Best practice in conducting RCTs: Lessons learnt from an independent evaluation of the Response-to-Intervention programme

Authors
Beng Huat See, Stephen Gorard and Nadia Siddiqui
School of Education, Durham University
Leazes Road
Durham
DH1 1TA
b.h.see@durham.ac.uk

Abstract

This paper presents the findings of the first independent UK evaluation of a large-scale randomised controlled trial of Response-to-Intervention, used as a catch-up literacy intervention. A total of 385 pupils in their final year of primary school (Year 6) were involved in the study (181 treatment and 204 control). These were identified as those at-risk of not achieving the threshold Level 4 in English at Key Stage 2. The pupils came from 49 schools across England. Twenty-seven schools were randomised to receive treatment immediately and 22 schools, which formed the control, were randomised to receive the intervention later. RTI was delivered in the summer term in preparation for pupil transfer to secondary school. The overall impact based on the standardised New Group Reading Test (NGRT) showed an 'effect' size of +0.19, and of +0.48 when considering only free school meal eligible pupils. However, these results must be viewed with considerable caution given the high attrition (over 25%) especially from the control group, and unclear identification of pupils eligible for the intervention. The fact that the evaluators did not have direct contact with schools when trying to identify eligible pupils, and that the developers wanted to use the pre-test to identify eligible pupils, led to this being a weak trial. The significance of the work therefore lies at least as much in the lessons learnt as in the headline figures. We learnt that ideally no more than two parties should be involved in communicating with schools, so that relevant instructions are passed quickly and acted upon promptly. This helps minimise the risk of misunderstanding and dropout post-allocation. Prior training on the technicalities of trials and research in general is necessary for both developers and any staff delivering the intervention so that all parties involved understand their commitment and the need to provide accurate and complete data. In future trials of RTI, it would be better for individual eligible pupils to be randomised rather than schools. RTI should ideally be given a whole year to allow the full cycle to be implemented, and it should be delivered daily for at least 30 minutes.

Introduction

This paper describes the findings of the first large-scale UK randomised controlled trial of a widely popular multi-tiered intervention known as Response-to-Intervention (RTI). The aim was to assess the impact of RTI as a literacy catch-up intervention. As this was an efficacy trial, a secondary aim was to assess the fidelity of implementation to identify barriers and so that lessons can be learnt which will inform future trials. The first part of the paper outlines the existing evidence on RTI and the rationale for this new study. This is followed by a discussion of the methods. The paper then presents the findings from the impact evaluation in terms of pupil test scores and the process evaluation. It finishes with a

discussion of the lessons learnt and makes recommendations for future implementation of RTI and the conduct of trials more generally.

Background

There have been increasing concerns that some children in England are moving from primary to secondary school without achieving the expected threshold level of literacy. Previous evidence suggests that these children are not likely to catch up and are more likely to continue to fall further behind their peers (Sainsbury et al. 1998, Galton et al. 1999). Since the 1970s, there has been considerable research looking into bridging this gap during the transition. And policies such as the National Curriculum and the National Literacy Strategy were introduced to try and ensure continuity of curriculum between the two phases.

Despite these efforts, the gap persists especially for those pupils from disadvantaged backgrounds. In May 2012, the Department for Education (UK) awarded the Education Endowment Foundation (EEF) a grant of £10 million to fund literacy catch-up projects. Catch-up literacy projects are a set of educational interventions intended for pupils struggling to reach what are officially deemed the age-appropriate levels in reading (and perhaps numeracy). They were intended to benefit pupil premium children (predominantly children who are eligible for free school meals) who would otherwise enter secondary school below Level 4 in literacy at Key Stage 2 (Gov.UK 2012).

A review of such interventions addressing catch-up literacy was commissioned by the EEF in 2011 (See and Gorard 2014). One of the more promising approaches identified in the review, and subsequently funded as an intervention by the EEF was Response-to-Intervention (RTI). Response-to-Intervention is a multi-tiered approach that involves initial screening to identify students' learning needs using research-based instructions with on-going monitoring of progress and with different levels of intensity (or tiers) to meet pupils' learning needs. It is a personalised and targeted intervention developed in the United States. The theoretical and empirical framework of the approach was based on work by Clay (1991) and Fountas and Pinnell (1996, 2006). According to Clay children learn literacy skills by developing an inner control of strategies for processing text. If a piece of text is too difficult, the child cannot develop this control. So any text used should be pitched at the right level. With effective and explicit teaching, the teacher can help the child build a strategy to enable them to process the text. Based on their work on Reading Recovery, Fountas and Pinnell developed an approach called Guided Reading using books matched to children's abilities employing differentiated instruction in small groups, gradually building up the child's inner control. This was the basis for the differentiated levels or tiers that formed the basis of the RTI approach.

Early evidence from the US suggested that this approach is effective with pupils in the transition period, defined in England as the stage when pupils move from Year 6 (final year of primary school) to Year 7 (first year of secondary school). One fairly large RCT in the US (Vaughn and Fletcher 2012), using RTI as a remedial intervention, reported positive effects on decoding, fluency and reading comprehension (d = +0.16) for those receiving both Tier 1 (whole class) and Tier 2 (small group) of the intervention compared to those who received only Tier 1. The study involved 784 pupils in 6th, 7th and 8th grades (English Years 7, 8 and 9). Another randomised study by some of the same authors, but with only 30 pupils in Grades 6 to 8, suggested that the intervention had a positive effect for pupils with severe

reading difficulties, although the gains were not big enough to close the gap with typically performing pupils (Leroux et al. 2011). The intervention was administered to treatment students every day in a 45 to 50 minute period.

A study by Graves et al. (2011) showed that RTI was particularly efficacious for pupils from disadvantaged backgrounds with learning difficulties, and was more effective for improving oral fluency than reading comprehension. All of the pupils involved were 'below' or 'far below' basic level in literacy. This was a quasi-experimental study that compared small group intensive reading instruction (Tier 2) with a control group ('business as usual') involving 6th Graders with and without learning disabilities. The duration was 30 hours over 10 weeks.

In another study, by Faggella-Luby and Wardwell (2011), positive results were reported for the small group (Tier 2) intervention for older children (Grade 6), but not for the younger ones (Grade 5). As the sample was quite small (n=86), it is possible that the result could be a reflection of teacher effectiveness or level of maturity of the pupils rather than due to the intervention. Study participants were identified as disadvantaged children in the 5th and 6th grades in an urban middle school. Students were randomly assigned to one of the three instruction practices: (a) experimental (Story Structure to improve reading comprehension), (b) comparison (Typical Practice) and (c) control (Sustained Silent Reading). Each session was 30 minutes and was administered 2-3 days per week for 18 weeks.

