Switch-on Reading

Evaluation Report and Executive Summary

February 2014

Independent evaluators:

Durham University

Professor Stephen Gorard

Beng Huat See

Nadia Siddiqui

The Education Endowment Foundation

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.

We aim 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 founded in 2011 by lead charity The Sutton Trust, in partnership with Impetus Trust (now part of Impetus – The Private Equity Foundation), with a £125m grant from the Department for Education. With investment and fundraising income, the EEF intends to award as much as £200m by 2026. Together, the EEF and Sutton Trust are the Government-designated What Works Centre for Improving Education Outcomes for School-Aged Children.

Literacy Catch-up

In May 2012 the Department for Education awarded the EEF a further £10 million for a grants round dedicated to literacy catch-up projects for children at the primary-secondary transition. The projects funded within this round aimed to identify effective ways to support pupils who do not achieve Level 4 in English by the end of Key Stage 2.

For more information 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 project was independently evaluated by a team from Durham University led by Professor Stephen Gorard.

Stephen Gorard is Professor of Education and Well-being, and Fellow of the Wolfson Research Institute at Durham University. He is a Methods Expert for the US government Institute of Education Science, member of the ESRC Grants Awarding Panel, and Academician of the Academy of Social Sciences.

His work concerns the robust evaluation of education as a lifelong process, focused on issues of equity and effectiveness. He regularly gives advice to governments and other policy-makers, including oral and written advice to the House of Commons Education Committee every year since 2003. He is also a widely read and cited methodologist, involved in international and regional capacity-building activities, and used regularly as an adviser on the design of evaluations by central and local governments, NGOs and charities. He is currently an evaluator for the European Commission Directorate-General for Education and Culture, the Department of Work and Pensions, the Food Standards Agency, the Learning and Skills Information Service, and the Education Endowment Foundation. He is author of nearly 1,000 books and papers.

Contact details:

Professor Stephen Gorard
School of Education
Durham University
Leazes Road
DH1 1TA

p: 0191 334 8419
e: s.a.c.gorard@durham.ac.uk

Contents

Executive Summary

The project

Switch-on Reading is an intensive 10-week literacy intervention. It is delivered on a one to one basis by staff, most commonly teaching assistants, who have been trained in the approach. The purpose of Switch-on is to achieve functional literacy for as many pupils as possible, and so to close the reading achievement gap for vulnerable children working below age-expected levels. It is inspired by the well-established intervention Reading Recovery, which is teacher led and delivered over a 12-20 week period.

In this evaluation, the programme involved regular sessions for pupils who had not achieved Level 4 English at Key Stage 2. The identified pupils in Year 7 attended regular 20-minute reading sessions over the course of the Spring term. The students were removed from class to attend the sessions, which aimed to improve their reading comprehension and fluency. Each session required students to read from four different books graded on the basis of their difficulty. Training and support for staff was provided by the Every Child a Reader staff of Nottinghamshire Local Authority.

What impact did it have?

The overall result was an effect size of +0.24, based on the pooled standard deviation of the post-test score for both groups, meaning that the programme made a noticeable positive impact. This effect can be envisaged as suggesting that on average pupils receiving the intervention would make approximately three additional months' progress over the course of a year compared to similar pupils who did not.

The evaluation identified positive results for all groups of pupils (defined by sex, first language, ethnicity, special educational needs, free school meal eligibility and measured attainment at the outset). However, it is important to note that conclusions about specific groups of pupils are more tentative than the overall finding, because the study was primarily designed to test the average impact on all identified children.

Pupils with low attainment prior to the intervention showed particularly positive results, making five additional months progress on average. Pupils eligible for free school meals and pupils identified as having special educational needs made four additional months progress on average. As such, this evaluation suggests that Switch-on can be an effective intervention for weak and disadvantaged readers at the stage of transition to secondary school.

The intervention was generally well-conducted and the pupils seemed to enjoy the sessions. Staff needed training and then some monitoring to ensure that they adhered to the protocol devised by the developers, which was necessary for the intervention to have the largest possible effect.

Group	Number of pupils	Effect size*	Estimated months' progress	Evidence strength**
All pupils	**308**	**+0.24**	**+3**	★ ★ ★ ☆ ☆
Lower attainers	156	+0.39	+5	
FSM-eligible	98	+0.36	+4	
SEN reported	225	+0.31	+4	

*Effect sizes are based on differences in post-test scores, except for the FSM-eligible results, which are based on raw gain scores. See full report for explanation.

Evidence ratings are a new measure under development based on a number of factors including study type, size and drop-out. Ratings are provisional and are not given for sub-group analyses, which will always be less secure than overall findings. For more information about ratings visit: **www.educationendowmentfoundation.org.uk/evaluation.

How secure are the findings?

The evaluation was set up as a small-scale efficacy trial to test the impact of Switch-on as delivered with the developer leading the training and overseeing the provision of the intervention. Efficacy trials seek to test evaluations in the best possible conditions to see if they hold promise, but do not seek to demonstrate that the findings hold at scale in all types of schools. The findings do not indicate the extent to which the intervention will be effective in all schools since the participating schools were intentionally selected within one local authority, and training was provided by the programme developers. To test this question, a future evaluation run on a larger scale in a wider variety of areas could be conducted.

The findings are based on a randomised controlled trial, with individual random allocation to groups and a waiting list for pupils who were initially not selected to receive the intervention. There was low dropout and no sign of post-allocation demoralisation, indicating that the findings are not biased. Though a trial involving 400 pupils would have been ideal, with 308 cases in 19 schools this is nonetheless the largest trial conducted to date of this kind of 10-week reading intervention.

The results were analysed in a number of ways. The same result appeared whether the data were analysed in terms of 'gain scores' (which are based on the improvement made by pupils who receive the intervention compared to those in the control group), or by only comparing the post-test scores of those who received the intervention with those in the control group. The same effect size also appeared for each specific group of pupils when the gain scores were used instead of the post-test scores. This is because these groups (such as boys and girls) were balanced in terms of pre-test scores between the treatment group and control group by the randomisation. The one exception was free school meal eligibility (FSM): FSM pupils in the treatment group had, by chance, a higher pre-test average, meaning that the post-test score is a less accurate assessment of the intervention's impact. It is also possible that the age-standardised pre-test scores may have distorted the progress made by very low-scoring pupils. As a consequence, the table above reports an effect size based on raw gain scores for FSM-eligible pupils.

The results found here are consistent with those of a similar 10-week reading intervention, and with the 12-20-week Reading Recovery intervention that has been trialled in the US. The existing evaluation of Switch-on prior to this study was a small, developer-led study with 100 pupils in Key Stage 2. That trial showed a larger effect size but followed a less robust methodology.

How much does it cost?

The cost of the approach is estimated at £627 per pupil. This estimate includes resources (estimated at £77 per pupil), direct salary costs of teaching assistants (£500), initial training (£32) and on-going monitoring and support (£18). Estimates are based on a school delivering the intervention to 24 pupils and training four teaching assistants.

Key Conclusions

1. Switch-on Reading appears to be effective for weak and disadvantaged readers at the stage of transition to secondary school.

2. It can be delivered by teaching assistants after two-days of training, and full training and support is required for all relevant staff.

3. Challenges to successful implementation may include timetabling and the availability of age-appropriate texts.

4. There is a tendency for some staff to stray away from the explicit schedule and this is likely to reduce the programme's impact. Regular monitoring will increase fidelity but may also increase cost.

5. Further research is required to understand: i) whether Switch-on can have an impact in all types of schools; ii) whether the impact lasts, iii) which are the essential components of the intervention, and iv) whether any harm is done to progress in other subject areas due to the time out of class required by the intervention.

Introduction

Intervention

This is a report of an evaluation of Switch-on Reading. The model of Switch-on being evaluated here was provided for Year 7 pupils in mainstream secondary schools in Nottinghamshire. The intervention is a short-term individualised reading programme for pupils who have not achieved Level 4 English at Key Stage 2 (KS2). The intervention was delivered over 10 weeks and consisted of regular 20-minute one to one reading sessions with Switch-on trained staff members such as teaching assistants and librarians. The purpose of Switch-on is to achieve functional literacy for as many pupils as possible, and so to close the reading achievement gap for vulnerable children working below age-expected levels. Switch-on reading is based on the principles of an intervention called Reading Recovery (RR), which is aimed only at children aged 5 years and 9 months to 6 years and 3 months, but is shorter in duration. As a result, if Switch-had a positive impact on pupil outcomes it could be a cost-effective alternative to Reading Recovery and several other one to one reading interventions.

Background

Switch-on is based on some of the principles of a long-standing intervention called Reading Recovery (RR), first developed in the 1970s by Marie Clay from the University of Auckland, New Zealand. Reading Recovery (RR) is an intensive one to one intervention for the lowest performing 20% of first graders, and has been used world-wide in the US, Australia, New Zealand and the UK. Each RR lesson incorporates 10 principles: phonological awareness, visual perception of letters, word recognition, phonics/decoding skills, phonics/structural analysis, fluency, comprehension, early intervention, individual tutoring and a balanced literacy approach.

