An Evaluation of the 'Switch-on Reading' Literacy Catch-up Programme

Stephen Gorard, Nadia Siddiqui and Beng Huat See
Durham University
s.a.c.gorard@durham.ac.uk

Abstract

This paper is based on one of the first completed studies funded by the Educational Endowment Foundation. EEF was set up in response to repeated demands for clearer evidence on school improvement. The paper presents the results of an intensive 10-week literacy intervention called Switch-on Reading. This was trialled in England as part of a government initiative to assist children below Level 4 literacy while at age 10 to catch-up with their peers on transfer to secondary school. Switch-on took place in 19 Nottinghamshire schools, with 314 Year 7 pupils individually randomised to treatment in the first or second term of school year 2012/13. This is the largest trial so far conducted of this kind of 10-week reading intervention. It was delivered on a one-to-one basis by trained school staff, mostly teaching assistants. The independent evaluation was based on pre and post administration of the New Group Reading Test, and on observations and interviews in schools. The overall result was an effect size of +0.24, based on the pooled standard deviation of the post-test score (and the gain score) for both groups, meaning that the programme made a noticeable positive impact. This effect can be envisaged as suggesting that on average a pupil receiving the intervention would make approximately three additional months' progress over the course of a year compared to similar pupils who did not, at a cost of around £600. 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). The trial also illustrates a key role for teaching assistants, and shows the feasibility of the EEF research programme.
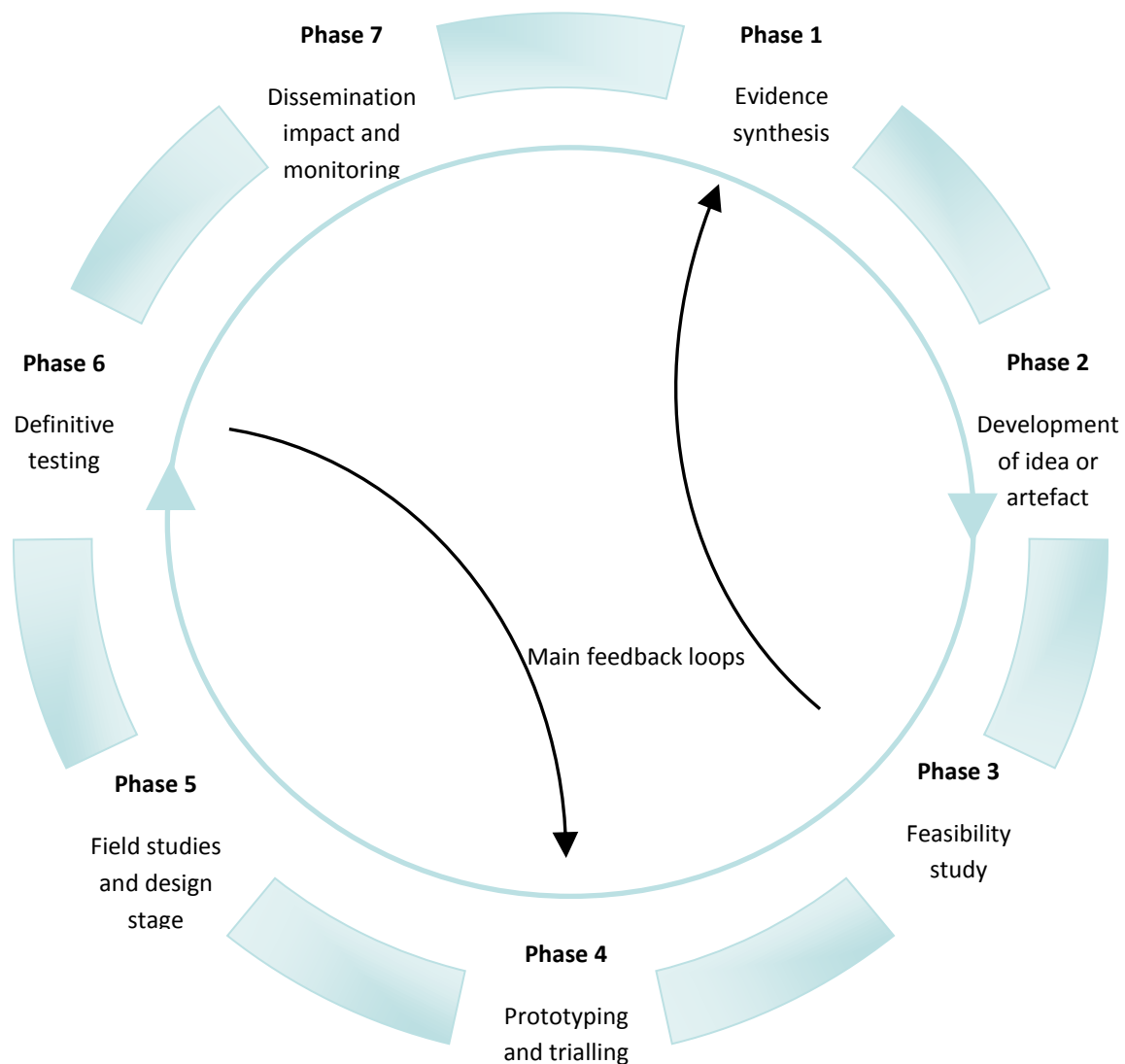
The need for robust evaluation