The evidence on RTI so far is predominantly from the US, and has mostly involved small samples or focused on those with learning disabilities. It was not yet clear whether RTI was suitable with 10 and 11 year olds struggling with literacy in the UK. In the US despite the range of resources, books, materials and training programmes available, there was little guidance on how RTI could be implemented within the framework of a regular classroom. In the UK, teaching materials have not been generally available until now, and there have hitherto not been any large-scale RCTs that test the programme in classroom conditions. There is therefore a case for an efficacy trial in the UK.

Study aims

The aim of this new study was to determine the effect of RTI on the literacy skills of children identified as not achieving the expected Level 4 at KS2 in England.

Since this was the first large-scale randomised-controlled trial in the UK, and previous implementations of the intervention had been patchy with underdeveloped resources, a secondary aim was to assess the fidelity of treatment, and to identify the challenges teachers may face in delivering the intervention in real classroom conditions. It is hoped that this will provide guidance for future applications of RTI in the UK.

The intervention

The RTI programme in this study was designed by the Centre for the Use of Research and Evidence in Education (CUREE) who developed the specialist tools and resources, and delivered the training. Training was conducted prior to the implementation of the programme and after schools had been

randomised. The training was a 3-day event which included an introduction to the concept of RTI, and the range of tools and protocols. Teachers were shown how to use these in screening pupils for eligibility and assessing their needs, and how to select appropriate research-based approaches. In addition, treatment teachers also received on-going support provided by another organization known as AfA3As (Achievement for All 3As) through in-school coaching using their Achievement Coaches as part of the AfA programme. Teachers in the control schools, on the other hand, did not receive any special training. They continued teaching as normal, including any interventions that they might have already been undertaking.

The programme description

The RTI programme used in this trial follows a sequence of stages. It begins with an initial screening to identify the individual needs of the pupils using the Close-Case Analysis Tool. The tool helps staff to determine the literacy areas to focus on for each pupil, such as phonics, fluency or comprehension. It also helps to determine the degree of intensity for the intervention - whether it would be Tier 1 (whole class) or Tier 2 (small group) intervention. The more intensive (Tier 3) is a one-to-one or pair work intervention. This is only recommended if the number of children with a particular issue in an area of literacy is small or if the pupils involved have already had interventions targeting this area, and the teachers do not think they are likely to make progress in a group setting.

Teachers then decide on the appropriate approaches to use on the targeted pupils using a set of 'Making Choices' information sheets. On these sheets is a menu of interventions or approaches, all of which are purportedly based on empirical research. For each intervention on the menu the appropriate levels of intensity or tiers is suggested. On the menu are strategies for promoting reading comprehension (for example, Peer-Assisted Learning Strategies or PALS), and for teaching paraphrasing and inference. There are also interventions for spelling, grammar, fluency, vocabulary and phonics.

For each aspect of literacy there is a Tracking Pupil Progress Tool to guide teachers in tracking pupils' progress and to assess how they are responding to the intervention, such as how well a child is doing in fluency, accuracy and speed. The tracking tool is in the form of a table where the teacher plots out the key features of the genre being taught. Then using a tick grid, the teacher notes whether each pupil on the programme has included or used each of these features. The teacher may develop a grading system to note how well or to what extent they have included the feature. Pupils' progress is monitored once in mid-intervention and once at the end of the intervention, based on teacher assessments and judgements. The mid-intervention monitoring is to enable teachers to decide whether to increase the intensity from small group to one-to-one for particular pupils.

The trial was conducted in the last term of the school year. The number of sessions and the length of each session varied from school to school. More details are given in the process evaluation (below).

Impact evaluation methods

This was a randomised controlled trial with a waiting-list design (Gorard 2013), of the RTI approach, developed by CUREE as a catch-up literacy intervention. The aim was to raise the literacy skills of Year

6 pupils (aged 10-11) who, at the point of transition to secondary school, were not expected to achieve the threshold level (or Level 4) at KS2.

Sample

Initially 91 schools were approached through the AfA3As network of schools. Of these 85 indicated interest, but 24 subsequently declined to participate when they realised what was expected of them, leaving only 61 schools. Traditional 'power' calculations make a number of assumptions that are not appropriate here (see below). However, using Lehr's approximation for an 80% chance of detecting a presumed effect size of 0.2 with 5% alpha, the minimum sample size needed per arm would be 400 individual cases (Gorard 2013). Traditionally, this would be higher because schools were randomised, not individuals, and then lowered again due to the high correlation between the pre- and post-test scores for individuals. And, of course, the effect size could have been higher than 0.2. Nevertheless, this trial would have had relatively low 'power' even if the intended sample had been achieved. Given that only 61 schools were recruited eventually, this reduced the scale of the trial considerably even before it began. The subsequent loss of three treatment and nine control schools makes the consideration of 'power' irrelevant.

Randomisation and identification of eligible pupils

After schools were recruited, all Year 6 pupils in the 61 schools (pupil N=2,352) took the New Group Reading (NGRT) pre-test (a standardised test of literacy). Schools were then randomised, with 30 allocated to receive treatment and 31 to a waiting-list control. Randomisation was carried out by the evaluators after the pre-test. This was to minimise bias since knowledge of group allocation can potentially affect how pupils perform in the test. To ensure blinding of the evaluators, the randomisation was carried out publicly, using a mechanically shuffled pack of cards to select the treatment or waiting school.

All schools (control and treatment) were then meant to identify their eligible pupils (those who were at risk of not achieving Level 4 and likely to benefit from the intervention) using a combination of teacher's judgement about which child or group of children would benefit from the treatment and the NGRT data. In general, six to eight vulnerable target pupils were to be identified for each Year 6 class. Although not ideal, this sequence relative to school randomisation and testing was adopted at the request of the developers who wanted to use the pre-test results to identify eligible pupils, and with the permission of the funders, but against the advice of the evaluators. What happened in practice was that the list of eligible pupils provided by schools was not clear. There were 79 pupils from the control schools and 37 from the treatment schools whose eligibility was unknown. This was partly a consequence of the sequence, and the school-level rather than individual or class-level design insisted upon, and partly due to the lack of direct communication between the evaluators and the schools (see below).