In a What Works Clearinghouse (WWC) evaluation of RR, a review of relevant evidence found only 4 of 78 studies that met minimal evidence standards, and even these randomised controlled trials (RCTs) were rather small in scale. Baenen et al. (1997) involved 168 first-grade pupils randomly assigned to RR or a control group. Pinnell et al. (1988) compared 38 RR first-grade pupils with a control group of 53 and a third group of 96 students taught RR outside regular class as well as receiving tuition from a RR-trained teacher in their regular class. Pinnell et al. (1994) evaluated RR in 10 schools, but the WWC review only included the results from 8 schools where students were randomly assigned (31 pupils in the intervention group and 48 in the control group). Schwartz (2005) compared the impact of RR on 37 students randomly assigned to intervention as opposed to a control (n=37). One other small study met WWC criteria with reservations because it was not a randomised controlled trial (Iverson and Tunmer 1993). It included 64 first-grade pupils (32 intervention and 32 comparison).

Of these five studies, four reported positive and 'statistically significant' effects of RR on first-grade general reading achievement. All four of these studies used the Observation Survey subtests for Dictation and Writing Vocabulary. Pinnell et al. (1994) also used the Gates-MacGinitie Test and the Woodcock Reading Mastery Test. Baenen et al. (1997) did not find statistically significant effects using the frequency of grade retention as an outcome measure. RR addresses four outcome measures of literacy – 'alphabetics', reading fluency, comprehension and general reading achievement. The results of the five WWC studies are rather mixed in terms of the impact on these four measures: two RCTs showed positive and statistically significant effects of RR on the overall alphabetics outcome measure (Schwartz 2005, Pinnell et al. 1988), although it should be noted that Schwartz did not control for pre-test differences between the two groups. When pre-test differences were taken into account by WWC the effect was neither significant nor large enough to be considered important. The same studies reported statistically positive effects for phonics (Schwartz 2005, Pinnell et al. 1988) using the Word

Recognition subtest of the Observation Survey, although subsequent WWC calculations showed that the Pinnell et al. result was not statistically significant.

Schwartz (2005) reported positive effects of RR on phonemic awareness (+0.25), but the effects were not statistically significant. The same study, however – the only study that examined the effects of RR on reading fluency – found that it had positive and significant effects on reading fluency of first graders.

Two studies evaluated the reading comprehension aspect of RR: one reported statistically positive effects used the Reading Comprehension subtest of the Comprehensive Test of Basic Skills (Pinnell et al. 1988), while the other (Schwartz 2005) found no statistically significant effect using the Degrees of Reading Power Test. This suggests that the effect of the intervention can vary depending on the kind of tests used. It is therefore worth emphasising that several outcome measures were assessed using the Clay Observation Survey which was created by the developer of the programme. Some of the studies also used Clay's book level scores. This use of advocate-generated tests can be a threat to rigorous evaluation and the wider validity of the results (Gorard 2013).

Center et al. (1995) conducted an evaluation of RR in 10 schools in New South Wales, Australia. Specific students in these schools were randomly assigned to treatment (n=31) or control (n=39) groups. There was also a matched comparison group from five other schools (n=39). This study was excluded by WWC because it did not establish baseline equivalence between the comparison and treatment groups; there was also considerable dropout after randomisation, and there were problems of diffusion. Results suggest that after 15 weeks, treatment students outperformed control students on all measures of reading achievement, but in only one out of three tests of metalinguistic skills. However, after 30 weeks no differences were observed between RR and control students on 7 of the 8 measures.

More recently, Hatcher et al. (2006) evaluated Reading Intervention on 77 children taken from 14 schools in the UK. This programme is similar to RR in that it involves daily 20-minute sessions, but alternates between small group and one to one sessions. The children were randomised to either the 10-week (n=39) or 20-week intervention (n=38). The study reported promising results with medium to large effect sizes on measures of reading, letter knowledge and phoneme awareness (British Ability Scales reading, +0.69; Early Word Recognition reading, +0.79; letter knowledge, +0.94; phoneme completion, +0.46). The study found that the progress made by children in the first phase of the study (20-week group) slowed down in the second 10 weeks of the study, and children in the second phase (who did not receive any special treatment in the first 10 weeks) caught up with the first group of children when they started the intervention. This might mean that around 10 weeks is the ideal length for the treatment. Regression analysis suggested that those who were in receipt of free school meals and those with very low initial reading skills were least likely to respond to the intervention.

Even more recently, Tanner et al. (2011) compared 57 RR schools with 54 other schools, and reported that pupils at the RR schools had performed better. However, the schools were not randomised to treatment and nor was baseline equivalence established. The comparator schools had more boys, more pupils in receipt for free school meals (FSM) and more pupils identified as having special educational needs (SEN).

Reading Recovery has therefore, and rather surprisingly, not previously been rigorously evaluated at scale, although studies with weaker designs have claimed an impact on reading age with an effect size as large as +1.3.

There has been even less evaluation of Switch-on itself. The Switch-on programme has previously been evaluated with Key Stage 2 primary age children (Coles 2012). Of 100 pupils randomised to treatment or control, 8 are unaccounted for (7 from the control). For the remaining 92, the effect size for Switch-on Reading was +0.8. It was the promise of success based on this evaluation that prompted the project leaders to apply for EEF funding to replicate the intervention with secondary school pupils. This new evaluation is different: it is at least twice the scale of anything done previously, it looks only at the reading element of Switch-on, and for the first time it is tried with pupils just arriving in secondary school (i.e. it is used as a transitional literacy catch-up scheme).

Evaluation Objectives

- To evaluate the impact of the Switch-on intervention on the reading performance of first year secondary students (Year 7) who had not previously achieved Level 4 English at KS2.
- To evaluate the impact of Switch-on intervention in relation to key background characteristics of the pupils.
- To record the fidelity of the Switch-on intervention through process evaluation, and assess the ease with which it could be scaled-up and implemented with a larger groups of schools.

Project team

The programme was developed by José Coles and Paula Burrell, the Every Child a Reader (ECaR) teacher leaders for the Nottinghamshire Local Authority. They trained all staff and oversaw delivery. The team also included trained staff members who provided extra support to the teachers in the schools.

Methodology

Trial Design

The evaluation was based on a relatively simple one term waiting list design. Nineteen schools were recruited to take part in the intervention. The schools agreed for half of their relevant pupils to be individually randomised to immediate intervention, or intervention after one term. The latter group formed the control. Eligible individual students were randomised to control or treatment conditions using the waiting list design. The Phase 1 intervention group of 157 pupils would be involved in reading every day, aiming for at least 40 sessions in the minimum of 10 weeks. The Phase 2 group of a further 157 pupils continued with normal lessons and any interventions or programmes that were also available to Phase 1 pupils and that would have been used anyway in the absence of this evaluation. After one term, the Phase 2 pupils received the intervention. The pre-test was conducted at the outset before randomisation, and the post-test was conducted before Phase 2 pupils received the intervention.

This design is ethical, reduces the dangers from post-allocation demoralisation, avoids bias caused by knowledge of grouping when taking the pre-test, and allows an unbiased estimate of the impact of one term of intervention. The major drawback is that it does not permit consideration of the longer term impact of the intervention.

Eligibility

The model of Switch-on being evaluated here was provided for Year 7 pupils in mainstream secondary school settings who had not achieved Level 4 English at Key Stage 2 (KS2). All local authority secondary schools in Nottinghamshire were eligible to take part. The schools which agreed to take part signed an agreement specifying that being part of the programme entails agreement to the evaluation. A consent form was also sent out to parents of pupils eligible for the programme by the school. The letter was drafted by the programme developer, and made it clear that consent to being part of the programme, as opposed to opting out, also involved being part of the evaluation. Only pupils with consenting parents were involved in this intervention.

Intervention

The programme began with a two-day training event for staff in order to equip them with the necessary skills to carry out the intervention. The training provided background information on the rationale and content of the Switch-On approach, as well as demonstration of the use of the running record, coding and analysis. Participants then practised the use of the protocol. There was a video demonstration of what a Switch-on reading session looks like. Eighty-three members of the staff from the 19 schools were trained to conduct Switch-on. Of this number 71 went on to deliver the intervention.

The programme was delivered by Nottinghamshire Local Authority and involved a project team with experience of delivering both Switch-on and Reading Recovery (at primary level). The aim was to reduce the reading achievement gap for at-risk children working below the age-expected levels in schools in Nottinghamshire. The programme was designed to last 10-12 weeks from the start of the Spring term 2012/13. In most schools some of this time was taken up with pre-testing, with the core intervention still taking place over a 10-week period. This was deemed sufficient time (see Hatcher et

al. 2006 above), especially as only the reading component was introduced and evaluated. The target children were identified in the first term of secondary school, and pre-tested using the GL New Group Reading Test A. They were randomised by the evaluation team to receive the intervention in either Phase 1 (Spring term) or Phase 2 (Summer term). The Phase 2 pupils undertook normal lessons during the first term, including any other educational interventions that were already in place.

