the IEE were responsible for conducting the main process evaluation the CEE also conducted an independent light touch evaluation through in-depth telephone interviews with a number of teaching assistants delivering the programme. All tutors from all schools were invited to take part and all respondents who agreed to take part were contacted and interviewed towards the end of the planned delivery of the programme (n=4).

2.9 Costs

Information on the cost of the programme, including costs of training and ongoing support as well as programme materials, software licence and special keyboards, was obtained from Success for All (UK). As the estimated cost was based on providing sufficient equipment to tutor six children, the cost per pupil was calculated by dividing the overall costs to schools by six.

Impact evaluation

3.1 Timeline

The programme was delivered in the final term of the school year; the timeline of key activities is set out in Table 3.1 below.

Table 3.1: Timeline of key activities of the impact evaluation

Month	Activity
March–mid May 2013	Recruitment of schools and collection of predicted KS2 results
Mid May 2013	Randomisation of eligible pupils to Tutoring with Alphie/Control
Late May 2013	Training of teaching assistants
May–July 2013	Programme delivery
8–17 July 2013	Post-testing (NGRTA test) collected

3.2 Participants

Schools were recruited throughout England, but although the majority of schools were based in the North (North East and Yorkshire and Humber) some were also recruited from the South East, South West and East regions. All schools were from urban settings from non-metropolitan areas. Among the majority of schools who had an Ofsted rating just over two-fifths of the schools (43%) were rated as 'Good', around a third of the sample were rated as 'Satisfactory' and 14% rated as 'Inadequate'.

Table 3.2 compares the percentages of pupils according to the proportion of pupils with special educational needs (SEN), with English as an additional language and those in receipt of free school meals within the sample of schools recruited to the study with the national averages. Overall the recruited sample had a similar proportion of pupils with SEN or an additional language compared to the national figures. In line with the recruitment strategy the sample contained a higher proportion of pupils in receipt of free school meals.

Table 3.2: School characteristics of sample compared with national data

	England national (State funded)	Sample
Percentage of pupils with SEN or on School Action Plus	7.7%	7.5%
Percentage of pupils with English not as a first language	18.1%	13.4%
Percentage of pupils eligible for free school meals	19.2%	32.3%

The majority of schools in the sample were Community schools, approximately three out of ten were Academy schools and one school was a Foundation school (Table 3.3).

Table 3.3: School type

Type	Number of schools
Foundation school	1
Voluntary aided school	0
Community school	14
Academy sponsor led	6

The CONSORT diagram (overleaf) provides details of the pupils recruited to the trial and those who were followed up at the end of the programme. In total, 252 pupils were recruited to the evaluation but

4 pupils were excluded on the basis of their assessed reading levels, which were below the intended use of the programme (below level 1a). Of the eligible pupils, 125 were allocated to receive Tutoring with Alphie and 123 to the control condition. Fifteen of the 21 participating schools failed to provide pupil data for the New Group Reading Test, the main outcome measure of reading comprehension. This represented a dropout rate of 72% of those recruited to the original trial. As the pupils who were allocated to Tutoring with Alphie and control groups were within the same schools the final achieved sample sizes for analysis were similar (37 and 35 respectively). Although the two samples remained relatively balanced between the Tutoring with Alphie group and their respective controls, a dropout level of this magnitude severely undermines the validity of the impact findings in the current evaluation.