Attrition

A total of 205 pupils from treatment schools and 312 from control schools were assessed as eligible. After randomisation, 11 schools (three treatment and eight control schools) dropped out, reportedly due to organisational issues as a result of changes in leadership. This is very high, and at a scale not
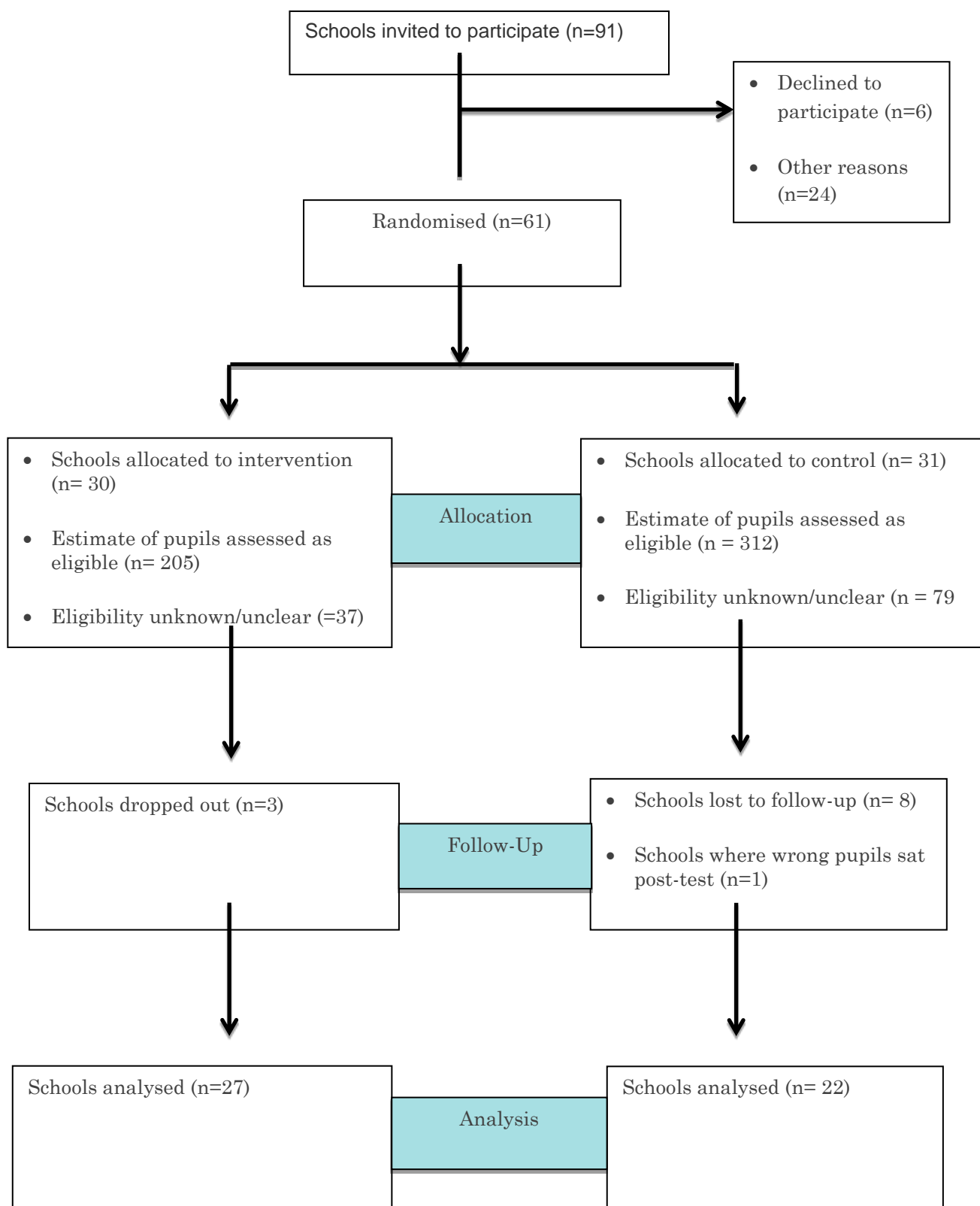
encountered by the evaluators before. In addition, one other control school conducted the post-test on the wrong year group of pupils. So valid data from only 49 schools was analysed (27 treatment and 22 control). The final number of eligible pupils with complete post-test scores was 181 (treatment) and 204 (control), making a total of 385. These were reasonably well balanced in terms of FSM-eligibility, sex and special educational needs (Table 1), and also prior attainment (see below). The control had more pupils of minority ethnic origin or speaking English as an additional language. This may or may not be relevant.

Table 1 – Percentage of pupils with known characteristics in each group, after attrition

	Treatment (N=181)	Control (N=204)
FSM-eligible	24	26
Males	54	55
Not White UK	20	32
English as additional language	15	22
Labelled SEN	39	40

Overall attrition was therefore in excess of 25%. Coupled with uncertainty about the identification of eligible pupils this means that the results of the evaluation must be treated as indicative only. The trial is therefore regarded as 'spoilt' as it has not met the CONSORT best practice recommendation regarding sticking to protocol, eligibility criteria and attrition. As such, the focus of the paper must be more on the lessons learnt about evaluating RTI in any future trials.

Participant flow diagram – Sample allocation and attrition

```
┌──────────────────────────────────────────┐
│ Schools invited to participate (n=91)      │
└──────────────────────────────────────────┘
                    │
                    │──────────────────────►  ┌────────────────────────────┐
                    │                          │ • Declined to               │
                    │                          │   participate (n=6)         │
                    │                          │                             │
                    ▼                          │ • Other reasons             │
┌──────────────────────────────┐              │   (n=24)                    │
│     Randomised (n=61)         │              └────────────────────────────┘
└──────────────────────────────┘
```

┌────────────────────────────────────┐   ┌──────────┐   ┌──────────────────────────────────────┐
│ • Schools allocated to intervention │   │Allocation│   │ • Schools allocated to control (n= 31)│
│   (n= 30)   │   └──────────┘   │   │
│   │   │ • Estimate of pupils assessed as   │
│ • Estimate of pupils assessed as   │   │   eligible (n = 312)   │
│   eligible (n= 205)   │   │   │
│   │   │ • Eligibility unknown/unclear (n = 79│
│ • Eligibility unknown/unclear (=37) │   │   │
└────────────────────────────────────┘   └──────────────────────────────────────┘

┌────────────────────────────────────┐   ┌──────────┐   ┌──────────────────────────────────────┐
│ Schools dropped out (n=3)   │   │Follow-Up │   │ • Schools lost to follow-up (n= 8)   │
│   │   └──────────┘   │   │
│   │   │ • Schools where wrong pupils sat   │
│   │   │   post-test (n=1)   │
└────────────────────────────────────┘   └──────────────────────────────────────┘

┌────────────────────────────────────┐   ┌──────────┐   ┌──────────────────────────────────────┐
│ Schools analysed (n=27)   │   │ Analysis │   │ Schools analysed (n= 22)   │
│   │   └──────────┘   │   │
└────────────────────────────────────┘   └──────────────────────────────────────┘

Analysis

Outcomes were measured using the New Group Reading Test (NGRT). NGRT is a standardised test of reading and comprehension. It is developed by GL Assessment and the National Foundation for Educational Research (NFER) and includes sentence completion and passage comprehension.

An 'intention-to-treat' analysis was agreed for this trial. This means that all pupils originally randomised and identified as eligible in both groups of schools should have been post-tested and their outcomes analysed, regardless of whether they remained in the trial or not. In practice, this was hindered by whole schools dropping out of the post-test, and by inconsistent identification of eligible pupils (due to the sequence of events above).