In Phase 1, each member of the treatment group was allocated to an individual member of staff. These staff included SENCOs, librarians, teachers, and teaching assistants. The clear majority were teaching assistants. Each member of staff looked after no more than four pupils. Each pupil was given a schedule in which to come out of one normal class per day for 20 minutes at a time for the Switch-on session. The schedule was arranged so that parts of different lessons were missed. The Switch-on session was delivered one to one by the trained member of staff.

In the first session, the materials used were selected to suit the reading age of the pupil as assessed by the pre-test and prior attainment. Switch-on Reading revolves around appropriately matched books that have been finely graded in bands and levels to provide small changes in challenge over time. Of course, these books had not been used with Year 7 pupils before and so one question was whether the pupils and staff found them suitable. Where there was a clear mismatch in the early sessions, the level was adjusted until the reading age required offered just enough challenge. The books themselves included fiction and non-fiction with lots of visual images to stimulate students' interest in reading and as well as providing clues for comprehension.

Each Switch-on Reading session should have consisted of:

- reading a familiar book (perhaps the first 100 words only)
- discussing the textual content of books (text, cover pages and blurbs)
- provoking student interest by discussing visual content
- reading the text and using the running record sheet for analysis of reading
- providing feedback to the student
- introducing a new book.

Each session thus incorporated the revision of a familiar text, the introduction of new vocabulary, practising phonics and also the improvement of comprehension through questions and discussion. In each session the student should have read excerpts from four books.

At some point in the 20-minute reading session the member of staff records the reading assessment of the pupil on a sheet, and makes an inventory of errors such as words missed, substituted with another, mispronounced or repeated; also recorded are any self-corrections and appeals for help. The form for recording these events and the rules for completion are standardised and an integral part of the intervention. Part of the intervention also involves analysis of errors. The average number of errors is calculated and this determines which book set is to be followed next. After each book is read, the adult trainer praises the child when an effective reading strategy is observed (e.g. 'you made it sound interesting'), and prompts the student to use new strategies where behaviour has not been effective or advice has been ignored (e.g. the advice to look for full stops). Pointing to the text for the child is to be avoided as the intention is to develop independent, active readers.

The rationale for use of the running record in the intervention is that it is pupil driven, provides evidence of pupil progress and achievement, helps ensure use of the appropriate level of books, and generally informs planning, choice of books, teaching focus and strategies. The rationale for the use of set books and for reading new books each time is that it engages pupils in texts, develops language and vocabulary through discussions, and provides an opportunity for problem-solving and new learning.

Outcomes

At the outset, all eligible pupils took the GL New Group Reading Test A, administered individually. The test was administered by the schools themselves in conjunction with the project team and assisted by GL. Because this pre-test took place before randomisation into two groups, the process was 'blind' as to the treatment group. In addition, the prior KS2 assessment results in literacy can be used as a secondary pre-test score.

Once the intervention was complete, both groups took the GL New Group Reading Test B. Again this was administered on an individual basis by schools. Because the staff and pupils were no longer blind as to who was in which group, the evaluators sampled this testing process and observed the tests in operation. Both the pre- and post-tests were conducted on-line to ensure standard timing and marking, to reduce the potential influence of staff, and to create instant results.

The primary outcome measure is the difference in the post-test scores between the arms of the trial, expressed as an effect size, based on balance between the groups in terms of prior scores (the pre-test), and pupil background characteristics. The result was confirmed by considering the difference in the gain scores between the arms of the trial, where the gain is the average difference between individual scores on tests A and B. A secondary outcome measure is the average residuals between the actual scores on test B and the predicted (modelled) scores based on prior KS2 literacy assessment and pupil background characteristics.

Sample size

The initial outline proposed by the developer of the intervention was for an individual-level randomisation of 480 Year 7 pupils to one of two groups: half of these would receive Switch-on from January 2013, and the other half a term later. This initial proposal preceded the recruitment of schools and the identification of eligible pupils within each. In fact, a total of 314 pupils were identified as eligible within the 19 schools agreeing to participate, giving an initial group size of 157 pupils per arm. This is still a large individually-randomised study but somewhat less than ideal in scale.

After pre-tests for all pupils, a pseudo-random number generator was used to assign each pupil to either the treatment or the waiting group. The initial outline also suggested an effect size (d) of 1.3 for Reading Recovery (no data was available at that time for Switch-on itself); the What Works Clearinghouse (2008) suggests a medium to large effect size. Using Lehr's approximation for an 80% chance of detecting a presumed effect size of 1.0 with 5% alpha, the minimum sample size required per arm is only 16 individuals (Gorard 2013). As a much larger individually-allocated sample, this study should have considerable power, assuming that the effect size is as indicated, or, put another way, 240 cases per arm, as proposed, would have a reasonable chance of detecting an effect size as low as 0.25.

Randomisation

The recruitment of schools in Nottinghamshire was conducted by the intervention project lead, and the student participants were identified and selected on the basis of weak performance in reading by the individual school teachers. The schools then each provided the evaluation team with the unique pupil numbers (UPNs) of those pupils who had achieved below secure level 4 in KS2 for English. The process of random allocation to treatment and control groups was conducted after the pre-test in all of the participating schools, and was carried out by the lead evaluator in the presence of another

researcher. The list of participating pupils was prepared. The procedure involved a set of playing cards with an equal number of odd and even cards, and one card per pupil. The cards were mechanically shuffled, and then dealt in turn to represent each pupil in that order in the list of participants. Odd cards represented Phase 1, and even cards Phase 2. The identities of relevant pupils with their group allocation were then sent to the respective schools and to the project lead. The order of the cards was retained for a short time in case of queries.

Analysis

The powerful design of the study means that any complex analysis is redundant, and there is no issue of 'statistical' generalisation to a larger group of schools, since the participating schools were selected purposively within one local authority. All eligible pupils in participating schools were allocated to a group, and therefore the study is of this population. The primary outcome measure for each pupil was the post-test score, and a secondary measure was the gain score or difference between the pre- and post-test measurement. The latter was used to compare with the post-test only scores and identify any situation where imbalance in pre-test scores between the original groups might have caused a problem. The post-test scores and gain scores were averaged for each group (or arm of the trial), and the difference between these averages was expressed as an 'effect' size. The effect size used was Hedges' g with correction, based on dividing the difference by the pooled standard deviation of the gain scores for both groups. This analysis was repeated for sub-groups of pupils separately, including those above the median pre-test score compared to those at or below the median, boys and girls, pupils with SEN and others, and pupils eligible for FSM and others.

Using the staff records of the number of sessions attended (zero for all Phase 2 pupils) it was possible to estimate any link between dosage and gain scores by running crossplots and calculating Pearson's R correlation coefficients. Pearson R correlations were also run for pre- and post-tests, gain scores, and background characteristics such as FSM. This was done in an attempt to explain some of the results.

Two simple multivariate regression models were also created: one model used the gain score as the 'dependent' variable, the other used the raw post-test score. The possible predictors added to the models included the pre-test score or prior attainment, the pupil age in months at pre-test, the precise time period between pre- and post-test for that individual, and the known pupil characteristics such as sex, FSM, SEN and EAL. These predictors were all added to the model in the first step. Once these values had been accounted for, the treatment group (a binary variable) and the number of sessions attended (dosage) were added as a second step.

Process evaluation methodology

The process evaluation was conducted jointly between Switch-on project leaders and the independent evaluators. The project leaders conducted the training of staff (see below), monitored the intervention and collected the formal records and the views of staff. The evaluators observed the training, the teaching and testing, used the texts and documents relating to the intervention, and conducted face-to-face interviews with staff, students and project members. These interviews were usually conducted without a formal schedule, and arose as the situation allowed. The observations of staff training and of the implementation of the programme in action were as simple, integrated and non-intrusive as possible. The schedule of visits was agreed with the intervention team and the schools, and some interviews were arranged at that time. Schools agreed to be part of the evaluation when agreeing to be part of the intervention.

Attention was given to the training of staff, participants' perceptions, their receptivity to the programme and the delivery of the intervention in the schools. Two evaluators attended both training days as participant observers during which they talked to other participants and project leaders to get their perceptions of the programme. Participants were mainly teaching assistants, SENCOs (Special Educational Needs Co-ordinators), librarians, literacy heads and school heads. The subsequent fieldwork in schools included observations of these staff delivering the intervention, noting inconsistencies or any departures from the programme protocol. The observation of the one to one sessions had to be carefully orchestrated so as not to disrupt the programme. Some members of staff requested that the evaluators did not sit in on some sessions, largely for educational or pastoral reasons. On each visit, the evaluators also interviewed the school leads, relevant teaching staff, and spoke with small groups of participating pupils. The evaluators also considered the resources used (such as story books, running records and pupil progress charts), and asked for staff and pupils' perceptions of these materials. The interviews and field notes were part-transcribed and shared between the evaluation team, all school and pupil names having been been altered to pseudonyms.

The process evaluation was useful in assessing fidelity to treatment. The perceptions of participants provided indications of any resentment or resistance to the programme, and were useful in identifying potential issues or barriers which should be addressed for any future scaling-up.