Figure 3.1: Participant flow diagram
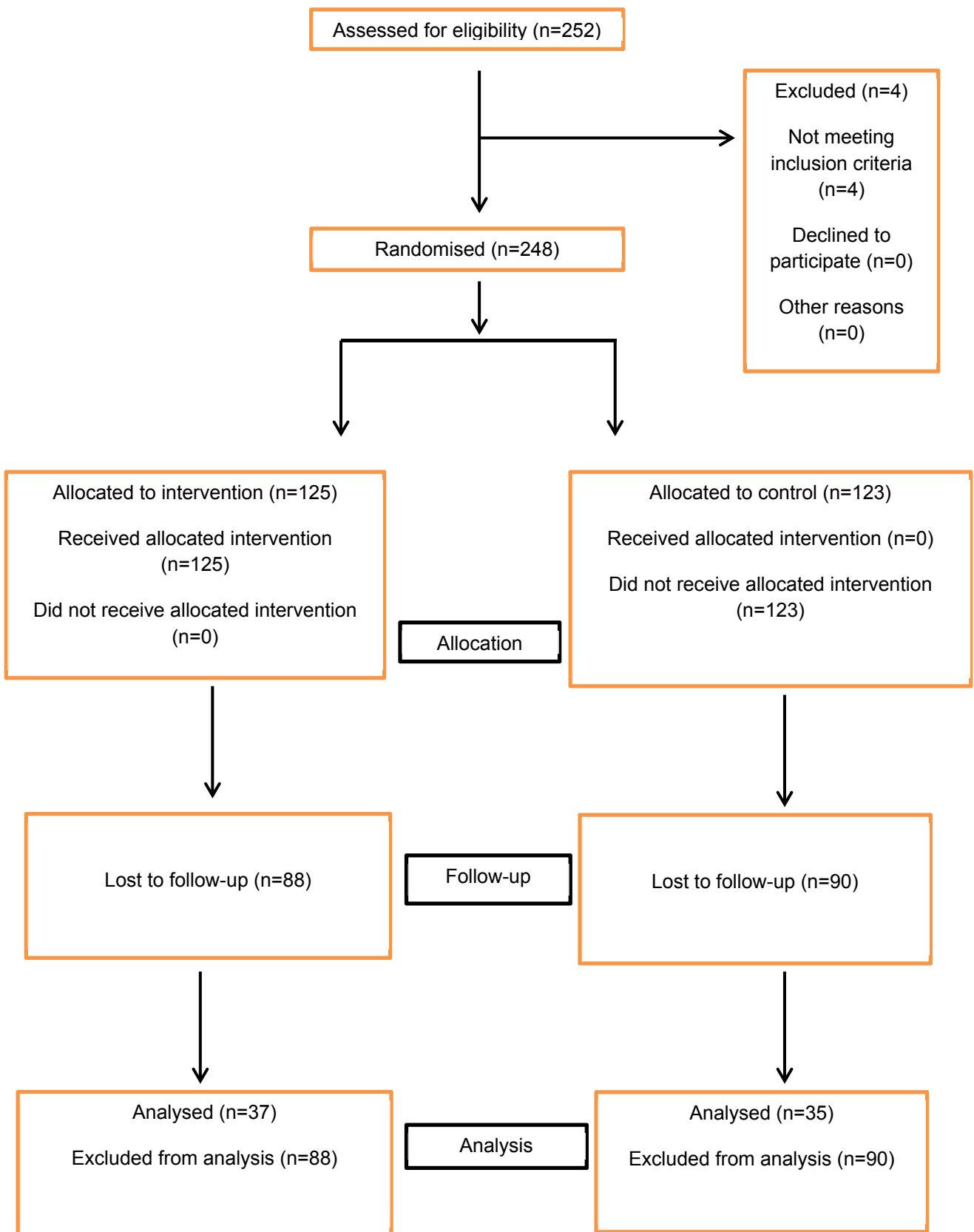
3.3 Pupil characteristics

Tables 3.4–3.7 provide an overview of the characteristics of the participants for both those who were selected to receive Tutoring with Alphie and the control group at the start of the programme (baseline) and at post-test follow-up, according to their gender, free school meal entitlement, whether they spoke English as an additional language and special educational need status. This information was gathered directly from schools prior to randomisation. These showed there were few major differences between the control group and the Tutoring with Alphie group at either baseline or at post-test. No statistically significant differences were found at either time point.

Table 3.4: Differences between the control and Tutoring with Alphie groups by gender

Gender	Pre-test (Baseline)		Post-test (Outcome)	
	Control Group	Intervention Group (Tutoring with Alphie)	Control Group	Intervention Group (Tutoring with Alphie)
	n (%)	n (%)	n (%)	n (%)
Male	65 (56%)	76 (64%)	19 (54%)	22 (60%)
Female	52 (44%)	43 (36%)	16 (46%)	15 (41%)
n	117	119	35	37
Pearson chi-square	1.69 (1)	p=0.19	.20 (1)	p=0.66
Missing values (n)	6	6	-	-

Table 3.5: Differences between the control and Tutoring with Alphie groups by free school meal entitlement

FSM	Pre-test (Baseline)		Post-test (Outcome)	
	Control Group	Intervention Group (Tutoring with Alphie)	Control Group	Intervention Group (Tutoring with Alphie)
	n (%)	n (%)	n (%)	n (%)
Yes	63 (54%)	70 (59%)	17 (49%)	20 (54%)
No	54 (46%)	48 (41%)	18 (51%)	17 (46%)
n	117	118	35	37
Pearson chi-square	0.72(1)	p=0.40	0.22(1)	p=0.64
Missing values (n)	6	7	-	-