For several decades (Hillage et al. 1998, Tooley with Darby 1998), and probably much longer (Gorard 2004), UK education research has been criticised for not providing the kind of evidence base necessary to raise attainment, especially for disadvantaged pupils. Despite resistance by threatened education researchers, in fact these concerns were and are shared by a wide range of stakeholders including policy-makers, funders, and senior academics (McIntyre and McIntyre 2000, Taylor 2002). Given the level of taxpayer and charitable funding, education research was described as failing to deliver answers to even some of the most basic educational questions of interest to policy-makers and practitioners. Similar concerns about the lack of usable research evidence emerged in the US (NRC 1999, NERPP 2000). In the US, this led to the establishment of federal funding programmes like the Institute of Education Science, with its separate goal schemes from preparing an idea to national effectiveness trials and monitoring rollout of successful interventions. Because the capacity to conduct such work was limited in traditional academic schools of education, most of this substantial funding is now sought and gained by not-for-profit research consultancies.

In the UK, one of the first major responses was the, then, huge ESRC Teaching and Learning Research Programme (TLRP), set up to permit applicants to be funded at a level that the generation of safe evidence on improving learning was deemed to require. The Chair was a trials advocate with experience in health sciences, and the Steering Committee included experts in field trials, experimental psychology, and psychometrics. However, as in the US, the capacity to conduct the kind of work needed did not exist in most schools of education. Unlike the US, the schools of education effectively took over the TLRP anyway. Within a few years all of the above had left the programme. They were replaced by a historian as Chair, theorists and qualitative researchers on the Committee, and the new Director was another 'qualitative' researcher. No one in charge of TLRP had ever run a randomised controlled trial or used an alternatively rigorous evaluation design such as regression discontinuity. In that, they represented the majority of UK education research, and hence the very gap in skills that the programme was intended to address (Walford 2002, Schuller 2007).

Since the failure of TLRP to have the kind of impact that Sir Ian Chalmers and others had wanted, there have been several other promising starts (including the brief National Educational Research Forum). Now, the Department of Education is citing Ben Goldacre in setting up a list of research priorities and requesting evidence of the same form that McIntyre and McIntyre (2000) and so many others described 25 or more years ago (Wrigley 1976, Broadfoot 1985, National Science Foundation 2002, Goldacre 2013). The main problems, as ever, are the lack of ability among traditional education researchers to conduct such work, and their apparent unwillingness to learn or adapt. Of course, there has never been any suggestion that all research should be of this type – merely that this is where there is a key gap in the research cycle (Figure 1).

Figure 1 – An outline of the full cycle of education research

The diagram shows a circular process with seven phases arranged clockwise:

Phase 1 — Evidence synthesis

Phase 2 — Development of idea or artefact

Phase 3 — Feasibility study

Phase 4 — Prototyping and trialling

Phase 5 — Field studies and design stage

Phase 6 — Definitive testing

Phase 7 — Dissemination impact and monitoring

Main feedback loops

(Source: Gorard 2013)

In 2010, the new UK government set up the quasi-independent Educational Endowment Foundation (EEF), funded by a combination of DfE research money, charitable contributions and co-operation with other funders. EEF is intended to meet the long-term demand for robust evidence on school improvement. And despite its deliberate limitations in scope, in ignoring pre-school and later life learning for example, this is a welcome innovation. This new paper describes the outcomes from one of its first and still relatively small studies, before returning to the implications for UK education research as a whole.