Impact evaluation

Timeline

December 2012

Schools were recruited. Training was conducted over two days and attended by teaching assistants. This took place on 10th and 12th December. Two members of the evaluation team attended and participated in the training.

January 2013

The pre-test was conducted between 7th and 12th January. Students rated at Level 3 and below on the NGRT test, or otherwise identified as working at below secure Level 4, were selected for the intervention. Students were randomised using a pack of cards to Phase 1 (treatment group) receiving immediate intervention or Phase 2 (control) to receive intervention a term later. Randomisation was completed before the Christmas break.

January to March 2013

Immediately after the pre-test and randomisation, the intervention began. Three members of the evaluation team visited schools as part of the light touch evaluation of the on-going process. Background data were collected.

February 2013

A follow-up meeting was conducted on 14th February by project leaders to review the progress of schools involved in the intervention. Schools fed back to project leaders about their students' and their own views about the books used. Student attendance was also discussed.

March to April 2013

The post-test was administered between 25th and 28th March. Background data were updated. Outcome measures were analysed as soon as tests were completed and results sent to evaluation team. Synthesis with process evaluation took place simultaneously with the analysis of test results..

Participants

The school recruitment process was conducted by the project lead members who work for the education department in Nottinghamshire Local Authority and are responsible for the schools support and improvement plans. The networking process for the recruitment involved meetings with the head teachers and school literacy coordinators. The aims of the Switch-on project were discussed and the process of RCT evaluation was also explained to the schools. Two schools initially agreed to participate but dropped out before the pre-test was conducted. It was reported that they were not prepared to adjust their timetable to accommodate the regular 10-week reading sessions. It must be noted here that the project timeline was determined by funding deadlines, and all stages of the trial had to be executed within this time scale.

Figure 1. Flowchart of sampling, allocation and attrition

```
                    ┌─────────────────────────────────┐
                    │ Assessed for eligibility (n=314) │
                    └─────────────────────────────────┘
                                   │
                                   │              ┌──────────────────────────┐
                                   │              │ Excluded (n=0)           │
                                   ├─────────────▶│                          │
                                   │              │ Not meeting              │
                                   │              │ inclusion criteria       │
                                   │              │ (n=0)                    │
                                   │              │                          │
                                   │              │ Declined to              │
                    ┌──────────────────────┐     │ participate (n=0)        │
                    │ Randomised (n=314)    │     │                          │
                    └──────────────────────┘     │ Other reasons (n=0)      │
                                   │              └──────────────────────────┘
```

Allocated to intervention (n=157)	**Allocation**	Allocated to control (n=157)
Received allocated intervention (n=157)		Received allocated intervention (n=157)
Did not received allocated intervention (n=0)		Did not received allocated intervention (n=0)

Lost to follow-up (n=2)	**Follow-Up**	Lost to follow-up (n=4)
Discontinued intervention (n=0)		Discontinued intervention (n=0)

Analysed (n=155)	**Analysis**	Analysed (n=153)
Excluded from analysis (n=0)		Excluded from analysis (n=0)

Table 1. School and pupil numbers for the trial

Category	Total number
Schools initially recruited	21
Schools participating in the trial	19
Pupils in the pre-test	314

Twenty-one of the 37 secondary schools in Nottinghamshire agreed to take part. Two schools dropped out before testing or randomisation, but then participated when the waiting-list students were given the Switch-on intervention. The 19 participating schools identified 314 pupils for the intervention. See flow chart for details (Figure 1). Around half were in the ex-mining areas of Nottinghamshire. The schools ranged in size from around 114 pupils to over 1,700. FSM eligibility ranged from 6% to over 30%, pupils not speaking English as their first language ranged from 1% to 12%, and pupils with statements of special educational need or receiving School Action Plus ranged from 3% to 11% (see Table 2 below).

Table 2. Participating schools

Category	Age range	No. pupils	% SEN	% EAL	% FSM
Academy Converter	11 to 18	775	4.2	12.1	22.8
Academy Converter	11 to 18	1730	7	6	11.1
Community School	11 to 18	749	6.5	1.5	16.6
Community School	11 to 18	918	8.3	1.3	17.4
Foundation School	11 to 18	953	5.1	7.9	23.2
Academy Sponsor led	11 to 18	939	3.6	0.3	32.6
Academy Sponsor led	11 to 16	448	6.9	1.8	16.3
Academy Converter – Mainstream	11 to 18	1063	2.4	8.7	9.4
Academy Sponsor led	11 to 18	1057	10.8		17.7
Academy Converter	11 to 18	1186	3.9	2.2	12.3
Academy Sponsor led	11 to 18	114	3.4	0.6	18
Academy Converter – Mainstream	11 to 18	576	6.8	1.4	28.7
Academy Sponsor led	11 to 18	900	4.4	2.6	14.6
Academy Converter – Mainstream	11 to 18	1004	3.5	2.3	17.5
Academy Converter – Mainstream	11 to 18	1424	2.8	9.3	8.8
Academy Sponsor Led	11 to 18	770	4.8	2.7	28.9
Academy Converter – Mainstream	11 to 18	1556	3.1	2.1	6
Academy Converter – Mainstream	11 to 16	664	5.3	8.9	16.3

Category	Age range	No. pupils	% SEN	% EAL	% FSM
Academy Sponsor led	11 to 16	448	6.9	1.8	16.3
Academy Sponsor led	11-16	No data available			

In total, 19 schools and 314 Year 7 pupils took part in the study: 157 pupils were assigned to treatment and 157 to control. The sample included students from disadvantaged backgrounds, defined as those eligible for free school meals, those with a range of learning disabilities (SEN) and those for whom English was a second language. In the final analysis six students dropped out for various reasons. One took the pre-test (repeatedly) but his school were unable to record the score. His post-test score was 78, and he would have been in the control. Five others took the pre-test but did not sit the post-test (Table 3). One left the school and could not be traced, initially scored 78 and would have been in treatment. One left the school and their new school was not able to arrange the post-test, initially scored 64 and would have been control. One changed schools, one could not get their score saved at pre-test, one refused to cooperate and one was persistently absent at post-test (perhaps excluded). Although this loss of data, and the reduction of the sample to 308 pupils, is unfortunate, there is no specific reason to believe that this dropout was biased or favoured one group over the other.

Table 3. Pupils allocated to groups but with no gain score, and reason for omission

Allocation	Pre-test score	Post-test score	Reason
Treatment group	78	-	Left school, not traced
Treatment group	73	-	Long-term sick during post-test
Control	74	-	Left school, new school would not test
Control	75	-	Withdrawn, personal reasons
Control	-	70	Pre-test not recorded
Control	73	-	Permanently excluded by school

There were instances of students leaving school, or of schools deciding to swap intervention and control students. One school was keen to prematurely test a boy who was in mid-programme (and doing well) but who was moving to a new school. It was made clear to the project leaders and to the school that the pupil was to be tested in the school he was moving to (if known) at the same time as all the other pupils, even though he might not have completed the full dosage of the intervention. There were also other cases where schools wanted to switch students round. One school requested that a student identified for intervention be allowed to switch with another student who had not been identified to receive treatment on the grounds that the latter student was weaker and would benefit. It transpired that the intervention student, although more able, had performed badly in the pre-test because he did not complete it and so was identified for treatment. Another pupil, although weaker, was not identified for treatment initially because he did not take the pre-test, but when subsequently tested he was found to be rather weak. So the school decided to swap the students around. In this case, we were fortunate that the school communicated their intention with the project lead, and the project lead consulted with us before making any decision. The evaluation would have gone awry if schools had made such decisions without consulting the evaluators. All pupils were analysed in terms of their initial allocation (intention-to-treat) as far as the post-test data permitted.

Pupil characteristics

The pupil characteristics in the achieved sample are provided below (Tables 4 to 8). The tables include basic demographics along with information on any missing data. In order to keep the pupil selection and group allocation procedures random and unbiased, the following characteristics were considered only after the randomisation procedure.

Table 4. Sex of participants in each treatment group

	Male	**Female**	**Missing data**
Treatment	93	62	0
Control	86	67	0
All pupils	179	129	0

Table 5. FSM status of participants in each treatment group

	FSM	**Not-FSM**	**Missing data**
Treatment	46	109	0
Control	52	101	0
All pupils	98	210	0

Table 6. SEN status of participants in each treatment group

	SEN	**Not-SEN**	**Missing data**
Treatment	115	40	0
Control	110	43	0
All pupils	225	83	0

Note: Variation in terminology and recording between schools means that the level of SEN may be over-estimated

Table 7. EAL status of participants in each treatment group

	English first language	**English not first language**	**Missing data**
Treatment	149	6	0
Control	147	6	0
All pupils	296	12	0

Table 8. Ethnicity of participants in each treatment group

	White British	**Not White British**	**Missing data**

	White British	Not White British	Missing data
Treatment	140	15	0
Control	141	12	0
All pupils	281	27	0

Overall, the two groups are reasonably well-balanced in terms of these background characteristics.