Table 3.6: Differences between the control and Tutoring with Alphie groups by English as an additional language

EAL	Pre-test (Baseline)		Post-test (Outcome)	
	Control Group	Intervention Group (Tutoring with Alphie)	Control Group	Intervention Group (Tutoring with Alphie)
	n (%)	n (%)	n (%)	n (%)
Yes	23 (20%)	23 (19%)	4 (11%)	5 (14%)
No	94 (80%)	96 (81%)	31 (89%)	32 (87%)
n	117	119	35	37
Pearson chi-square	0.004(1)	p=0.94	0.01(1)	p=0.79
Missing values (n)	6	6	-	-

Table 3.7: Differences between the control and Tutoring with Alphie groups by special educational need

SEN	Pre-test (Baseline)		Post-test (Outcome)	
	Control Group	Intervention Group (Tutoring with Alphie)	Control Group	Intervention Group (Tutoring with Alphie)
	n (%)	n (%)	n (%)	n (%)
Yes	66 (56%)	69 (58%)	29 (83%)	26 (70%)
No	51 (44%)	50 (42%)	6 (17%)	11 (30%)
n	117	119	35	37
Pearson chi-square	0.06(1)	p=0.81	1.58(1)	p=0.21
Missing values (n)	6	6	-	-

3.4 Outcomes and analysis

The analysis was designed to test the efficacy of Tutoring with Alphie in improving children's reading comprehension. The outcome variable (NGRT) was used to assess the differences between those in the control group and those who received Tutoring with Alphie using linear multiple regression. The model controlled for prior reading attainment at Key Stage 2 (pre-test), gender, free school meal eligibility and the child's age at post-test[2], using robust standard errors with school identifier as the grouping variable. This adjusts the standard errors to take account of the fact the programme was delivered in different schools by different tutors. The full results are reported in Table 3.8**.**

Table 3.9 summarises the main outcome findings outlining the adjusted post-test means[3] for reading comprehension for the control and Tutoring with Alphie groups and the associated effect size. In this table, the coefficient refers to the mean change in the dependent variable (i.e. the outcome—reading comprehension) for one unit change in the predictor variable (e.g. Intervention), while holding the other predictors in the model constant. Coefficients can be positive, zero, or negative. A positive coefficient

2 All covariates were standardised prior to analysis.

3 These were calculated through the regression coefficients.

indicates a higher score on the outcome variable; for example, if a positive coefficient is seen for the 'Intervention' variable, then it indicates that those in the intervention group score higher than those in the control group. Similarly, a negative score indicates a lower score on the outcome variable, and a coefficient of zero indicates no difference in the outcome variable. The standard error associated with each coefficient is also reported, and this gives an indication of how precise the estimated coefficient term is (with small values reflecting greater precision). Statistically significant coefficients are indicated by an * indicating a p-value of less than 0,05, which means that the finding is unlikely to have been caused by random error (i.e. chance).

A small positive effect of the programme was found (*g*=0.11); however, this was not statistically significant with a wide confidence interval. Therefore, the results do not constitute strong evidence that Tutoring with Alphie had an impact on the pupils' reading comprehension. As indicated in Section 2.3, subgroup analyses were planned but with the diminished sample size and only a small number of schools completing the NGRT assessments the risk of spurious and misleading findings was potentially high from such analyses. These results are presented in Appendix 1.

Table 3.8: Main effects for primary outcome—reading comprehension (NGRT)

	Coefficient	Robust Standard Error	[Confidence Interval]
Intervention	0.11	0.13	[-0.21, +0.44]
Pre-test (KS2 reading)	0.30*	0.09	[+0.06, +0.55]
Gender (male)	3.85	4.29	[-7.17, +14.88]
Age in months	-0.23	0.11	[-.53, +0.07]
FSM eligibility	-4.26	5.90	[-19.43, +10.91]
Constant	0.09	0.71	[-1.72, +1.91]
Number of observations	72		

*Notes: Significance *p<0.05,*

Intervention n=37, Control n=35

Table 3.9: Summary of main effects at post-test

	Adjusted post-test means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (sd)	n (missing)	Mean (sd)	n in model (intervention; control)	Hedges g** [95% CI]	p-value
Reading comprehension	37 (88)	0.21 (0.05)	35 (90)	0.10 (0.95)	72 (37;32)	0.11 [-0.13, +0.36]	0.63