The impact of the intervention was measured using the 'effect' size (Hedges' g) for the post–test only results of NGRTB. However, due to school dropouts, the groups may no longer be equal. To ensure baseline equivalence the effect size for gain scores from NGRTA (pre-test) to NGRTB (post-test) was also assessed. Here the gain scores are taken to mean the average difference in scores between the pre-test and the post-test. The two tests are designed to be equivalent, and permit re-testing to assess progress over time.

To assess whether the finding is robust, that is whether it could occur by chance or due to bias as a result of attrition, we calculated the 'number needed to disturb the finding' of the effect size (NNTD). This is calculated by creating a counterfactual score consisting of the mean score for the control group, and subtracting its standard deviation. The number of such scores that can be added to the treatment group before the effect size disappears is the 'number needed to disturb the finding' (Gorard and Gorard 2015). The control counterfactual is used here because the treatment group is slightly smaller (and so is easier to 'disturb').

The analysis deliberately does not use significance testing or confidence intervals (Gorard 2015). They do not make sense anyway (Carver 1978), are routinely misinterpreted (Watts 1991), and can lead to serious mistakes for policy and practice (Falk and Greenbaum 1995). But even more obviously, they are mathematically predicated on complete random allocation (Berk and Freedman 2001), and take no account of sample quality or attrition (Lipsey et al. 2012). Given the level of attrition and possible bias in this study, such standard-error based calculations would be inappropriate, since there is no standard error here, even if they worked as envisaged by their remaining advocates.

Further analysis was carried out to see if the intervention would benefit disadvantaged children identified as those eligible for free school meals (FSM). The results are illustrated via a two-step multivariate regression analysis with NGRTB scores as the outcome and pupil background characteristics, including sex, ethnicity, first language, SEN and FSM, plus prior attainment (NGRTA) as predictor variables. These variables were all entered in one step. The binary variable representing allocation to treatment or control was entered in a second step.

Process evaluation methods

To assess the fidelity to implementation in the delivery of the intervention an evaluation of the process of implementation was carried out. This also included the observation of the post-testing to ensure that no special attention was given to either group that would bias the results since schools now had knowledge of their group allocation.

The process evaluation took the form of observation visits to training sessions, and field visits to schools. The developers arranged with five schools for us, the independent evaluators, to visit. One visit to each school was scheduled at the beginning of the intervention and another towards the end to record progress (if any) and changes in pupils' and teachers' behaviour and attitude. These visits were informal and non-intrusive involving participant observations, face-to-face interviews with pupils, trainers (Achievement Coaches, CUREE/AfA3As trainers) and teaching staff. Pupil interviews were either in pairs or small groups. These were loosely structured to get a sense of pupils' perceptions of the programme and to allow respondents to talk about what they thought was important, and not be constrained by a more structured interview protocol. Resources and protocol tools used in the programme were also collected. A total of 20 person visits were made to five treatment schools to observe the delivery of the intervention. Further visits to a sample of treatment and control schools were made to observe the conduct of the post test.

How effective is RTI as a catch-up intervention?

The two groups were reasonably well-balanced in terms of pre-test scores, as assessed by NGRTA, with the intervention group already slightly ahead (pre-test 'effect' size of just under +0.05). A post-test only comparison based on the NGRTB test (Table 2) suggests that RTI had a positive effect on reading and comprehension (effect size of +0.19). However, the results have to be taken as indicative because of the level of school dropout after allocation (25%), the number of schools in the control group which did not carry out post-testing, and the inconsistency with which pupils who were eligible for the intervention were identified.

Table 2 - Effect size based on all achieved NGRTB post-tests

	N	Mean score NGRTA pre-	Standard Deviation	Mean score NGRTB post-	Standard Deviation	Effect Size
Intervention	181	264.1	54.0	287.9	53.0	+0.19
Control	204	261.4	60.1	276.5	67.0	-
Overall	385	262.7	57.2	281.8	61.0	-

Note: 27 intervention and 22 control schools were involved in this analysis

The NNTD analysis shows that these results are quite likely to be due to chance or bias caused by attrition. For example, taking the mean post-test score for the control group and subtracting one standard deviation creates a score that would be counterfactual to the treatment group. It would take only 21 such scores added to the existing treatment group to eliminate the post-test effect size reported in Table 1. This is considerably less than the level of dropout, which amounted to at least 132 cases.

In theory, the randomised groups should have been equivalent. However, there was a slight initial imbalance (Table 2). This could be due to attrition after randomization, especially from the control group. So, a second analysis of impact was conducted using gain scores of NGRTA (pre-test) to NGRTB (post test). This showed an effect size of +0.15, which is similar to that using only the post-test scores (Table 3), and different in magnitude by the amount that might be predicted from the pre-test difference.

Table 3 - Effect size based on all achieved gain scores – NGRTA pre-test to NGRTB post-test

	N	Mean	Standard Deviation	Effect Size
Intervention	178	25.0	50.2	+0.15
Control	195	18.0	42.0	-
Overall	373	21.4	46.2	-

Note: the N differs slightly from Table 1 as some pupils did not have a corresponding pre-test score for a number of reasons (including being newly arrived in the school).

Although only the outcomes of pupils identified as eligible to receive the intervention were supposed to be compared, in reality it was not clear who these pupils were. Some schools did not provide this information, and some of the pupils who were not identified as eligible also took the post-test. It was not clear if any of these pupils did in fact receive the intervention despite being reported as ineligible. Further, some pupils who were identified as eligible had the highest pre-test scores in their schools, and even across all schools. According to the published criteria of eligibility they would not be deemed eligible for RTI, yet schools considered them as suitable for the intervention. So another analysis was carried out which ignored the incomplete list of reported eligibility provided by the developer, and considering only those pupils who were graded at Level 4C and below at pre-test (and should thus be deemed eligible for the intervention). However, this means that we cannot be sure that those in the 'treatment' group labelled below did in fact receive the intervention, which might dampen the impact. However, using these limited figures, the 'effect' was again positive (Table 4).

Table 4 - Effect size based on NGRTB post-test, cases initially below secure Level 4

	N	Mean	Standard Deviation	Effect Size
Intervention	171	286.6	51.3	+0.29
Control	180	270.3	59.5	-
Overall	351	278.3	56.2	-

Further analysis was carried out to see if the intervention benefited disadvantaged children in particular, defined as those eligible for free school meals (FSM). The results suggest that the intervention may be especially beneficial for FSM pupils (Table 5).