Outcomes and analysis

The overall effect size of the intervention was +0.24 standard deviations of the overall gain score, meaning that it has shown a small noticeable positive impact (Table 9). Both randomised groups had very similar scores at the outset (NGRTA), which suggests that the randomisation was effective and so the test of the intervention was fair in that respect. The headline finding of this study is therefore that the intervention, as conducted, is effective. It is unlikely that the gain scores of the missing six pupils (see above) would have been so divergent between groups that they would have altered the order of magnitude of this effect size.

Table 9. Estimated impact of Switch-on Reading Programme – overall with gain score

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	155	76.53	80.93	4.40	8.18	-
Control	153	76.14	78.73	2.59	6.53	-
Overall	308	76.33	79.84	3.50	7.45	+0.24

Note: the 95% confidence interval for this effect size lies from 0.02 to 0.47.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

Table 10 shows the same results assessed as though for a post-test only design. This confirms the same substantive result of a small positive impact. This version of the analysis is suggested by concerns about the potential propagation of initial errors in the pre- and post-test design (Gorard 2013). It also allows the inclusion of any cases with post-test scores but missing pre-test data. The similarity of results is reassuring.

Table 10. Estimated impact of Switch-on Reading Programme – overall with post-test only

Treatment group	N	NGRTB	Standard deviation	'Effect' size
Switch-on	155	80.93	9.23	-
Control	153	78.73	9.29	-
Overall	308	79.84	9.33	+0.24

Note: the 95% confidence interval for this effect size lies from 0.021 to 0.46.

Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

If the pupils are separated into those eligible for free school meals (FSM) and others, the effect size assessed on the basis of gain scores is smaller for FSM pupils (Tables 11 and 12). This finding does not affect the headline result, and is necessarily based on a smaller number of cases in each Table. The finding is also not from a randomised-controlled trial, since the design used here was for all identified low attaining readers. A trial for the impact on FSM pupils only would have been different. The finding may, however, suggest that Switch-on is not particularly effective for the most disadvantaged pupils, as previously also reported for a similar catch-up scheme by Hatcher et al. (2006). One tentative explanation for this could be a difference in the training and experience of the teaching assistants who delivered the intervention. In south Nottinghamshire, with a greater proportion of areas with low FSM, the TAs employed from the local area may be more likely to be graduates and even trained teachers working as TAs. This could have led to differential implementation between north and south, but there is no direct evidence to support this. Another explanation could be that FSM-eligible pupils have other problems that are not addressed by this intervention.

Table 11. Estimated impact of Switch-on Reading Programme – gain scores, FSM pupils

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	46	77.30	80.15	2.85	7.71	-
Control	52	74.88	77.37	2.48	5.91	-
Overall	98	76.02	78.67	2.65	6.78	+0.05

Note: the 95% confidence interval for this effect size lies from -0.34 to 0.45.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

Table 12. Estimated impact of Switch-on Reading Programme – gain scores not FSM pupils

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	109	76.20	81.26	5.06	8.32	-
Control	101	76.78	79.44	2.65	6.85	-
Overall	210	76.48	80.38	3.90	7.72	+0.31

Another possible explanation comes from consideration of the relationship between pre- and post-test scores (Figure 2). The crossplot below shows a reasonably strong near-linear relationship between the two scores for any individual, with the exception of scores at the very lowest level for the pre-test. There are pupils with a pre-test score of 69 with post-test scores ranging from 69 to 105. The flattening of the graph shows that some pre-test scores have been constrained by an artificial threshold that is higher than in reality. This threshold effect has two implications: first, the overall result for the trial could be an underestimate of the effect size, since the 'true' gain score for some low achievers should be higher; second, if FSM-eligible pupils were disproportionately among those with an initial score of 69, then this issue would disproportionately affect them, and so depress their apparent gain scores.

Figure 2. Crossplot of pre- and post-test results
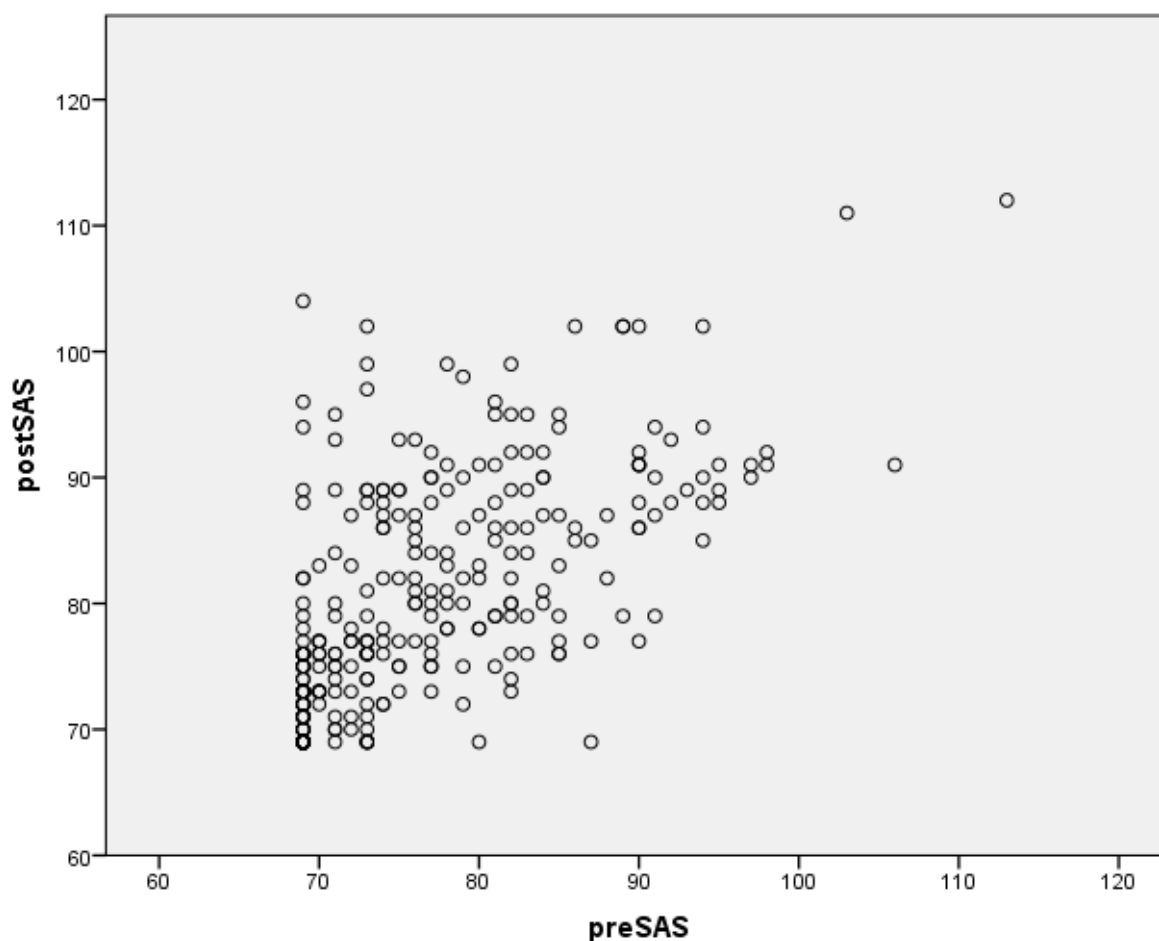
Neither of these proposed speculations have much else to support them. Of the overall 308 pupils in the trial, 94 achieved a pre-test score of 69 (apparently the lowest score available in practice for this age group). This is too many, and means that the progress of some of them will have been underestimated. However, these 98 are reasonably balanced in terms of treatment group (48% were in Phase 1, with 52% in Phase 2) and FSM (35% were FSM eligible compared to 32% in the overall 308 pupils). This is confirmed by the level of correlation between pupil characteristics and the test outcomes (Table 13). There is no linear relationship between FSM and either pre-test or gain scores, nor is there a clear link between pupil eligibility for FSM and having a special educational need.

Table 13. Pearson's R coefficient for test outcomes and pupil characteristics

	FSM	Sex	SEN	EAL
FSM	1	+0.04	+0.08	-0.07
Pre-test score	+0.03	-0.10	+0.27	-0.08
Gain score	+0.08	-0.06	+0.11	+0.10

It is noticeable in Table 13 that the FSM-eligible pupils in the control group had a lower pre-test score than the intervention group. This initial imbalance (the FSM-eligible pupils were not randomised separately, nor was the trial designed to give results for FSM-eligible pupils only) is the reason why

the gain scores are used in the considerations above. If the data are analysed in terms of post-test scores only (Table 14), thus ignoring how far the control group had to progress just to catch up, the result for FSM-eligible pupils is of the same order of magnitude as the overall result.

Table 14. Estimated impact of Switch-on Reading Programme – FSM pupils only

Treatment group	N	NGRTA	NGRTB	Standard deviation	'Effect' size
Switch-on	46	77.30	80.15	9.28	-
Control	52	74.88	77.37	9.29	-
Overall	98	76.02	78.67	6.78	+0.30

Note: the 95% confidence interval for this effect size lies from -0.10 to 0.71.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

In order to assess any limitation created by the lower threshold score of 69 in the pre- and post-tests, the analysis was also re-run using the raw scores (age before standardisation and the lower limit of 69 was imposed). This produced an effect size of +0.36, which is actually larger than the overall result using the standardised scores. It suggests that the 69 threshold led to an underestimate of the gains for FSM-eligible children.