Cost

The cost of the approach as delivered in the trial is estimated at £582 per pupil, or £3,490 per school. This estimate is based on a teaching assistant supporting six students per school. The estimate includes 3.5 days of externally provided training and support from Success for All (UK) delivered on-site (£3,030), website licence fee (£300) and an initial start-up pack, including a programme manual and four electronic key pads (£160) to cater for six pupils and an adult tutor in a group. It does not include direct salary costs, supply cover for training or costs associated with the provision of computers. Given that a large proportion of the cost incurred is related to staff training, the per pupil cost to schools could be

substantially reduced if a greater number of pupils followed the programme and were supported by the same teaching assistant, or if a larger number of teaching assistants attended training at the same time.

The process evaluation highlighted that some schools may require additional computing facilities and that releasing the teaching assistant for the time required for training and completing the initial assessments was an issue in some schools. Therefore the availability of such existing resources would need to be fully considered as to whether or not they constituted additional costs to the school. Ongoing costs would be reduced, dependent on the need for training and due to the initial costs for equipment.

The schools involved in the current evaluation were provided with the training and resources at no charge, but provided the time in kind from existing teaching assistants within the school.

Process evaluation

This section presents the key findings from the process evaluation, bringing together the findings from the various components of this aspect of the evaluation (i.e. the telephone interviews with teaching assistants delivering the programme, the teacher and student online survey, and an analysis of email communication of Success for All trainers). Pupil returns to the survey were received from 11 of the 21 schools and teaching assistant responses from 8 schools.

4.1 Implementation

A number of serious issues with the Tutoring with Alphie software emerged early on in the implementation of the programme. Software development was still ongoing at the start of the programme and the version of the software introduced into schools had not been sufficiently piloted by the developers prior to its introduction to schools in England. Therefore the intervention cannot be considered to have been delivered under optimal conditions owing to the problems that were encountered.

This lack of readiness of the software clearly had a significant impact on the overall evaluation and undoubtedly played a major role in the lack of outcome data obtained in the current study. Some of the key issues that emerged during implementation are outlined below.

Software problems

A wide range of problems were encountered with the software particularly in the early stage of implementation. Examples of the problems experienced included: difficulties in accessing the online site, screen freezes, slow running of the program, reading material being obscured by on-screen icons, and a failure of the software to record prior pupil progress resulting in a lack of progression and repetition of material. While the programme is intended to operate across a range of reading levels based on an initial assessment of the children's reading abilities, at the start of the programme it was only operating at one of these levels. Although this level was considered appropriate by tutors for the majority of pupils in the intervention who were at KS2 Level 3, it was considered as unsuitable for a number of pupils with more serious reading difficulties (Level 2). It was clear that the adaptations made to the software for use with Year 6 pupils in England had not had sufficient time for extensive piloting and a number of major bugs remained at the time of implementation.

Exposure and dosage

The independent evaluation team have not been provided with any detailed measure of implementation or the amount of the programme that individual pupils received within schools. The software program should have recorded accurate data on the programme's use, but process data suggested that there were problems with the recording of this data by the software during the early stages of the intervention, making such data, if available, unreliable. Therefore the extent to which individual schools continued to deliver Tutoring with Alphie over the six-week period remained unclear. Process data highlights that three schools withdrew completely from the programme at an early stage having become disillusioned with the programme and it would also appear that a number of other schools may have ended the programme prematurely, or did not manage to get the programme fully up and running on a regular basis. Other schools, however, persisted with the programme with varying degrees of success. It was clear that Success for All support staff worked extensively with these schools that continued with the programme and did their best, in conjunction with the developers in the US, to resolve the issues as they emerged.

A small number of pupils reading at around KS2 Level 2 were also withdrawn from the programme by tutors as the reading material and activities were considered too advanced and therefore unsuitable for their current reading level as a result of the limited functionality of the software (noted above).

Those schools that were identified as delivering the programme to the end of the school year were still considerably behind in the completion of the intended number of sessions. A number of schools who have provided process data were continuing with the programme beyond the end of the initial six-week initial period, and by this stage most of the major issues with the software appear to have been largely resolved. Some of these more persistent schools indicated that the software was suggesting that they had delivered less than half the intended number of sessions, but they highlighted that there had been some difficulties with the software recording progress in the early stages of implementation. Therefore, overall the intervention cannot be considered to have been implemented at the intensity, or according to the intended number of sessions, expected to be delivered.