Catch-up literacy programmes

Catch-up literacy projects are educational interventions intended for pupils struggling to reach what are officially deemed the age appropriate levels in reading. They are important because struggling pupils entering secondary school are more likely to remain behind, or fall further behind, their classmates, which can also lead to other issues such as disruptive

classroom behaviour (Gorard et al. 2011). In May 2012 the government in England made an extra £10 million available to the Educational Endowment Foundation (EEF) via the Department for Education, for a grants round dedicated to literacy catch-up projects for children at the primary-secondary transition. It was intended to benefit pupil premium children who enter the secondary school with below Level 4 in literacy (Gov.Uk 2012). Some reading interventions appear to be effective, at least for some struggling readers, but some do not (Hatcher et al. 2006, Cantrell et al. 2013). A prior review of existing evidence in this area for the EEF showed that one-to-one structured support was an area of promise (See and Gorard 2014).

One application for the EEF catch-up grants round was to conduct a trial of Switch-on Reading, as described in the next section. Switch-on is derived from a long-standing intervention called Reading Recovery (RR). This 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. What Works Clearinghouse (2013) found only four out of 78 evaluations of RR that met minimal evidence standards, and even these RCTs were rather small in scale. They involved 168, 91, 79 and 74 students respectively (Baenen et al. 1997, Pinnell et al. 1988, 1994 and Schwartz 2005). One other study of 64 students met WWC criteria with reservations because it was not a randomised controlled trial (Iverson and Tunmer 1993). Of these five studies, four reported positive effects for RR on first-grade general reading achievement, using the Observation Survey subtests for Dictation and Writing Vocabulary. Baenen et al. (1997) did not find positive effects using grade retention as an outcome measure.

In addition, 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 FSM and more SEN pupils. May et al. (2013), on the other hand, reported an effect size of +0.68 for RR with 866 randomly assigned low achieving first graders, and based on measurements using the Iowa Tests of Basic Skills.

Reading Recovery has therefore rarely been rigorously and successfully evaluated at scale, although several studies with weaker designs or rather small samples have claimed an impact on reading age. There has been even less evaluation of Switch-on Reading itself than of RR. 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 from this project that led to EEF funding to replicate the intervention with secondary school pupils and at a considerably larger scale. The existing evidence base was used to justify a rapid efficacy trial to test the impact of Switch-on with the developer leading the training and overseeing the delivery of the intervention, and the authors of this paper as an independent evaluation team.

The new evaluation described in this paper is different to those described so far. It is over 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). Switch-on is shorter in duration than the traditional Reading Recovery and, if found to have positive impacts, could represent a cost-effective way of rolling out an intervention like this. The evaluation was conducted by the

authors, appointed by EEF independently of the project team from Nottinghamshire LA who conducted the training and ran the intervention itself.

Intervention

The model of Switch-on Reading being evaluated was provided for Year 7 pupils in mainstream secondary school settings in Nottinghamshire. The intervention is a short-term individual 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. The purpose was for as many pupils as possible to achieve functional literacy, and so to close the reading achievement gap for vulnerable children working below the age-expected levels.

The intervention was conducted by staff including SENCOs, librarians, teachers, and teaching assistants who were the clear majority. Each member of staff was trained, and looked after no more than four pupils. Each pupil was given a schedule in which to come out of one standard 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.

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. 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 was just challenging enough. The books themselves included fiction and non-fiction with lots of visual images meant to encourage students' interest in reading 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)
 • Discussion on the material, visuals, cover pages and blurbs of the books
 • Invoking interest of students by involving them in talking about visual content
 • Reading of the text and using the running record sheet for analysis of reading
 • Feedback to the student
 • Introduction to a new book

Therefore, each session incorporated revision of a familiar text, introducing new vocabulary, practicing phonics and also improving comprehension through questions and talking about the texts. In each session the student should read excerpts of text from four books.