Table 5 - Effect size based on achieved NGRTB post-tests, FSM only

	N	Mean	Standard Deviation	Effect Size
Intervention	44	284.6	70.5	+0.48
Control	52	251.3	68.3	-
Overall	96	266.6	70.9	-

The two-step regression model shows that the bulk of the variation in NGRTB scores (main outcome measure) that can be explained is explained by background variables and prior attainment (NGRTA). These variables were entered at Step 1. Step 2 shows that being in the treatment group makes much less difference to the outcome (Table 6). It has to be mentioned that the regression does not represent a test of causation but does provide a caution about the possible strength and importance of the intervention in relation to pupil characteristics.

Table 6 -Variation explained (adjusted R-squared) in two-stage regression model, NGRTB scores

	NGRTB outcome scores
Step 1 – background and prior attainment	0.48
Step 2 – intervention or not	0.48 (i.e. no discernible increase)

Table 7 presents the standardised coefficients for all variables retained in the model. The largest of these by some way is the pre-test score (NGRTA). This is the best single predictor of the post-test score. There is only a very slight positive standardised coefficient for being in the treatment group (+0.05), again suggesting that the intervention may have contributed a little to pupils' performance.

Table 7 - Standardised Exp(B) coefficients for the regression model in Table 6

	NGRTB outcome scores
Step 1	
FSM	+0.03
Sex (female)	+0.06
SEN	-0.02
EAL	-0.05
Ethnicity (White UK)	+0.05
NGRTA score (pre-test)	+0.65
Step 2	

	NGRTB outcome scores
Treatment (or not)	+0.05

In summary, the trial results suggest a positive effect from RTI for NGRT, and that the intervention may be particularly beneficial for FSM-eligible pupils. However, the results can only be indicative given that it was not clear which pupils who were identified as eligible actually received the intervention, and a lot of the variation in outcomes could be related to the characteristics of the pupils who remained in the study. Nevertheless, there is some promise here for a bigger and better controlled trial, taking into account the lessons that follow.

Implementing RTI and the testing

As part of the process evaluation the team visited five schools to observe the programme in action and to gather information about pupils' and teachers' perceptions of the intervention. Interviews with staff and pupils were conducted during these visits. Twenty-six pupils were interviewed either in pairs or in small groups. In total we interviewed six teachers, two headteachers and two deputy heads, five Achievement Coaches and one higher level Teaching Assistant. Visits to the schools were arranged and agreed with CUREE (the developer) and their Achievement Coaches. As such, there is potential of biased selection in that these schools could be the more successful implementers. Therefore, the views here may not be representative of all of the schools in the trial.

The process evaluation revealed wide variations in the intensity and frequency with which schools implemented RTI. Some schools managed only five sessions of 20 to 30 minutes each in total, and most schools struggled to fit in only two to three sessions instead of five a week. This was too short for proper monitoring and tracking and adjustments to intensity, and would not be deemed sufficient in the protocol for impact to be realised. Many teachers were unable to conduct as many sessions as they had intended. This was largely due to the timing of the programme, being introduced in the last few weeks of the final term after the National Curriculum assessment also known as SATs (Standard Assessment Tests), which was in turn a result of political timing behind the extra funding. It was also not clear whether every teacher used the monitoring and progress tracking tools. There was no evidence that any school had completed enough sessions for a complete cycle of RTI. No school had indicated adjusting the level of intensity as a result of the monitoring and tracking process. All of this should cast further doubt on the results of the impact evaluation.

There were also variations in the levels of implementation; some schools were able to deliver all the three levels or tiers (whole class, small group and one-to-one) while others decided to deliver only one level of RTI.

Almost all previous studies of RTI highlighted the importance of fidelity of implementation, such that the amount of time allocated to treatment corresponded with the actual amount of time children received the treatment. This did not happen in this study. Important lessons were learnt from this experience which will inform future implementation of RTI in the UK. These will be discussed in detail in the section on recommendations. The following section discusses the main barriers to implementation.

Barriers to implementation

The biggest challenge to successful implementation was the very short duration given to the intervention. This was primarily because of the funder's requirement that it should be implemented in the last few weeks of the summer term. This short duration meant that teachers could not go through the complete cycle from needs analysis, delivery of the intervention to tracking and monitoring. The last two stages were given the least attention. Some schools conducted so few sessions in all that teachers could not realistically monitor and make adjustments in the middle of the intervention. To do so would mean that teachers had to make the assessments by the second session. One teacher remarked that there was just 'not enough time to do it properly.'

Another issue repeatedly raised by the schools was the timing of the intervention. As the intervention was implemented in the last few weeks of term, there were lots of distractions taking away time for the intervention – for example end-of-term activities like sports days, prize-giving and rehearsals for end-of-year productions. A number of other transition programmes were also arranged for Year 6 pupils, such as visits to secondary schools, school camps and secondary school integration programmes. In one primary school the pupils spent a whole week in their secondary school, thus losing a whole week of instruction. Feedback from staff suggests that the programme would have had greater impact if it had started in Year 5 or at the beginning of Year 6 to allow time for adjustments and for the intervention to develop.

Schools reported a high rate of absence in the last term after SATs, meaning that a number of planned sessions were missed.

Finding appropriate short texts for teaching proved challenging for teachers. As the resources were not fully developed, teachers had to look for relevant materials for each session. It was not only time-consuming but finding texts that were both age-appropriate and interesting was also not easy.

Another challenge for teachers was planning multiple lessons for each session where there was more than one tier of intervention. Even with only one tier, teachers still had to prepare one lesson for the intervention children and one for those not targeted for the intervention. In one school the teacher had three preparations: one for the whole class, one for the RTI pupils and one for the weaker pupils. So each literacy period could involve two to three sets of materials and instructional strategies.

Finding space to accommodate different groups of children receiving different intensities of treatment within one period could be a potential barrier to effective implementation. Some schools could only have one tier because of the logistics of finding rooms to accommodate pupils in other tiers.

Despite these challenges, the intervention was well-received. The participants - teachers, pupils and Achievement Coaches – had positive things to say about the intervention. The way schools were willing to work out the complex logistics of arranging classes was one indication of their commitment to the programme. Additional support staff were also enlisted.

What were the perceived benefits of RTI?

Accounts from teachers, pupils and Achievement Coaches suggested that RTI has beneficial effects on pupils' literacy as a catch-up literacy programme. One school claimed that the data they collected showed that their pupils had made an equivalent of five months' progress in comprehension and spelling in the four weeks, and in some cases as much as a year's progress. The teacher reported how two pupils improved their reading fluency from reading 200 words in three minutes to nearly half of that time. Dramatic improvements in reading were also observed. At a later assessment, all of their targeted pupils achieved Level 4. According to the teacher:

'This has never happened in the history of the school before.'

However, teachers added that such progress could be due to the increased attention given.

The teacher in one school attributed the marked progress made by her pupils to the repeated reading intervention. She explained how pupils were now able to read faster, achieving better results with each new text. On the day of the observation, one read the text in 1.4 minutes, aiming for 1.2 the following day. The other pupil had improved in his spelling which boosted his self-esteem.