Table 15. Estimated impact of Switch-on Reading Programme – raw gain scores, FSM pupils

Treatment group	N	Raw gain score	Standard deviation	'Effect' size
Switch-on	46	201.46	37.61	-
Control	52	186.29	44.87	-
Overall	98	193.41	42.11	+0.36

Note: the 95% confidence interval for this effect size lies from -0.04 to 0.76.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

A similar analysis to that for FSM, using the gain scores, shows that the intervention is almost equally effective for boys and girls (Table 16 and 17). Again the NGRTA scores are similar for both randomised groups at the outset, and both show improvement over time, but the gain is always greater for the intervention group.

Table 16. Estimated impact of Switch-on Reading Programme – Girls only

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	62	77.65	82.56	4.92	8.46	-
Control	67	76.97	80.18	3.21	6.60	-
Overall	129	77.29	81.33	4.03	7.57	+0.23

Table 17. Estimated impact of Switch-on Reading Programme – Boys only

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	93	75.78	79.84	4.05	8.02	-
Control	86	75.49	77.60	2.12	6.47	-
Overall	179	75.64	78.77	3.12	7.36	+0.26

Note: the effect size based on post-test scores only is +0.25.
Note: the 95% confidence interval for this effect size lies from -0.05 to 054.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

A slightly different picture emerges from consideration of pupils reported by their schools to have special needs or specific learning challenges (SEN). The effect size for these pupils is considerably higher than for the rest (Tables 18 and 19). It is understandable that a high proportion of children selected for this study will have SEN, meaning that the groups are unbalanced in size, nevertheless, it is intriguing that the intervention appears to be especially effective for them.

Table 18. Estimated impact of Switch-on Reading Programme – SEN only

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	115	75.21	79.27	4.06	7.53	-
Control	110	74.74	76.67	1.94	4.90	-
Overall	225	74.98	78.00	3.02	6.46	+0.33

Note: the effect size based on post-test scores only is +0.31.
Note: the 95% confidence interval for this effect size lies from 0.04 to 0.57.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

Table 19. Estimated impact of Switch-on Reading Programme – non-SEN only

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	40	80.33	85.70	5.38	9.86	-
Control	43	79.72	84.00	4.28	9.37	-
Overall	83	80.01	84.82	4.81	9.56	+0.12

In order to investigate these patterns further, the cases were notionally sub-divided into two groups around the median pre-test score of 73. There were 156 pupils scoring 73 or less, and 152 scoring more than 73 initially, and these were again almost evenly divided between treatment and control (Tables 20 and 21). It is clear that, despite the likely impact of the lower threshold at pre-test score 69, the intervention is considerably more important for the lowest attainers among those selected for intervention.

Table 20. Estimated impact of Switch-on Reading Programme – Lower attainers only

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	76	69.96	75.96	6.00	7.81	-
Control	80	70.19	73.13	2.94	5.64	-
Overall	156	70.08	74.51	4.43	6.93	+0.44

Note: the effect size based on post-test scores only is +0.39.
Note: the 95% confidence interval for this effect size lies from 0.07 to 0.71.
Note: this confidence interval is NOT an estimate of the scale of the true effect size. It is predicated on the reported effect being the 'true' effect.

Table 21. Estimated impact of Switch-on Reading Programme – Higher attainers only

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size
Switch-on	79	82.85	85.71	2.86	8.28	-
Control	73	82.66	84.88	2.22	7.40	-
Overall	152	82.76	85.31	2.55	7.85	+0.13

As expected, there is a reasonably strong correlation between pupils' scores on the pre-test and post-test, and between these scores and their overall gain during the period of the intervention (Table 22). More surprisingly, given the overall result, there is no such clear linear link between the test and gain scores and the number of Switch-on sessions attended by each pupil (ranging from zero for the control group to over 40 sessions). Assuming that there is impact from the intervention (above), then being on the intervention or not is what seems to make the overall difference, rather than the precise number of sessions completed by each pupil. To assess this further, dosage and treatment phase are entered into a regression model.

Table 22. Correlations between pre- and post-tests and 'dosage'

	N of sessions	Gain score	NGRTA	NGRTB
N of sessions	-	+0.10	-0.03	+0.06
Gain score		-	-0.32	+0.57
NGRTA			-	+0.60

N=245 (some pupils are missing scores for the number of sessions attended)

Table 23 presents the R values for two regression models, each based on two steps. The first model is used to try and explain variation in the gain score; the second concerns the post-test score only. In step 1, the pupil background and pre-test scores are included, and then in step 2 the binary variable for being in the treatment group or control is added along with the number of sessions attended. The model is better at explaining variation in the post-test outcome than the gain score. But for both models the bulk of the variation that is 'explained' by the variables in the model is explained at step 1. Once pupil background and prior attainment is accounted for, very little difference is made by knowing whether a pupil was in the treatment group or not. This model is not, in itself, any test of causation but it does provide a caution about the strength and importance of the intervention in relation to pupil characteristics.

Table 23. Variation explained (R) in two-stage regression model, using two possible outcomes

	Gain score outcome	Post-test outcome
Step 1: background and prior attainment	0.41	0.68
Step 2: intervention and dosage	0.42	0.69

For completeness, Table 24 presents the coefficients for all variables retained in either model. The largest of these by some way is the pre-test score. This is the best single predictor of the post-test score, followed by having English as an additional language, and then the treatment group.

Table 24. Standardised coefficients for the regression model in Table 9

	Gain score outcome	Post-test outcome
FSM	-0.07	-0.05
Sex (female)	+0.09	+0.07
SEN	-0.21	-0.17
EAL	-0.17	-0.13
Ethnicity (White UK)	+0.12	+0.10
Age at pre-test	-0.09	+0.01
NGRTA score	-0.36	+0.59
Step 2: N of sessions	+0.07	+0.06
Step 2: Treatment (or not)	+0.12	+0.10

Note: the precise period between pre- and post-test was not related to either outcome

It is important to recall that the evaluation was designed as a straightforward comparison of the outcomes between two groups for all pupils. Modelling and the analysis of sub-groups can yield interesting insights but do not themselves have the power of a trial.

Cost

The project leaders have provided an estimate of the full costs of implementing Switch-on Reading from scratch. Based on a school with 24 participating pupils taught by four teaching assistants, the average cost per pupil would be £627. The cost would tend to decrease in the following year, but would increase if any of the staff involved were paid more than teaching assistants. The resources (banded books and equipment) would be £1,840. The initial training would be £760 (£95 per person per day for two days). The on-going monitoring and professional development would be £450, and the direct salary costs could be £12,000. The total is £15,050 (or £627 times 24).

Process evaluation

Fidelity

In general, the members of staff observed as part of the evaluation conducted the sessions as they were trained to. The use of the reading protocol was largely adhered to although there were a couple of deviations. The main elements of the training that were most often observed in the reading sessions were:

- using a variety of text books in a 20-minute session
- talking about the text
- discussing comprehension questions posed by the teacher
- independent and out loud reading of text (at least 100 words) by the pupil
- teacher's feedback to the pupil.

One of the elements of Switch-on is the focus on individualised attention and a private space for the conduct of the intervention. In practice, however, this was not always followed and there were some variations in the quality of intervention provided by different members of staff. In one school, the librarian who conducted the session chose a quiet corner at the back of the library with a screen on the table shielding the child from others. Phonics symbols were pinned to the screen to help the pupil when she had difficulty. The librarian was encouraging and spoke to the child in a supportive voice, using praises and prompts appropriately. The pupil looked comfortable and confident in her reading.

In the same school, the teaching assistant delivering Switch-on was observed not to be following the Switch-on protocol. The TA chose to conduct her session in the middle of the library in full view of everyone using it, and with no resources available or visible. When asked why she did this, the evaluators were told that she wanted to have the session 'in an open space'. The pupil appeared inhibited and conscious of other students walking around him. This TA also did not adhere to the Running Record protocol. She spent a lot of time talking about the book and the difficult words listed on the inside front cover page. Although she asked lots of questions to check for comprehension, the answers she received from the pupil were often monosyllabic. The teaching assistant did more talking than the student did reading, and the session lasted more than 30 minutes rather than 20 minutes set out in the guidelines.. The teaching assistant in this example was replaced immediately, perhaps an indication of how seriously the monitoring was taken, however this action is somewhat unrealistic since it is unlikely to happen if the intervention were to be rolled out more widely. In fact, such instances of unwillingness to stick to the protocol or of becoming confused between different programmes would become more common.

There was clearly variation in settings, the quality of the teaching, the adherence to the protocol, and the behaviour of children both within and between schools. Overall, the impression was that children enjoyed the sessions, and that those revisited had generally made progress between bands and levels of books.