At some point in the 20-minute reading session the member of staff recorded the reading assessment of the pupil on a sheet, and makes an inventory of errors such as words missed, substituted with another, mispronounced, repeated, plus self-corrections and appeals for help. The form for recording these events and the rules for completion were standardised, and an integral part of the intervention. Part of the intervention also involved analysis of errors. The average number of errors was calculated, and determined which book set was followed next. After each book, the adult trainer praised the child when an effective reading strategy was

observed, and prompted the student to use new strategies where behaviour had not been effective or advice had been ignored.

Evaluation Methods

Trial Design

The evaluation was based on a relatively simple one-term waiting-list design (Gorard 2013a). 19 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, and half to intervention after one-term. The latter group formed the control. 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, and the post-test was conducted before Phase 2 pupils received the intervention.

This design is ethical since all schools received the intervention, 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.

Outcomes

At the outset, all eligible pupils took the GL New Group Reading Test A. Because this pre-test took place before randomisation to the two groups, the process was 'blind' as to the treatment group.

Once the intervention was complete, both groups took the GL New Group Reading Test B, 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 observed the tests in operation. Both the pre- and post-tests were conducted on-line to encourage standard format and timing, to reduce the potential influence of staff, and to create instant results for the schools and evaluators.

Participants

All local authority secondary schools in Nottinghamshire were eligible to take part. The school recruitment process was conducted by the project lead members who work for the Local Authority. The networking process for the recruitment involved meetings with the head teachers and school literacy coordinators. Two schools that initially agreed to participate dropped out before the pre-test was conducted. The reason cited was that they were reported as not being prepared to adjust their timetable to accommodate the regular 10-week reading sessions.

The eventual 19 schools which agreed to take part signed an agreement that being part of the programme entailed agreement to the evaluation. A consent form was sent to parents of

pupils eligible for the programme. Around half of the schools were in the ex-mining areas of Nottinghamshire, ranging in size from around 600 pupils to over 1,500. FSM eligibility ranged from 6% to 30%, pupils not speaking English as their first language ranged from 1% to 10%, and pupils with statements of special educational need or receiving School Action Plus ranged from 3% to 7%.

The students to be a part of the trial were identified and selected by the individual school teachers on the basis of weak performance in reading (below secure level 4 in KS2 for English). The process of random allocation to treatment and control groups was conducted after the pre-test, by the lead evaluator in the presence of another researcher. 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 back to the respective schools. The order of the cards was retained for a short time in case of queries. In total, 314 individual Year 7 pupils took part in the study. 157 pupils were assigned to treatment and 157 to control. Their characteristics are shown in Table 1. Overall, the two groups were reasonably well-balanced in terms of these background characteristics.

Table 1 – Background characteristics of the pupil sample, percentage of pupils in each group

	Treatment (%)	Control (%)
Male	60	56
FSM	30	34
SEN (statement or school action)	74	72
EAL	4	4
Non-White UK	10	8

By the final analysis six students had missing scores for various reasons. One took the pre-test (repeatedly) but his school were unable to record the score. Five others took the pre-test but did not sit the post-test (Table 2). 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 2 - Pupils allocated to groups but with no gain score, and reason for omission

Allocation of pupil	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

All pupils were analysed in terms of their initial allocation (intention-to-treat) as far as the post-test data permitted.

Analysis

The powerful design of the study means that the analysis is simple (Author Gorard 2013). 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 cause 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, 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.

Process evaluation

The Switch-on project leaders conducted the training of staff, 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. Fieldwork in schools included observations of staff delivering the intervention, noting inconsistencies or any departures from the programme protocol. On each visit, the evaluators also interviewed the school leads, relevant teaching staff, and spoke with small groups of participating pupils. The evaluators considered the resources used (such as story books, running records and pupil progress charts), and asked for staff and pupils' perceptions of these materials. Interviews were usually conducted without a formal schedule, and arose as the situation allowed. The interviews and field notes were part-transcribed and shared between the evaluation team. Schools agreed to be part of the evaluation when agreeing to be part of the intervention.

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 also useful in identifying potential issues or barriers which could be addressed for any future scaling-up.