Stakeholder perspectives

The data obtained from the tutor and pupil survey provided a mixed response in terms of stakeholders' perspectives. In light of the software difficulties, it is not surprising that most of the negative comments were centred upon the technical issues experienced and it was clear that both pupils and the teaching assistants experienced considerable frustration with these aspects of the programme. Nevertheless, among both groups of stakeholders positive aspects of the programme were highlighted. A number of pupils highlighted how they enjoyed spending time on the computers and working collaboratively in pairs. Others suggested that they felt it had helped them with their reading and spelling and had boosted their confidence in reading. More negative responses from the pupils highlighted frustration with screen freezes, and the 'done' button not being displayed on the page to highlight that they had finished an activity. Aside from these specific technical issues, criticism over the programme content was related to the programme graphics which a number of pupils felt were too 'babyish' for their age group.

Responses from the teaching assistants who delivered the programme highlighted that they felt that overall the training provided was good, although for at least some of the sessions conducted at individual schools no password to access to the programme was available at the time so they did not manage to get direct hands-on experience of the programme. One respondent, while happy with the overall quality of the training, highlighted that it was based on the US version of the programme, rather than the adapted version. As a result, they felt they had oversold the programme activities to their pupils in advance of starting the programme. They felt that the range of activities provided by the programme was more limited than they had first been led to believe. While the teaching assistants also reported many of the same technical issues as the pupils, other issues included: pupils getting stuck on a particular activity and lack of progression, that some of the graphics could be updated to be more attractive to Year 6 pupils, and the lack of adaption from American to English spellings of words.

These difficulties clearly led to disengagement among both staff and pupils; however, it was apparent that some schools worked closely with Success for All staff and the developers to try and resolve these issues. Among these schools a more positive experience of Tutoring with Alphie developed over time as some of the initial difficulties began to be resolved. Clearly, significant efforts were made by Success for All staff in the UK to try and support schools to try and overcome these difficulties. Towards the end of the evaluation a number of the teaching assistants who had persisted with the programme felt that with many of the software problems resolved they could see the potential benefits of the approach and perceived it as having a positive impact on reading among their pupils.

The timing of the programme, towards the end of the summer term, was another barrier to smooth programme delivery. The reasons provided were related to competing activities and the wider benefits to be derived from such a programme to the school. Respondents highlighted a number of activities and competing priorities at the end of Year 6. These included transition days, end of year productions and some pupil absence through families taking advantage of discounted holidays in the final stages of the school term. In addition to these activities several respondents highlighted that due to the timing the school would not benefit directly from any improvements in pupil attainment through improved overall SATs results. In this respect, and taking account of the competing activities at this time of year, it was suggested that if the programme was to be delivered to Year 6 again it would be preferable to schedule

it during the autumn term. A number of the teaching assistants who had worked closely with Success for All to try and resolve the software problems, and had become more positively disposed towards the programme, felt that with the technical difficulties resolved it would be more ideal for a younger group of pupils (*'perhaps Year 3 or 4'*).

Not all the technical difficulties experienced were down to the readiness of the software. The IT capacities of some primary schools to deliver an online computer program also led to some difficulties. This could be as simple as the initial unavailability of the Google Chrome browser which was required to run the program, which compounded the difficulties in getting the software installed and approved for use on school computers. Some schools experienced difficulty in gaining access to computers during the school day and as a result one school ran the programme between 8.30 and 9.00am as this was the only time they could freely access computers. Some of the wider barriers to successful implementation highlighted included teaching assistants finding sufficient time to be released from class or gaining timely general technical support with the computers which in some cases had been outsourced to external companies.

4.2 Outcomes

Perceptions of outcomes among both pupils and teaching assistants suggested that despite the significant problems that were experienced with Tutoring with Alphie, there was a feeling among some respondents that the programme was having a positive impact on the pupil's reading and confidence in reading.

The main unintended consequences of the programme related to the frustration experienced with the software, by both teaching assistants and pupils. While an element of repetition is built into the programme until mastery at a given level is achieved, some teaching assistants felt that the level of repetition was too high and could lead to pupil boredom in the event of unsuccessfully completing a task. It was unclear whether this was a result of programme design or some of the difficulties with progression of the software (noted above).