Pupils also remarked on how the activities had helped them.

'I remember things that I learn now.'

'I can do a lot more – better with spelling, reading and writing.'

The programme also appeared to help develop independent learning. Pupils said they were taught to mark their own work and check for mistakes.

A number of schools reported improvements in reading. One pupil told us how she used to hate reading but since the programme was introduced she was enjoying it. Some pupils said they now preferred books to films as books allowed them to use their imagination and there was also an element of suspense. They explained how they were excited to find out what happened in the books. Some claimed that they were reading more books than in previous years and were enjoying reading more now than before. Of course, this could be the result of maturation, or even of coaching. When asked how they rate the programme and what they liked, some said:

'I would give 10 out of 10 for PALS and 5 out of 5 for the Kenusuke's Kingdom.'

'Others won't laugh at me because I can read better.'

In another school a teacher reported how he had to switch to 'harder' spelling lists because of the rapid improvements his pupils made. One teacher gave an account of how one of the pupils, who had no concept of decoding and had problems with speech and was unable to spell properly, had shown remarkable improvement in spelling using the phonic intervention activity despite the short duration of the intervention. The pupil said:

'Now if I don't know a word I slow down and think.'

Other pupils explained how the use of mirrors also helped them with their spelling as they could see all the letters when mouthing the words. They could see how their mouths shaped the words and how seeing the shape of the word had added 'texture' so when they read they could feel the word.

'With a tricky word, now I say it and I write the sounds down.'

Pupils gave accounts of how the RTI strategy had helped them learn spelling. For example, one pupil said he realised that to spell the word 'sweets' he only needed to change one letter of the word 'tweets'.

Improvements in reading comprehension and writing were also reported. Pupils enjoyed the intervention on inference reading, and how it had helped their writing. They said they were now able to write in a more interesting and descriptive way. When they looked back at their previous year's writings they could see how much their writing had improved. Pupils said that feedback from their teachers had suggested that their writing had improved. For example:

'The teacher said my writing is better this year.'

'I have more good comments from my teacher like "well done".'

The inference intervention was believed to facilitate reading comprehension. Pupils reported how the strategy helped them to understand what they read in story books, gave them more ideas and helped them to be more imaginative in their writing. Other pupils reported improvements in the use of better expressions and a wider vocabulary, which had enhanced their writing.

Beyond the actual outcomes, the intervention apparently also empowered some pupils, raising their self-esteem and confidence. From the 21 schools which sent completed tools to CUREE, there was evidence of variation in the intensity of use and this broadly corresponded with teacher-observed effectiveness on pupil outcomes. The tools that were used in more detail and depth showed high levels of progress in a variety of areas, including pupil confidence, and some of these schools reported increases in spelling and reading ages of 9 to 10 months.

Generally, RTI as a literacy intervention was perceived by schools as having impact on both pupils' literacy and non-academic outcomes. One school perceived the intervention as such a success that they were considering a home version of the programme to monitor further results at Year 7. The headteacher of another school was so confident in the intervention, having seen the results it had produced among their Year 6 pupils, that she planned to roll out part of the RTI programme to other year groups. In another school, the teacher was so encouraged by the impact she saw that she was already preparing to introduce the intervention to her new Year 6 class. Schools were generally so confident in the effectiveness of the intervention that they had asked for a similar programme for numeracy. When asked if they would continue with the programme after the trial, almost all the teachers interviewed said they definitely would.

Conclusions

It is quite common for participants in trials to report that everything is going well even when the impact evaluation shows that the intervention has failed or is harmful (see, for example, Khan and Gorard 2012). However, that is not quite the situation here because the impact evaluation does have some suggestion of promise, especially for disadvantaged children. It is possible that the optimistic assessment of participants could be accurate, and that the evidence provided by the impact evaluation could not be definitive due to lack of clarity about pupil eligibility and school and pupil dropout from both arms of the trial. The most important lessons from this study are therefore not about impact but about the conduct of such a trial, and the implementation of RTI in schools.

Lessons for future trials

One of the reasons for the 'spoilt' trial was the unclear or inconsistent identification of eligible pupils. It is therefore recommended that in future trials schools would have to identify eligible pupils prior to randomisation in order to ensure that the pupils in the control group were identified under the same conditions and timing as those in the intervention group. Although this sequencing was originally proposed by the evaluator, it was agreed that pupil identification could take place after randomisation due to the short timeframe in which the intervention was to be delivered. This would not have compromised the trial if the eligible control pupils had been identified early, and this data shared with the evaluator. In practice, however, the data on the eligible control pupils were not made available to the evaluator until after the delivery of the intervention. This would not have been a problem either if the data eventually given were reliable. It was not, so as a result the findings of the evaluation were substantially weakened.

The security of the findings was also undermined by the high dropout, especially from the control schools which did not provide test data. This is reasonably common, especially in large-scale randomised controlled trials. For example, in an evaluation of Quest, a literacy intervention (Biggart 2015a), 45% of pupils in the treatment group and 15% in the control group did not take the post-test. This arose due to miscommunication with schools where they only tested pupils who received the intervention. In addition, close to half of the schools originally committed to the programme dropped out after allocation due to issues with the programme. In another trial of an intervention called Tutoring with Alphie (Biggart 2015b), also funded by the EEF, there was an even higher attrition (70% from treatment group and 73% from control). Again this was due to teachers' dissatisfaction with the programme.

Four recommendations are proposed to minimise the risk of such events happening:

Effective communication between schools and evaluators
Experience from other trials suggests that direct communication between evaluators and schools was a more efficient way of communication. Information and instructions got passed quickly and acted upon promptly.

Have no more than two parties involved
The number of parties involved added further complication. In the RTI trial there were five parties involved: the EEF, the evaluators, the developers (CUREE), AFA3As and the schools. Ideally future studies will have as few parties involved as possible as this additional layer of communication between the evaluators and the schools was not helpful.

Our evaluations of the Fresh Start (Siddiqui et al. 2015a) and Accelerated Reader programmes (Siddiqui et al. 2015b) demonstrated that when schools themselves conduct the trial and deliver the intervention, with no involvement with a third party (the programme developer), full and complete data were obtained with little or no attrition. This is also partly because schools received training prior to the trial on the craft of conducting a trial and the importance of data collection, and the implications of attrition were impressed upon them.

Ensure that schools understand their commitment to the trial
This is necessary to ensure that schools agreeing to take part in the trial are committed to provide all relevant data. In this study, some schools were under the impression that because they were not continuing with the intervention, there was no need for their pupils to take the post-test. Similarly, in an evaluation of the World and World Reading programme (See et al. 2015), schools in both the control and treatment groups wrongly thought that since they did not intend to take up the programme after the trial, it was not necessary to post-test the pupils. Apparently, schools discussed this with the developer who agreed for them not to take the post-test. Some schools also routinely excluded pupils with special educational needs from tests. In the example of the Quest programme, schools thought that only pupils directly involved in Quest needed to take the post-test. In an ongoing evaluation of a Youth Social Action trial by the authors of this paper, some schools only tested pupils who opted to participate in the uniformed group even though the purpose of the trial was to compare the outcomes of those who were in uniformed group and those who were not. This was despite the schools receiving individually named tests for all eligible pupils.