Summative evaluation results

Overall

The effect size of the intervention was +0.24 standard deviations of the overall gain score, showing a noticeable positive impact (Table 3). 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 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 (Gorard 2013a).

Table 3- Estimated impact of Switch-on Reading Programme – overall with gain score

Treatment group	N	NGRTA	NGRTB	Gain	Standard deviation	'Effect' size

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

Table 4 shows the same results assessed as though for a post-test only design. This confirms the 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 standard pre- and post-test design (Gorard 2013b). It also allows the inclusion of any cases with post-test scores but missing pre-test data. The similarity of results to Table 3 is reassuring about the initial balance between the groups.

Table 4 - 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	154	78.73	9.29	-
Overall	308	79.84	9.33	+0.24

Sub-groups

Children with special needs, boys, lower attainers, and poorer pupils tend to do worse than average in reading and so were more likely to be selected by their schools for either phase of Switch-on. All of these groups are therefore over-represented in the cases eligible for Switch-on. 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 like the other sub-groups these were almost evenly divided between treatment and control.

If the cases are separated into sub-groups of only those pupils sharing a particular characteristic, all have a positive 'effect' size (Table 5). Three of the sub-groups actually have a higher effect size than overall, and this is so whether the gain scores or the post-test only scores are used. This suggests that Switch-on is effective for the lowest attainers (those below the median pre-test score of 73), even among those selected for the intervention. The effect size for FSM pupils, however, is considerably lower and almost indistinguishable from zero based on the standardised ages scores.

Table 5 - Estimated impact of Switch-on Reading Programme, using gain scores for sub-groups

Sub- group	N	'Effect' size
FSM eligible	98	+0.05

Sub- group	N	'Effect' size
Boys	179	+0.26
SEN (statement or School Action)	225	+0.33
Lower attainers	156	+0.44

One possible explanation for the anomalous FSM result comes from consideration of the relationship between pre- and post-test scores. Of the overall 308 pupils in the trial, 94 achieved a pre-test score of 69 (the lowest standardised age 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. There are pupils with a pre-test score of 69 with post-test scores ranging from 69 to 105. This suggests 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.

In order to assess any limitation created by the lower threshold score of 69 in the pre- and post-tests, the analysis for FSM-eligible pupils was also re-run using the raw scores (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 (Table 6). It suggests that a floor effect in the age standardised scores did lead to an underestimate of the gains for FSM-eligible children.

Table 67 - 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

The overall analysis was also re-run using the raw post-test scores (before age standardisation). This produced an effect size of +0.22, which is about the same as the overall result using the standardised scores (Table 78). It suggests that the overall impact of the 69 threshold is low (although it may still matter for analysis of specific groups, such as FSM pupils). There is a general warning here that analysts need to check for 'floor' effects, especially when dealing with standardised test scores.

Table 78 - Estimated impact of Switch-on Reading Programme – Post-test only raw scores

Treatment group	N	NGRTB raw score	Standard deviation	'Effect' size

Treatment group	N	NGRTB raw score	Standard deviation	'Effect' size
Switch-on	155	260.65	54.77	-
Control	153	248.18	58.99	-
Overall	308	254.45	57.16	+0.22

The intervention in action

The training events

The intervention began with a two-day training event for 83 staff from 19 schools, of which 71 went on to implement the intervention. The trained adults included teaching assistants (TAs) SENCOs (Special Educational Needs Officers), librarians, literacy teachers and other members of the teaching staff. Training involved project leads demonstrating 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. Switch-on resources, including specially selected colour-banded books, calculators, clocks and running record sheets, were distributed to schools. Eight trainers were specially trained by the project leaders to support the schools. Their job was to visit schools and support the members of staff who delivered the intervention. This was to help ensure fidelity to treatment.

The project leads were clear in communicating the objectives, rationale and process of the intervention method. The process and procedures were explicitly demonstrated and the trainees were given enough time and chance to ask questions and practice the protocol such as recording and analysing errors, calculating the average reading scores and making decisions about the levels of books for their pupils. There was a follow-up session to review the progress of the schools involved in the intervention.