In addition, the teething problems experienced by some of the schools had a clear impact on the broader evaluation and the failure of schools to conduct post-test evaluations. Some schools felt they had wasted enough time with the programme at a busy time in the Year 6 calendar, or that there was little point in providing outcome data as pupil progress would be limited due to the low level of implementation.

4.3 Formative findings

The evaluation highlights that successful implementation of the programme would require:

- Full development of the programme and extensive piloting prior to implementation. This should include an assessment of reading material, adjustments to materials to revise American spellings and the development of materials for the full range of abilities that the programme was intended to cover.
- The availability of sufficient IT capacity within the school, including the availability of basic software (Goggle Chrome) and sufficient access to dedicated computers at appropriate times.
- A consideration of when to use the programme. Running the programme at the end of Year 6 may inhibit the full buy in of schools as there is a perception that the school will not see the benefits of any improvements in terms of improved SATs results. Competing priorities at the end of the school year with transition events also hamper implementing a programme at the end of Year 6.
- Translation of a previously effective programme to a different age group requires careful consideration. For example, characters embedded within the software should be age appropriate, attractive and engaging to the age group the programme is aimed it.

- A consideration of the appropriate age group for the programme. The current software may be better suited to a younger age group. A number of the teaching assistants felt that it would be better suited to struggling readers in Year 3 or 4, although this may be a reflection of the content and graphics of the current programme.

4.4 Control group activity

The programme was delivered as a pull-out activity and control group children attended their classes as normal (business as usual). As an individually randomised trial within schools there is always the risk of contamination effects; however, as a computer-based programme run through pull-out sessions overseen by an adult tutor with login passwords it is unlikely that any of the control group were exposed to the programme.

Conclusion

5.1 Key conclusions

Key conclusions
1. The evaluation was unable to provide a secure estimate of the programme's impact on literacy, primarily due to the large number of schools that dropped out of the trial.
2. The Tutoring with Alphie programme was not sufficiently developed or at a stage to be evaluated in English schools at the time of the project.
3. A number of schools perceived the programme as having some promise and worked with the developers and Success for All to overcome initial difficulties with the software.
4. The programme had been adapted as a catch-up literacy programme for Year 6, but the process evaluation suggested that it may be better targeted at a younger age group.
5. To assess the efficacy of approaches that combine collaborative learning with computer-assisted technology to improve reading comprehension, further evaluation would be required.

Evidence from previous research has demonstrated that the approach underpinning Tutoring with Alphie can be effective in improving reading outcomes among a younger age group of beginning readers in the US. The current evaluation aimed to assess whether an adapted version of the programme for an older age group delivered in the English context could also lead to demonstrable improvements in Year 6 pupils' reading comprehension for those who were assessed at reading below the expected Key Stage level. The current evaluation did not provide a secure estimate of the programme's impact reading comprehension outcomes among Year 6 pupils in England.

The significant problems experienced with implementing the software and the lack of post-test data provided by schools has seriously undermined the integrity of the current trial. As a result we can say little about whether the programme, if implemented as intended, could make a discernible change in the intended outcomes. While some schools became frustrated and disillusioned with the programme and did not want to waste any more time with the programme or the wider evaluation, a number of schools persisted and worked with Success for All staff to overcome the initial difficulties in implementation. Some of the teaching assistants who delivered the programme became positively disposed to the approach as a reading intervention. However, they felt that Tutoring with Alphie, at least in its current formulation, would be better placed as a reading intervention for younger children than children aged 10 to 11 as was the case in the current evaluation.

The evaluation highlights a number of important considerations when implementing software programmes, especially those which are being adapted from other successful programmes developed in other national contexts. It is clear that the US developers of the programme underestimated the extent of development that would be required to adapt the programme to a group of older readers in a different national context, at least within the timescales of the current evaluation. Careful pilot work is required to ensure both the functionality of the software and the appropriateness of materials to the intended audience.

5.2 Limitations

There are a number of limitations to the current impact evaluation. The extent of attrition among schools providing post-test data represents a serious threat to the validity of the outcome analysis with over two-thirds of those randomised lost at follow-up. Although as an individually randomised trial within schools the drop-out rate between the intervention and control group remained balanced, measurement attrition on this scale is likely to have introduced serious bias into the results. As the source of attrition was also directly related to the programme this undermines the representativeness of its delivery across a range of schools. In other words, the results do not reflect its delivery under realistic conditions, as

schools who continued with the programme are likely to differ in important respects from those that abandoned the programme.