Training of programme deliverers and developers
Developers may be expert in their own intervention, but not necessarily expert in conducting evaluations. They also have a clear conflict of interest. Where evaluations were conducted by developers themselves or commissioned by developers, important aspects of a trial concerning validity were not heeded, such as lack of randomisation of participants, but using volunteers for the programme instead (e.g. Brodsky 1994; Fagan & Iglesias 1999), having no comparison groups (e.g. Fiala & Sheridan 2003; Mobley 2012; McDonald & Fitzroy 2010; McDonald et al. 2006) or high or unreported attrition. In some cases how participants were selected were not known. In one example, the developer of a computer-based reading programme claimed that their trial showed that their software could improve pupils' reading after 6 weeks of exposure to the programme. An independent evaluation (Khan and Gorard 2012) revealed that although the developers were right in their claim, the control pupils made an even bigger improvement in that time compared to the treatment group. As noted above, where schools conduct their own trial training helped minimise incidents of dropout and missing data.

The examples given clearly show that it cannot be assumed that programme developers understand how to conduct trials. Both the developers and those delivering the intervention need to be trained to understand the importance of complete and accurate data.

Future evaluations could consider a pre-intervention workshop for schools about the process of evaluation, the importance of complete data, keeping attendance records (for assessing impact of dosage) and commitment to testing. The implications of failure to comply would need to be spelt out.

Close monitoring of the programme

Fidelity of treatment could be improved with close monitoring and supervision of teachers to ensure that the required number of sessions was carried out as proposed. In Vaughn and Fletcher (2012), researchers were hired to supervise all the treatment teachers and documented the actual treatment time. Although this study involved Achievement Coaches who were hired by AfA3As to support schools, their role was not a supervisory one.

Where possible the post-testing should be monitored closely by independent evaluators to try and ensure that the correct pupils are tested.

Genuine interest in the quality of the trial
It may also be the case that in the Fresh Start and Accelerated Reader trials (above) the schools have ownership of the trial and genuinely want to know if the programme works. It is therefore, important for anyone conducting a trial to care more about the quality, validity and reliability of the findings than what the results are.

Such pre-emptive measures not only improve the security of findings, but also prevent unnecessary waste of funders' money and time and effort put in by everyone concerned – pupils, teachers, staff, developers and evaluators. Any missing data or missing cases potentially bias the results and render the findings insecure. This means that all the money and efforts put in will be wasted if we cannot make any meaningful conclusions from the study. It is unethical to allow such things to happen, so any effort to prevent such things from happening should be taken.

Lessons for successful implementation of RTI

The biggest challenge to implementation of RTI was the timing and the short duration. Not enough time was allowed for the programme to develop. The process evaluation suggests that the intervention would have been more effective had it been started at the beginning of the year and allowed to run over a longer period of time, rather than as a catch-up intervention in the busy period at the end of Year 6. Some schools were only able to implement a small number of sessions, which ~~made~~ meant it was not possible to implement, monitor and adjust the intervention as intended.

Another constraint was inadequate preparation time for teachers to plan additional lessons and to source for suitable teaching resources and texts.

The following are recommendations for anyone wishing to implement RTI in their schools:

1. Support from the management is key to successful implementation to allow teachers time to plan and prepare lessons, collect and develop instructional materials. This is necessary for teachers to deliver the programme satisfactorily. Time is also needed for the testing process, which some schools see as quite disruptive. Support from the top management is therefore necessary. As has been shown in the examples given above, if the headteacher does not believe in a programme, teachers will find it hard to justify the time for testing.

   In this study, it was observed that where schools had the full support of the management, the programme was implemented on a more frequent and regular basis, and where schools did not have the support of the management, implementation tended to be sporadic. For example, in one

school the headteacher personally arranged for rooms to be available for the different tiers. In another, the head deployed teaching assistants to assist the teachers in implementing the programme, and one headteacher arranged for a higher level teaching assistant to take the class while the form teacher took the Tier 2 small groups. A deputy head monitored pupil progress using the data collected and personally arranged the schedules for the testing. She even went out to buy sweets for the children to thank them for taking the post-test. Where the support of the management was absent or not strong, schools were more likely to drop out of the trial. For example, four schools in the study withdrew from the trial due to changes in leadership and one school could not continue due to organisational problems. In these cases when a new head arrives, they have a different agenda and priority. This can be detrimental to the integrity of any trial. The importance of management support can be seen in the other trials that we were involved in. For example, in our evaluation of the programme Philosophy for Children (Siddiqui et al. 2015c), some schools completely abandoned the programme and some implemented it only occasionally. Similar experience was seen with the Word and World Reading programme (above). Our evaluation of Fresh Start, Accelerated Reader and the Anglican School Partnership feedback trials (See et al. 2014) were successful with no dropouts and had full commitment from teachers only because the schools themselves had asked to run the trial. They were all initiated by the headteachers and thus had the support of the management.

2. If the programme was to be used in schools it should be introduced at the beginning of a school year, running through the whole year to allow the full cycle of the intervention to be implemented. The duration for the intervention should be long enough to allow for the full cycle of RTI to be observed, from identifying target pupils to determining the level or intensity of intervention and implementation. Time is needed to allow the intervention to develop before teachers can track, monitor the progress of the pupils and adjust the intensity. Elapsed time is therefore essential for a proper round of tracking and monitoring.

3. Ideally, it should be introduced in the penultimate year of transition (that is Year 5 rather than Year 6).

4. To get optimal benefit from the intervention, lessons should ideally be carried out on a more frequent and regular basis. Positive studies in the US (e.g. Vaughn et al. 2010; Vaughn and Fletcher 2012) involved pupils receiving RTI on a daily basis for 50 minutes over a one-year period. If the study was to be scaled up, one recommendation would be to increase the intensity and duration of the programme.

5. The RTI programme implemented in this trial did not come with pre-packaged resources, although sample materials were provided to start teachers off. Therefore, for successful implementation of the programme, suitable materials for teaching the different components of reading and writing should be made readily available and easily accessible. This could help cut preparation time and make it more attractive to teachers who may not be sold on the idea in the first place. Teachers are already overwhelmed with work, so anything that could make it easier for them to do their job is likely to be welcomed.

Many teachers shared the same view that it was time-consuming to look for appropriate and interesting materials. However, once developed these resources could be used in subsequent

years. It would be a good idea if all the schools could come together and share these resources and talk about successful lessons. In future a library or bank of resources or a website could be developed although supporting access to the resources would need to be carefully considered.