Implementation

Conditions for successful implementation

The successful implementation of the programme hinges on the following factors:

- The quality of training and the ability of staff to use the protocols suggested.
- The commitment of the school: schools must be willing to redeploy teaching assistants and/or librarians to conduct the reading session and to provide the space for this.
- The motivation of teachers: teachers must be keen to try the programme and adhere to the protocol.
- Pupil attendance: pupils need to receive a minimum number of sessions for the impact to be observed.
- Adequate monitoring: close monitoring of teachers is necessary to ensure that teachers conduct the sessions as trained. (The intervention project team had a system to monitor and support schools. Each school was assigned a Switch-on trainer, themselves trained by the project leads, whose role was to monitor and support the schools in implementing the programme: they visited the schools regularly to provide extra support to the members of staff, and had regular online communication for feedback between these expert trainers and the project leaders. The enthusiasm of the project and teaching staff was a noticeable feature of this intervention.)
- The appropriate levels of books and the professional judgement of the teachers in deciding when pupils should move to the next level is also essential.

Challenges to successful implementation

Members of staff were generally positive about the programme, and many were enthusiastic and excited about the progress they had observed among their students. In most cases they reported that students generally enjoyed the one to one attention, something they would not otherwise have had. However, a number of challenges to successful implementation could be noted, including:

- Finding appropriate books for secondary school age children with a reading age of below 8 is a challenge. The stories and pictures in the books sometimes seemed more appropriate for children in Key Stage 1. Several teachers raised concerns that the stories were 'too babyish' in style and content (not reading age necessarily) for their Year 7 students.

- Scheduling the intervention within the school time-table. One of the main issues expressed by students was that they were missing lessons on a daily basis. Some seemed to have missed maths lessons more often than other lessons despite the Switch-on sessions being arranged so as to even out absence from different lessons. This resulted in a loss of learning which the students found difficult to cope with, potentially leading to unintended and damaging consequences for this intervention. In one school the sessions were not conducted at the same time every day (to try and avoid this problem), but this meant that students were expected to remember the different times for each day. As a result, several missed sessions, and some students simply dropped out as they could not remember the timings.

- Irregular attendance or non-attendance was another barrier to successful delivery. There were instances when students who were identified for intervention decided to take themselves out of the programme. One student did not attend all the sessions because he said his mother did not think he needed it as his primary school teacher had told her that he was doing very well in reading. This was despite his KS2 reading score being only level 3 and his first NGRT assessment (conducted by the school in September) registering his reading age at 8.5 (pen and paper test) and his NGRT test in December recording his reading age as 8. This boy, however, also said that he would have attended the reading sessions if his teacher had gone to collect him.

- Doubts about the validity of the running record as an assessment tool. One member of staff reported that one of his students was fixated with errors. He said that when he did the running records the student felt very nervous about making mistakes. The evaluators observed that the student was so aware of making mistakes that he seemed to make more mistakes. Another member of staff said:

'Success of 90% or even above on running records does not mean that these students can comprehend it completely. The tick sheets are irrelevant and do not show the true reflection of children's ability to comprehend.'

- Interference with other on-going literacy interventions. Some pupils were also on other reading intervention programmes (e.g. Accelerated Reader), so one cannot be certain that the improvement in the post-test was due to Switch-on or to the combined effect of Switch-on and Accelerated Reader. Analysis of the results, however, suggested no bias in take-up of other interventions between the treatment and control groups. In this design, other interventions are part of normal practice for both treatment and control. Nevertheless, there could be an interaction effect.

- Concerns regarding the validity of the pre-test results. Some members of staff raised concerns about this, reporting that some of their students were placed in a lower age band than they thought would be appropriate. Their observation was based on the fact that the students appeared to be reading above the colour banded books provided for the children for their reading age, perhaps skipping two or more colour bands in a short time. There are several possible reasons for this: perhaps the test is not accurate in assessing the children, or staff were noticing fluency in decoding whereas the test also included comprehension. Alternatively, the grading of the books may not have been as precise as was intended. So a book may appear unsuitable in topic or style for an eleven year old with a reading age of five, for example. On the other hand, there were members of staff who reported that although their students were moving up the coloured bands, they did not think that the reading was really that good. Perhaps it is the staff impressions that are incorrect.

Attractiveness of intervention to stakeholders

- Pupils liked the chance either to miss lessons or receive individual attention (perhaps both). One pupil was so concerned about missing a session due to a clash that he asked if he could do it at lunchtime and take the book home also. Another said that he felt very good when he realised that he had read so many books in the previous few weeks. For some students leaving the class for the reading session was a stigma, but for others it was a relief from the mundane activity of regular classes. A group interview with students highlighted similar issues.

- Teachers thought that students liked the individual attention as they rarely had this. This helped to boost their confidence as they felt special.

- One teaching assistant who was interviewed said that she found Switch-on training very helpful, especially in giving her the confidence to conduct individual reading sessions.

Outcomes

Perceived outcomes

Evaluators and teachers perceived that the intervention had a positive impact on reading progress. Teachers thought that students liked the individual attention as they rarely had this. This helped to boost their confidence as they felt special.

For example, the evaluators observed the pupil Samina on her thirteenth reading session: she had shown remarkable progress since the first session, moving on to lime books (the highest band), and according to the reading scores her reading age was 11.5 – very close to her real age. The TA thought Samina was ready to move to a new level, although there were no appropriate books available at that time.

The pupil Sarah also made some progress. She was on the white band of books and her reading age was estimated at between 7 and 10: she read fairly well with quite good expression, indicating that she understood the punctuation and sentence structure well, however she was still making mistakes and often looking to the member of staff for help. On a second visit, there were signs of more progress.

The pupil Mike seemed to enjoy the reading sessions because he considered them as a break from regular classes. He had thought the books he read during the first visit were babyish but by the second visit he was reading a book about the life and career of Wayne Rooney, a book that was appropriate for him as he was interested in football, and so something that he could identify with. He read very clearly and fluently. The member of staff asked comprehension questions and his answers were fairly accurate. The member of staff had noted that Mike was likely to benefit from the individualized attention. According to her, Mike had made significant improvements, and had become more confident.

Another pupil, Josh, with the same member of staff, was not as communicative as Mike and rather erratic in reading. He was new to the school and may have needed time to adjust. According to the member of staff, there were days when Josh was withdrawn or would miss the reading session and other days when he was very relaxed. Sometimes he could be defiant. At the observed session, he did not seem keen to engage with the member of staff or the books. The member of staff found it hard to establish a rapport with him, describing him as a closed box, not divulging much information about home or where he lived. Although he was good, his comprehension was not good. The member of staff noted that Josh had difficulty in registering any connections to make meaning of the text.

In School D, the members of staff reported similar improvements in reading for both of their students.

Another member of staff in School O reported that her pupils had made progress in reading, and now showed greater interest in reading. She also saw improvements in the pupils' confidence and trust. She claimed that there was now definitely a better relationship between staff and pupil.

One of the main issues expressed by students was that they were missing lessons on a daily basis. Some seemed to have missed maths lessons more often than other lessons despite the Switch-on sessions being arranged such that they happened at different times on different days to avoid children missing the same lesson. This was a loss of learning that the students found difficult to cope with, and could lead to unintended and damaging consequences for this intervention.

Formative findings

Since the evaluation was only of the intervention as proposed by the developer, what follows can only be suggestions based on the experiences of participants. Research evidence suggests that participants in an intervention, even where they are professionals, can be misled about what works and what does not.

The running records for assessment

Feedback from some teachers suggested that the running record sheets were not very helpful. They claimed that the running records were assessing students' gradual speed in reading but did not indicate anything much about decoding skills, and nothing about comprehension. One of the teachers suggested using the running record only once a week instead of every session

Range of books

A larger number of books for each band would have been appreciated, specifically with more developed for older pupils with low reading ages (i.e. suitable topics but easy to read).

Number of books per session

Teaching staff found it a rush to deal with four books in each 20-minute session, and many believed that the pupils would have coped better with two – a continued one and a new one – on each occasion.

Timings of the Switch-on sessions

The timings of the sessions could be arranged to avoid a clash with regular classes, or to coincide with a time when the rest of the class undertakes a communal reading intervention appropriate to their level. The students reported missing 20-30 minutes from their regular classes about four times per week.

Capacity building for teaching assistants and other staff

In terms of capacity building, Switch-on was seen as a very productive way of engaging teaching assistants in more professional roles, given appropriate training.

Control group activity

There was no evidence of post-allocation demoralisation, presumably because all schools, and all eligible pupils were 'treatment' pupils. The control group carried on with the usual routine of the schools, attending lessons as normal and continued with whatever intervention was in place. It was business as usual. The evaluators' concern was more with the unintended effects of missing lessons. Also if pupils said they liked to miss lessons because they were boring, this is even more worrying. Why are pupils not enjoying lessons? Is it because lessons are badly taught or is it because teachers are uninspiring?

Conclusion and implications

Limitations

Design

This was a simple and successful evaluation which functioned as intended. Pupils were taken out of classes for the intervention which followed a precise schedule and used forms and approaches not otherwise used in schools. Therefore, despite the within-school allocation of individuals to both groups there was minimal danger of treatment diffusion. Even if this had taken place to some extent, it would have had the effect of reducing the apparent effect size.