The extent of attrition also means that the study became severely underpowered to detect significant differences between the control and intervention groups. The reported effect size for reading comprehension was not statistically significant, with particularly wide confidence intervals. Overall, it cannot be concluded that the observed effect represents a secure estimate of the effect of the programme.

5.3 Interpretation

While there is a lack of rigorous studies evaluating the impact of computer-assisted programmes to improve the reading outcomes of pupils, previous syntheses of the evidence have concluded that overall they can have a positive effect, although one slightly smaller than other research-based interventions (Higgins et al., 2012). The current evaluation aimed to assess a tutoring intervention with previous evidence of the effectiveness of the approach which combined computer-aided instruction with cooperative learning. Previous research has suggested that traditional computer-assisted software has a limited impact on reading outcomes (Slavin et al., 2009), although when combined with other pedagogical approaches such as cooperative learning it may represent a more promising approach. Unfortunately, the programme as delivered to schools in the current evaluation was insufficiently developed, particularly at the beginning of the evaluation. This led to considerable frustration and demoralisation with the programme on the part of both pupils and teaching assistants with subsequent drop-out or severe difficulties with implementation in many schools. This highlights the particular importance of extensive testing of computer software prior to rigorous evaluation, especially when it involves translational programmes from other national contexts.

Although considerable efforts were made on the part of Success for All staff to resolve these issues, the programme was not delivered as originally intended and this undermined the current evaluation in a number of ways. The outcome results of the current evaluation did show a small positive improvement in reading comprehension (NGRT) among those who received Tutoring with Alphie and provided outcome data. However, many schools failed to provide NGRT data, and the difference between the control group and those who received Tutoring with Alphie was not statistically significant. In other words, we cannot confidently attribute the observed positive impact to the programme itself as the observed differences may simply have occurred by chance. The current evaluation therefore provides limited evidence in terms of the effectiveness of a computer-based tutoring programme combined with cooperative learning to assist with improving reading comprehension among struggling Year 6 pupils.

The process evaluation highlighted how those schools who persisted with implementing the programme perceived that it had potential as a reading intervention, but they felt that further adaptation would be required. A number of teaching assistants who delivered the programme suggested that the programme would be better targeted at younger struggling readers than those in the current evaluation and that this would assist in school buy-in through the potential rewards of the programme being reflected in improved SATs results for the school.

5.4 Future research

There is lack of rigorous evaluations of programmes that combine computer-assisted instruction with cooperative learning on reading outcomes. Unfortunately the current evaluation failed to provide further evidence of the efficacy of this approach to improving reading comprehension. However, the study provides a number of important lessons for future research evaluations. First, it highlights the importance of programmes being fully developed and piloted prior to being subject to an effectiveness trial. This may be particularly salient in the case of computer software, where timely adaptations may be difficult to achieve. The current evaluation also highlights some potential issues over the import of successful programmes evaluated in other national contexts. While there may be a tendency on the

part of programme developers to wish to ensure programme fidelity to a positively evaluated programme, there is a need to carefully consider translation to a different age and national context. This was most evident in the current evaluation in relation to American spellings in the programme resources, but other cultural and age-related adaptions may also be required. This, again, underlines the importance of careful initial pilot work among the age range and national context when importing successful programmes from other national contexts.

References

Brooks, G., Miles, J.N.V., Torgerson, C.J. and Torgerson, D.J. (2006) 'Is an intervention using computer software effective in literacy learning? A randomised controlled trial'. *Educational Studies*, 32: 2,133–43.

Chambers, B., Slavin, R.E., Madden, N., Abrami, P., Logan, M., and Gifford, R. (2011) 'Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders'. *The Elementary School Journal*, 111: 4, 625–40.

Cheung, A. and Slavin, R.E. (2012) 'How features of educational technology applications affect student reading outcomes: A meta-analysis'. *Educational Research Review*, 7: 3, 198–215.

Higgins, S., Xiao, Z. and Katsipataki, M. (2012) *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation*. Durham: Durham University.

Johnson, D.W. and Johnson, R.T. (2009) 'An educational psychology success story: Social interdependence theory and co-operative learning'. *Educational Researcher*, 38: 5, 365–79.

Slavin, R.E., Lake, C., Davis, S., and Madden, N. (2009) *Effective Programs for Struggling Readers: A Best Evidence Synthesis*, Baltimore MD: John Hopkins University, Center for Research and Reform in Education (CRRE).