Observations of Switch-on lessons

In general, the members of staff conducted the sessions as they were trained to. The main elements of the trainings that were most often observed in the reading sessions were:

- Using a variety of text books in a 20 minutes session
- Talk about text
- Comprehension questions by teaching staff
- Independent reading aloud of text (at least 100 words) by students

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. 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. Another member of staff in the same school 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. 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. In fact, she did more talking than the student did reading. The session lasted more than 30 minutes rather than 20 minutes. It was clear she failed to use the clock provided. The pupil appeared bored and was using his finger to point at every word in the book. The book was also placed at an angle instead of being directly in front of him. When asked why she was spending so much time talking about the book, the TA explained in front of the student that he was rather weak. Overall, this TA was not implementing Switch-on as it was intended.

In another school, it was noted that staff completed the attendance sheets and running records, and these were carefully filed. The evaluators observed most pupils having made considerable progress both in terms of the band of books and their reported reading age. In fact, one pupil had no more appropriate books at the highest level (at that time). One pupil seemed to have enjoyed the reading sessions because he considered them 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. This was appropriate for him as his interest was in football. He read clearly and fluently. The member of staff stated that he had made significant improvements, and become more confident. Another pupil was enjoying it so much that he requested the member of staff to let him take the books home to read them with his mum.

One pupil had a recorded reading age of five, and was on the yellow band of books (for reading age 5-6). His reading was slow and he made errors such as omissions and substitutions. He was regular in attendance and showed interest in the visual contents of the books. He especially enjoyed discussions about simple things in the text. Although reading the text orally was a little challenging for him, he was very keen to talk about the stories, and draw comparisons with other people and real life. He said that he preferred coming to the reading sessions as he enjoyed reading with 'Miss', and the other classes were boring for him.

One pupil was not communicative and rather erratic in reading. He was new to the school and may have needed time to adjust. Sometimes he could be defiant. 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. One other pupil was resistant in going to the reading sessions and had missed some. Other schools showed similar variation in motivation and improvement.

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 considerable progress between bands and levels of books.

The views of members of staff

Members of staff were generally positive about the programme. 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, as something they would not otherwise have had. A number felt that the Switch-on sessions gave them the opportunity to get to know more about the students, their attitudes and family background. In some cases

the sessions had helped develop a mutual relationship between the staff and students which, according to the staff, improved students' confidence and trust more widely. One member of staff commented that in the ideal world they would like to make this intervention permanent for their students.

However, there were aspects about the programme that some members of staff would like to change. Several felt that although Switch-on allowed them the flexibility to adapt the intervention according to the interest and levels of the students, the requirement of the intervention to use four different books in one brief reading session was distracting to the students. Students often did not get to finish the rest of the story, and so could not engage with the story fully. One member of staff had a pupil with autism, and reported that she specifically resisted changing the texts before completion with that child.

Several teachers raised concerns that the stories were 'too babyish' in style and content (not reading age necessarily) for their Year 7 students. This criticism of the books was well taken by the project team members and they responded that only students with low reading age (Level 3 and below) had been selected for this intervention. It was assumed that introducing this kind of material to them in individualised settings would probably not adversely affect the confidence of the students. In the training it was clearly explained that the Switch-on reading sessions would not be discussed generally in the classrooms. Most members of staff felt the second batch of Switch-on books were more appropriate for the age group of the students in terms of topics and level of difficulty. Children appeared to take more interest in the newer books.

Some doubts were expressed about the validity of the running record as an assessment tool, since it does not record comprehension. One member of staff said that decoding was not the problem. Another reported that one of his students was fixated with errors; because of the running records the student felt very nervous about making mistakes.

The views of pupils

The main challenge to implementing the intervention was the constraint of the school time-table. One of the issues raised 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 it happened at different times on different days to avoid children missing the same lesson. This could lead to an unintended a loss of learning. 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. For example, one boy told the evaluators he decided not to bother attending because he could never remember when to go for the session. Another girl explained that the sessions often clashed with her favourite subject and that this was the reason why she decided not to attend.

A few intervention students who had to leave their classes on a regular basis reported that they felt conscious as their peers knew that they were the ones singled out for the intervention. For this reason, such sessions might have been better conducted during break times or after school, or when entire classes were broken up for different activities. On the other hand, and as shown in the section on staff views, most pupils liked the chance either to miss lessons or receive individual attention. One pupil was so concerned about missing a

Switch-on session due to a clash he asked if he could do it at lunchtime and take the book home also.