Long-term impact

The waiting list design, with all pupils having started the intervention by the second term, meant that there was no possibility of follow-up to see how far the impact eroded over time. However, as presented, the intervention is intended to be for the immediate transition from primary to secondary school and so this limitation is not a great concern.

The schools involved were volunteers within one local authority for an intervention run by a team who worked for that local authority; they were not intended to be representative of schools in any larger area. Only two schools dropped out, and this occurred before testing and randomisation. Six pupils in the remaining schools had either a pre- or a post-test score missing: in general, good reasons were given for each. The analysis has been run as a post-test only (with missing pre-test pupils included), and envisaged as though the missing six had extreme scores. In each case the substantive results are the same. The internal validity of the study is good. However, there were indications from the process evaluation that the intervention was mis-applied in some settings, even with close oversight and an accompanying evaluation. Therefore, problems could arise in trying to roll out this intervention to other areas and schools.

Testing

There were some issues regarding the use of online NGRT pre-tests. A number of schools had already used the pen and paper version of the test with their pupils in the autumn term as a matter of course: this could have had an effect on the performance of some pupils due to boredom or familiarity. Two schools reported that they had pupils who obtained quite different scores on what is essentially the same test, performing worse the second time round, however, the trial involved individually randomised pupils and so this is unlikely to have any systematic impact on the findings. According to the Chief Publisher at GL assessments, any anomaly could

> *'…well be an issue of testing too often (this may well have a negative impact on weaker readers) and there could be an effect of paper to digital. In any group of students you will see some fluctuation in scores, even those receiving an intervention, and a small number may appear to have gone backwards. I think this is just the nature of testing.'*

The pre- and post- tests were taken three months apart, whereas GL assessment advice is that re-testing is carried out at least six months later. This makes it more likely for a pupil to perform worse the second time round.

Members of staff in a number of schools also suspected that some students were simply guessing and rushing through. There was evidence from anecdotal reports of members of staff that some pupils used the test as a game to see who could finish it first. One pupil, for example, rushed to complete the test – clicking on what he thought was the answer without reading carefully to find the correct answer – because he did not want to be seen to be the last to complete it.

The tests proved a challenge to some pupils with learning difficulties such as problems with short attention span and Asperger's; there were also problems with visual impairment and speech impediment, and reports of students 'deliberately sabotaging the test', and 'messing around'.

Some schools reported that the test itself had been a barrier:

> *'The questions started off too difficult, so [students] lost heart and guessed a lot of answers.'* (School B)
>
> *'The NGRT test is far too difficult for students on a reading age below the pink books – (-5). Alternative tests need to be given to those with reading ages under 7.'* (School D)
>
> *'Due to the testing method – the test started too difficult and then adjusted accordingly to the answer – it should have started easier and then adjusted for the individual by becoming harder. Pupils not helped by the layout of the test – needed to scroll down to see questions and then back up to see text.'* (School H)

Several schools also reported technical problems with their computer systems which caused some confusion among pupils.

Interpretation

The overall finding, confirmed in several ways, is that the intervention as conducted was effective with these pupils, with an average effect size of +0.24. This is equivalent, in very approximate terms, to around three months extra improvement in reading age over the course of a year, at an estimated cost of £627 per pupil (or less than £2,240 per standard deviation of gain score). There is a suggestion from the raw score analysis that 0.24 may even be an underestimate of average progress, since a relatively large number of very low attainers had the same artificial threshold score for the age-standardised pre-test.

The intervention was as effective with boys as with girls, and appeared to be especially effective for low-attaining pupils and those with recognised special educational needs (although it must be noted that the quality of this indicator varied between schools). As a caution it should be noted that the intervention did not initially produce the same kind of gain scores for children living in poverty (a finding that has been noted with similar interventions before), the results were not linked to the precise number of sessions attended (greater than zero), and the impact appears to be relatively small once pupil background and especially prior attainment is accounted for. In fact, after prior attainment, the clearest determinant of the outcome score is the child's background as described by FSM, ethnicity, exact age, and first language. The data here provide no evidence regarding the active elements of the interventions, and no evidence on any unintended consequences or 'side-effects'.

This was a well-run intervention, generally supported by teaching staff and pupils. To achieve this level of co-ordination and support required considerable energy and monitoring. It is not clear that such a level of commitment could be maintained if the intervention was used more widely. Despite the generally supportive attitude, some teaching staff nevertheless made their own adjustments to the protocol or, in at least one case, ignored it almost entirely. The overall effect size of +0.24 is large enough to be taken seriously as a sign of success for the intervention – and the programme could be refined and developed for use with Year 7 pupils, partly as a result of the findings here – nevertheless, +0.24 is likely to be the upper limit of what is possible if the intervention were used more widely.

Future research and publications

The intervention was not fully developed at the outset, particularly for the better readers (the prior evaluations had been restricted to primary age pupils). So to some extent it developed as it went along, and the introduction of more advanced books for pupils during the intervention was greatly appreciated by members of staff. The two areas of greatest concern emerging from the process evaluation are the books themselves, and the timing/attendance schedule. These lead to two further research questions.

What is the 'active' ingredient of Switch-on? For example, does success depend on these precise books, on the reading record, on the length, number or frequency of the sessions? Does it depend on the rigid use of four books on each occasion? Or would almost any programme of one to one reading with a trusted member of staff be equally effective? Assuming that the overall effectiveness of Switch-on is accepted as promising, a multi-group trial could be designed to address such questions. In addition, a review of all one to one literacy evaluations would help identify the unique and common elements of success.

What are the possible 'side-effects' of Switch-on? For example, attending around 40 sessions during normal lesson times means that pupils have 40 lessons per term disrupted. The evaluation reported here only picked up the benefits of attending the sessions for reading, but there may also be harm done to progress in other areas of the curriculum, even though this may be 'scattered' among many curricular areas. Can this potential damage be measured? Is it possible for all children in a class to concurrently have a 20-minute session of a programme tailored to their needs (i.e. not individual attention for all), all at the same time? For some, this could be Switch-on. This could reduce any sense of shame (if it exists), encourage full attendance, and equalise any impact on the wider curriculum. An alternative would be to schedule one to one sessions outside of lesson time.

Such questions mean that there is more work to be done with Switch-on to make it more effective, efficient and as low-cost as possible, as well as minimising disruption to the life of a school.

In conclusion, it is worth repeating that the intervention was a success, and that it was partly conducted by teaching assistants (TAs). The future funding of TAs in England is unclear, and the evidence so far is that just having TAs, or using them as substitute teachers, is rather costly and largely ineffective. Switch-on is an example of one way in which TAs might be deployed in schools to follow a set protocol and make a useful difference to the reading of pupils in transition from primary to secondary education.

In the near future, the evaluators will prepare a summary paper describing the evaluation for a peer-reviewed journal.

References

Baenen, N., Bernhole, A., Dulaney, C. and Banks, K. (1997). Reading Recovery: Long-term progress after three cohorts, *Journal of Education for Student s Placed at Risk*, 2, 2, 161.

Centre, Y., Wheldall, K., Freeman, L., Outhred, L. and McNaught, M. (1995). An evaluation of Reading Recovery, *Reading Research Quarterly*, 30, 2. 240-263.

Coles, J. (2012). *An evaluation of the teaching assistant led Switch-on literacy intervention*, Unpublished MA thesis, University of London Institute of Education.

Gorard, S. (2013). *Research Design: Robust approaches for the social sciences*, London: SAGE.

Gorard, S. (2013). The propagation of errors in experimental data analysis: a comparison of pre- and post-test designs, *International Journal of Research and Method in Education*, http://dx.doi.org/10.1080/1743727X.2012.741117.

Hatcher, P., Hulme, C., Miles, J., Carroll, J., Hatcher, J., Gibbs, S., Smith, G., Bowyer-Crane, C. and Snowling, M. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised controlled trial, *Journal of Child Psychology and Psychiatry*, 47, 8, 820–827.

Iverson, S. and Tunmer, W. (1993). Phonological processing skills and the Reading Recovery program, *Journal of Educational Psychology*, 85, 1, 112-126.

Pinnell, G., DeFord, D. and Lyons, C. (1988). *Reading Recovery: Early intervention for at-risk first graders*, Educational Research Service Monograph, Arlington, VA: Educational Research Service.

Pinnell, G., Lyons, C., DeFord, D., Bryk, A. and Seltzer, M. (1994). Comparing instructional models for the literacy education of high risk first graders, *Reading Research Quarterly*, 29, 1, 8-39.

Schwartz, R. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention, *Journal of Educational Psychology*, 97, N2, 257-267.

Tanner, E., Brown, A., Day, N., Kotecha, M., Low, N., Morrell, G., Turczuk, O., Brown, V., Collingwood, A., Chowdry, H., Greaves, E., Harrison, C., Johnson, G. and Purdon, S. (2011). *Evaluation of Every Child a Reader*, London: NatCen.

Education
Endowment
Foundation