Appendix 1: Interaction models

Interaction Models									
	Pre-test			**Gender**			**FSM**		
	coef.	**s.e.**	**95% CI**	**coef.**	**s.e.**	**95% CI**	**coef.**	**s.e.**	**95% CI**
Intervention	0.12	0.13	[-0.22, +0.45]	3.18	2.40	[2.99, +9.36]	-3.66	1.50	[-7.50, +0.19]
Pre-test (KS2 reading)	0.29	0.14	[-0.08, +0.66]	0.31*	0.09	[+0.06, +0.55]	0.28*	0.10	[+0.02, +0.53]
Interaction (Pre-test×Intervention)	0.02	0.33	[-0.82, +0.87]	-	-		-	-	
Gender (male)	3.83	4.23	[-7.05,- +14.71	9.85	6.06	[-5.74, +25.43]	2.31	4.03	[-8.04, +12.67]
Interaction (Gender×Intervention)	-			-11.83	9.32	[35.79, +12.12]	-	-	
FSM eligibility	-4.33	6.31	[-20.55, +11.89]	-3.35	5.76	[-18.15, +11.45]	-11.42	7.31	[-30.20, +7.36]
Interaction (FSM×Intervention)	-			-	-		13.95*	5.22	[+0.52, +27.37]
Age at post-test	-.23	.12	[-0.54, +0.08]	-0.23	0.12	[-0.53, +0.06]	-0.24	0.12	[-.55, +0.06]
Constant	-.12	0.87	[-2.12, +2.36]	-1.71	1.07	[-4.49, +1.06]	2.42	1.16	[-0.56, +5.40]
No. of observations	72			72			72		

*p<0.05

Appendix 2: Security classification of trial findings
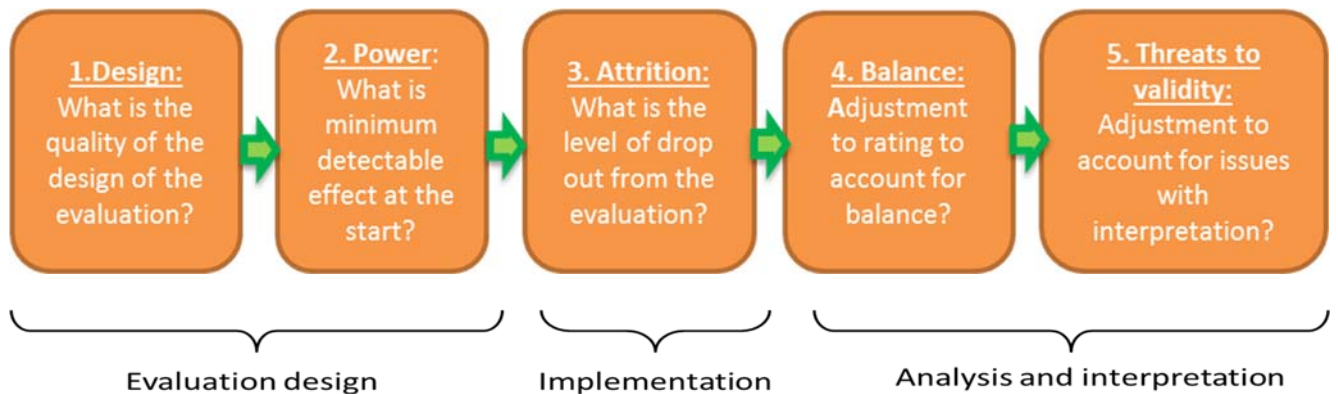
Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5 🔒	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4 🔒	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3 🔒	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		
2 🔒	Matched comparison (quasi-experiment)	< 0.5	< 40%		
1 🔒	Comparison group with poor or no matching	< 0.6	< 50%		
0 🔒	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 0 🔒. This means that the conclusions have very low security.

The trial was designed as an efficacy trial and could achieve a maximum of 5 🔒. The trial was moderately underpowered because fewer pupils were recruited than intended for the MDES of 0.2 (248 pupils instead of 290). There was substantial attrition of more than 70%. This resulted in zero padlocks, and rendered consideration of Balance and Threats to Validity redundant. Therefore, the final padlock rating is 0 🔒.

Appendix 3: Cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found on the EEF website. Cost ratings are awarded as follows:

Cost rating	Description
£	*Very low:* less than £80 per pupil per year.
£ £	*Low:* up to about £200 per pupil per year.
£ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.