In two schools observed, there was general apathy among the students. They did not seem particularly excited about the programme. When asked if they would like to continue the sessions the following term, many said they would not. Interestingly, almost all of these pupils said their favourite subjects were PE, dancing, drama, Art or ICT (DT), subjects that did not require much reading. It was also observed that many of these students did little reading on their own outside school. When asked how they read instructions and notices in public places like bus stops or train stations, almost all said they relied on older siblings and parents for help. The only kind of reading they did outside school was magazines such as 'Hello' magazine for girls and sports pages from the newspapers for the boys.

Observation of the testing

This evaluation differed from many of those outlined at the start in using a generic standardised test of outcomes, agreed with the project team, but which was not devised by that team or explicitly taught in the programme. It was administered on-line, marked independently and date-stamped. At the pre-test no one knew which pupils would be randomised to which groups. At the post-test, the evaluators sent observers to a sample of schools. The evaluators have reasonable confidence in the quality of the testing procedures (and of the results in raw-score terms).

Some members of staff raised concerns regarding the validity of the pre-test results. They reported that some of their students were placed on 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 the staff were noticing fluency in decoding whereas the test also included comprehension, or the grading of the books was not as precise as was intended. 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.

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.

Members of staff in a number of schools also suspected some students were simply guessing and rushing through. One pupil 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 the test. Most schools reported similar attitudes from among some of their pupils. The tests proved a challenge to some pupils with learning difficulties. In one school a pupil with hearing difficulties struggled to hear through the head phones because she just had a new hearing aid fitted, and so became agitated. There were also reports of students 'deliberately sabotaging the test', and 'messing around'. Six schools also reported

technical problems with their computer systems, and this caused a bit of confusion among pupils.

Despite these problems, there is no reason to assume that they caused any systematic bias for the impact evaluation, since there were an equal number of treatment and control pupils in each school.

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 interventions were already in place. It was business as usual.

Conclusions

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. This was an efficacy trial, set up rapidly in response to a political timetable, to test the impact of Switch-on as delivered with the developer leading the training and overseeing the provision of the intervention. Efficacy trials test evaluations in the best possible conditions to see if they hold promise, but do not demonstrate that the findings hold at scale in all types of schools. The findings do not necessarily indicate the extent to which the intervention will be effective in all schools since the participating schools were selected purposively within one local authority, and training was provided by the programme developers. The intervention was generally well-conducted and most pupils seemed very happy with their reading sessions. Staff needed training and then some monitoring to ensure that they adhere to the protocol in order that the intervention has the largest possible effect. There were indications 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. However, this also suggests that the estimated effect size is realistic and not inflated by the artificial situation of an evaluation.

The overall finding, confirmed in several ways, is that the intervention as conducted was effective with these pupils, with an effect size of +0.24. This is equivalent, in very approximate terms, to around three months extra improvement in reading-age over three months, at an estimated cost of £627 per pupil (for a school to set it up, including staff costs and books). The intervention was as effective with boys as girls, and was especially effective for pupils with recognised special educational needs (although it must be noted that the quality of this indicator varied between schools), and lower attainers. The intervention was effective for FSM-eligible pupils, based on raw-score outcomes.

The intervention was largely conducted by teaching assistants (TAs). The future funding of TAs in England is unclear, and the evidence so far had been that just having TAs or using them as substitute teachers is rather costly and largely ineffective (Blatchford et al. 2012). 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.

The data provide no evidence on what the active elements of the interventions are, and no evidence on any unintended consequences or 'side-effects'. For example, does it 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 process 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.

Also 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 have 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.

Such questions mean that there is more work to be done with Switch-on to make it more effective, as efficient and low-cost as possible, and presenting the least disruption to the life of a school. Meantime the results can be added to a growing synthesis of evidence of what works, such as that represented by the Pupil Premium Toolkit (EEF 2014). Although relatively small compared to future plans, this trial shows again that RCTs are feasible and useful, and that the EEF approach of filling in the existing gaps in Phases 6 and 7 of the research cycle (Figure 1) is possible. The evaluation itself was inexpensive (around £30k), since the main cost was that of the intervention. The intervention was to happen anyway, as so many interventions do every year, and the phasing-in was needed to ensure individual attention. Therefore, the RCT simply 'piggy-backed' on the kind of activity that happens regularly in schools anyway. It generated no specific ethical or practical difficulties of the kind that threatened researchers claim are intrinsic to rigorous evaluations. This work therefore forms part of the belated response to McIntyre and McIntyre (2000) and others.