6. More importantly, teachers and the school as a whole need to be committed to the intervention. It is essential that teachers have a positive attitude and are willing to adapt to changes in routines. The successful implementation of intervention relies on the general support given by the head and the ability of schools to use the data (NGRT pre-test data, teacher assessment data and SATs data) to inform decisions. It was very apparent, for example, in some schools, how the achievement data was feeding into the general work and the interventions that were already in place.

Endword

Although research in the US suggests that RTI is a promising approach in improving literacy, the current evaluation, while showing some promising outcomes, is unable to show conclusively if this approach is effective in UK schools. The process evaluation shows that the approach was popular with teachers and pupils, and was considered likely to have had positive benefits on wider outcomes such as confidence and self-esteem. Important lessons have also been learnt from the process evaluation which can help inform future trials. In fact more may be learnt from this damaged trial than some more successful ones.

Acknowledgements

References

Berk, R. and Freedman, D. (2001) *Statistical assumptions as empirical commitments*, http://www.stat.berkeley.edu/~census/berk2.pdf, accessed 030714

Biggart, A. (2015a) An evaluation of Quest, a whole-group literacy intervention. London: Education Endowment Foundation

Biggart, A. (2015b) An evaluation of Tutoring with Alphie, a computer literacy-tutoring programme. London: Education Endowment Foundation

Brodsky, S. (1994) *An Urban Family Math Collaborative* (CASE-09094). New York: City

Carver, R. (1978) The case against statistical significance testing, *Harvard Educational Review*, 48, 378-399

Clay, M. (1991) *Becoming literate: The construction of inner control*. Auckland: Heinemann

Fagan, J. and Iglesias, A. (1999) Father involvement program effects on fathers, father figures, and their head start children: A quasi-experimental study, *Early Childhood Research Quarterly*, 14, 2, 243-269

Faggella-Luby, M. and Wardwell, M. (2011) RTI in a Middle School: Findings and Practical Implications of a Tier 2 Reading Comprehension Study*, Learning Disability Quarterly*, 34, 1, 35-49.

Falk, R. and Greenbaum, C. (1995) Significance tests die hard: the amazing persistence of a probabilistic misconception, *Theory and Psychology*, 5, 75-98

Fiala, C. and Sheridan, S. M. (2003) Parent involvement and reading: Using curriculum-based measurement to assess the effects of paired reading, *Psychology in the Schools*, 40, 6, 613-626. DOI: 10.1002/pits.10128

Fountas, I. and Pinnell, G. (1996) *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinenmann

Fuchs, L., Fuchs, D., Compton, D. (2010) Rethinking Response to Intervention at Middle and High School, *School Psychology Review*, 39, 1, 22-28

Galton M, Gray J, Ruddock J. (1999) *The impact of school transitions and transfers on pupil progress and attainment.* DfEE Research Report No. 131. Norwich: HM's Stationery Office; 1999

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

Gorard, S. (2015) Rethinking "quantitative" methods and the development of new researchers, *Review of Education*, 3, 1, 72-96, doi: 10.1002/rev3.3041

Gorard, S. and Gorard, J. (2015) What to do instead of significance testing? Calculating the 'number of counterfactual cases needed to disturb a finding', *International Journal of Social Research Methodology*, (forthcoming)

Gov.UK (2012) *£10 million to boost literacy*, https://www.gov.uk/government/news/10-million-to-boost-literacy-for-year-sevens, accessed 14/7/14

Graves, A., Brandon, R., Duesbery, L. McIntosh, A. and Pyle, N. (2011) The Effects of Tier 2 Literacy Instruction in Sixth Grade: Toward the Development of a Response-to-Intervention Model in Middle School, *Learning Disability Quarterly*, 34: 1, 73-86 (full paper not available, analysis based on abstracts)

Hargreaves A. and Galton M. (1999) *Moving from the primary classroom: 20 years on.* London: Routledge

Khan and Gorard (2012) A randomised controlled trial of the use of a piece of commercial software for the acquisition of reading skills, *Educational Review*, 64, 1, 21-36

Leroux, A., Vaughn, S., Roberts, G. and Fletcher, J. (2011) *Findings from a Three-Year Treatment within a Response to Intervention Framework for Students in Grades 6 with Reading Difficulties*, Paper presented at the Society for Research on Educational Effectiveness Conference, (no details about place and date of conference) http://www.eric.ed.gov/PDFS/ED518866.pdf.

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*, Washington DC: Institute of Education Sciences

McDonald L. and Fitzroy, S. (2010) *Families and Schools Together (FAST): Aggregate FASTUK evaluation report of 15 schools in 15 local education authorities (LEAs) across the UK*. London: Middlesex University

McDonald, L., Moberg, D.P., Brown, R., Rodriguez-Espiricueta I., Flores, N.I., Burke, M.P. and Coover, G. (2006) After-School Multi-Family groups: A randomised controlled trial involving low-income, urban, Latino children, *Children and Schools*, Vol. 28, No 1, p25-34.

Mobley, J. (2012) The impact of home-school collaboration on student achievement in K-5 reading and math. EdD thesis. Walden UniversitySainsbury M, Whetton C, Keith M, and Schagen, I. (1998) Attainment on transfer at age 11: evidence from the Summer Literacy Schools evaluation. *Educational Research*, 40, 1, 73-81

See, B.H. and Gorard, S. (2014) What works for literacy catch-up during transition to secondary school? *British Journal of Education, Society and Behavioural Science,* 4, 6, 739-754

See, B.H., Gorard, S. and Siddiqui, N. (2014) Evaluation of the Anglican School Partnership Effective Feedback trial. London: Education Endowment Foundation

See, B.H., Gorard, S. and Siddiqui, N. (2015) Evaluation of the Word and World Reading programme. London: Education Endowment Foundation

Siddiqui, N., Gorard, S. and See, B.H. (2015a) Evaluation of Fresh Start, a catch-up literacy intervention. London: Education Endowment Foundation

Siddiqui, N., Gorard, S. and See, BH (2015b) Accelerated Reader as a literacy catch-up intervention during the primary to secondary school transition phase, *Educational Review,* 10.1080/00131911.2015.1067883

Siddiqui, N., Gorard, S. and See, B.H. (2015c) Evaluation of Philosophy for Children. London: Education Endowment Foundation

Vaughn, S. and Fletcher, J. (2012) Response to intervention with secondary school students with reading difficulties, *Journal of Learning Disabilities*, 45, 3, 244-256

Vaughn, S., Cirino, P., Wanzek, J., Wexler, J., Fletcher, J. and Denton, C. (2010) Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review*, 39, 3-21

Watts, D. (1991) Why is introductory statistics difficult to learn?, *The American Statistician*, 45, 4, 290-291