Acknowledgements

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

Blatchford, P., Webster, R. and Russell, A. (2012) *Challenging the role and deployment of teaching assistants in mainstream schools*, Report to the Esmee Fairbairn Foundation

Broadfoot, B. (1985) The impact of educational research, *Research Intelligence* May 1985, p.10-11

Brown, S. (2002) *TLRP: The significance of the Programme*, paper presented to TLRP Annual Conference, Cambridge, September 2002

Cantrell, S., Almasi, J., Rintamaa, M., Carter, J., Pennington, J. and Buckman, D. (2013) The impact of supplemental instruction on low-achieving adolescents' reading engagement, The Journal of Educational Research, DOI: 10.1080/00220671.2012.753859

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

EEF (2014) Toolkit, http://educationendowmentfoundation.org.uk/toolkit/, Accessed 18/3/14

Goldacre, B. (2013) Building evidence into education, DfE, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/193913/Building_evidence_into_education.pdf, accessed 23/5/14

Gorard, S. (2004) The British Educational Research Association and the future of educational research, *Educational Studies*, 30, 1, 65-76

Gorard, S. (2005b) Current contexts for research in educational leadership and management, *Educational Management Administration and Leadership*, 33, 2, 155-164

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

Gorard, S. (2013b) 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

Gorard, S., See, BH and Davies, P. (2011) *Do attitudes and aspirations matter in education?: A review of the research evidence*, Saarbrucken: Lambert Academic Publishing

Gov.UK (2012) £10 million to boost literacy for year sevens, https://www.gov.uk/government/news/10-million-to-boost-literacy-for-year-sevens, Accessed 18/3/14

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

May, H., Gray, A., Gillespie, J., Sirinides, P., Sam, C., Goldsworthy, H., Armijo, M. and Tognatta, N. (2013) *Evaluation of the i3 scale-up of Reading Recovery*, University of Delaware

McIntyre, D. and McIntyre, A. (2000) *Capacity for research into teaching and learning*, Swindon: Report to the ESRC Teaching and Learning Research Programme

National Educational Research Policy and Priorities Board (2000) *Second policy statement with recommendations on research in education*, Washington DC: NERPP

National Research Council (1999) *Improving student learning: a strategic plan for educational research and its utilization*, Washington DC: National Academy Press

National Science Foundation (2002) *Scientific research in education*, Washington DC: National Academy Press

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

Schuller, T. (2007) Capacity building in educational research: sketching an international picture, http://www.scotedreview.org.uk/pdf/198.pdf, accessed 23/5/14

Schwartz, R. (2005) Literacy learning of at-risk first-grade students in the Reading Recovery early intervention, *Journal of Educational Psychology*, 97, N2, 257-26

~~See, BH and Author, S. (2014) Improving literacy in the transition period: a review of the existing evidence on what works, *British Journal of Education, Society and Behavioural Sciences*, 4, 6, 739-754, http://www.sciencedomain.org/issue.php?iid=431&id=21~~

See, BH and Gorard, S. (2014) Improving literacy in the transition period: a review of the existing evidence on what works, *British Journal of Education, Society and Behavioural Sciences*, 4, 6, 739-754, http://www.sciencedomain.org/issue.php?iid=431&id=21

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

Taylor, C. (2002) The RCBN Consultation Exercise: Stakeholder Report, *Occasional Paper 50, Cardiff University School of Social Sciences*

Tooley, J. with Darby, D. (1998) *Educational research: a critique*, London: OFSTED

Walford, G. (2002) Editorial, *British Journal of Educational Studies*, 50, 4, 415-418

What Works Clearinghouse (2013) Reading Recovery, http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=420

Wrigley, J. (1976) Pitfalls in educational research, *Research Intelligence* Autumn 1976 Vol. 2, No. 2, p.2-4