



National College for
Teaching & Leadership

Closing the gap: test and learn

Executive summary

Winter 2016

Richard Churches - CfBT Education Trust

Contents

| | |
|--|----|
| List of figures | 3 |
| 1 What is closing the gap: test and learn? | 4 |
| 2 Programme phases | 5 |
| 2.1 Consultation phase | 5 |
| 2.1.1 Intervention selection process | 5 |
| 2.2 Capability phase | 5 |
| 2.3 Dissemination phase | 6 |
| 3 Results | 8 |
| 3.1 Conclusions with regard to the large-scale trials | 8 |
| 3.2 Evaluation of the extent to which the pupils' attainment gap (compared to expected pupil progress) had been closed | 9 |
| 3.3 Limitations and important considerations when interpreting the results | 11 |
| 4 Findings from the 'early adopter' strand | 12 |
| 4.1 The effectiveness of teacher-led randomised controlled trials and other forms of experimental research | 12 |
| 5 Conclusions | 13 |
| 5.1 Summary of programme findings from all the areas of delivery and engagement (including qualitative data from teacher surveys during the life of the programme) | 13 |
| 5.2 Policy implications and recommendations | 14 |

List of figures

| | |
|---|----|
| Figure 3.1.1: Combined effect sizes for all pupils involved in the trials | 8 |
| Figure 3.1.2: Combined effect sizes for pupils eligible for FSMs | 9 |
| Figure 3.2.1: SAS points reduction in attainment gaps for the year 1 trials and RLS (trialled for the first time in year 2 of the programme) – control and intervention | 10 |
| Figure 3.2.2: SAS points reduction in attainment gaps for AFA (years 1 and 2) – control and intervention | 10 |
| Figure 3.2.3: SAS points reduction in attainment gaps for the year 2 replicated trials – control and intervention | 11 |

1 What is closing the gap: test and learn?

Closing the gap: test and learn is the first programme in the world to trial multiple interventions simultaneously using a wholly collaborative approach across a large number of schools. Seven interventions were chosen through an extensive and systematic consultation and review of interventions seen as most likely to close the attainment gap for pupils with achievement below the national average in literacy and numeracy.

Collaborative randomised controlled trials (RCTs) were then conducted to evaluate these interventions and four of the interventions were replicated. Alongside this, teachers were trained in a range of research methods. A total of 50 teacher-led experimental studies (including RCTs designed and conducted by schools) were grant funded.

This report describes the programme and its delivery. It also outlines the findings from the large-scale trials, learning from developing teachers' scientific literacy through the school-led research programme and conclusions regarding the efficacy of system-led research approaches such as the ones embedded within the initiative.

2 Programme phases

There were three phases to the programme:

- a consultation phase (January to August 2013)
- a capability phase in which the research programmes and training of schools took place (September 2013 to July 2015)
- a dissemination phase involving an event for early adopters of teacher-led experimental research (as they became known) and a national dissemination event involving focus groups (October and November 2015)

2.1 Consultation phase

The purpose of the consultation phase was to identify a set of interventions which the current evidence supported as being effective in closing the attainment gap for lower-performing pupils, with a view to evaluating them using large-scale RCTs. It was also during the consultation phase that the research design for the first phase of trialling and timelines for pre- and post-testing were determined.

2.1.1 Intervention selection process

The first part of the consultation took the form of online surveys and focus groups with partner teaching schools and schools who had expressed an interest in contributing. The teaching schools' research & development advisory group then carried out a ranking exercise for the long-list of interventions, considering them in terms of likely take-up and manageability for large-scale trialling. The final list of interventions selected was as follows:

- 1stClass@Number (1stClass)
- Achievement for All (AfA)
- Growth mindsets
- Inference training
- Numicon intervention programme (NIP)
- Research lesson study (RLS)
- Response to intervention(breakthroughs in literacy) (RTI).

2.2 Capability phase

There were three elements to the capability phase of the programme:

1. the delivery of comprehensive training for teaching schools
2. the provision of support to teaching school trial co-ordinators
3. teaching school testing and intervention delivery within trial site schools

The delivery of comprehensive training for teaching schools participating in the closing the gap: test and learn programme covered rigorous and robust research methods appropriate for use in schools, including quantitative research methods such as RCTs, so that teachers gained an awareness of research methodologies (set-up, design and evaluation) and were able to contribute effectively to the trials. This also ensured that teachers in different contexts were able to deliver the interventions under trial in a consistent manner. The strand of work delivered through the research development and networking events (RDNE) focused on training teachers in the delivery of small-scale RCTs (and other forms of experimental research) and immediately yielded school-level activity. In response to this, the National College for Teaching and Leadership (NCTL) made available 50 'early adopter' grants to support participating teaching schools and their alliances in delivering their own small-scale RCTs. A total of 48 of these studies were presented at a conference poster event at the NCTL in Nottingham on 21 October 2015.

Participating teaching schools and trial site schools carried out a programme of testing over a period of two academic years (September 2013 – July 2015). This focused on assessing whether the seven selected interventions made a positive difference and whether such effects may be replicable and transferable.

Unlike many of the other RCTs that have taken place in education, schools were given a grant to cover the cost of the training and testing. It was the schools themselves who contacted the commercial providers, booked their teachers onto the training (within prescribed windows), and attended the intervention training courses, without Department for Education (DfE)/NCTL direct oversight. The same approach was taken with regard to the purchasing of the relevant pre- and post-tests from GL Assessment.

2.3 Dissemination phase

Two dissemination events took place – on 21 October 2015 in Nottingham, and on 18 November in London. The first event was attended by teachers who had conducted their own teacher-led RCTs (and other forms of experimental research) ('early adopters') together with some invited guests. The second event was open to all participating teaching schools and included presentations summarising the large-scale trial findings and the small-scale teacher-led studies alongside two teacher research presentations

Both of these events contained focus group sessions, the findings from which are discussed below.

Finally, this report was drafted with contributions from the Education Development Trust (formerly CfBT Education Trust), the University of Oxford Department of Education (OUDE) partner team, Centre for the Use of Research and Evidence in Education (CUREE), Durham University and NCTL. Large-scale trial results analyses were undertaken by analysts from the DfE.

3 Results

3.1 Conclusions with regard to the large-scale trials

Within the limitations of the research design and its implementation, the following conclusions can be drawn:

Overall, teaching school associated existing practice (the control conditions for all of the large-scale trials) appears to be at least equal to six of the top seven interventions identified in the consultation but better than growth mindsets when used with an average group of pupils. Existing teaching school practice may also be better than the first year of AfA with regard to the exposure of free school meals (FSM) pupils to this treatment.

The exception to the above was NIP which consistently appeared to improve mathematics attainment and progress rates, particularly for FSM pupils, and irrespective of the analytical model used to assess its efficacy. There may also be gap closure benefits in the use of RLS; however, in the context of the present study design none of the assessments produced significant results.

Although finding that the majority of the interventions showed no effect greater than existing good practice is useful, it is but a starting point for further investigation. The established practice in other fields (e.g. medicine) would be to undertake further trials in different circumstances to see if the benefits of these interventions are revealed in different contexts (for example, in struggling schools, with a more tightly defined group of students, or particular age groups).

Figures 3.1.1 and 3.1.2 show the effect sizes and 95 per cent confidence intervals for each trial. Y1 refers to the first year of the programme, while Y2 refers to the second year.

Figure 3.1.1: Combined effect sizes for all pupils involved in the trials

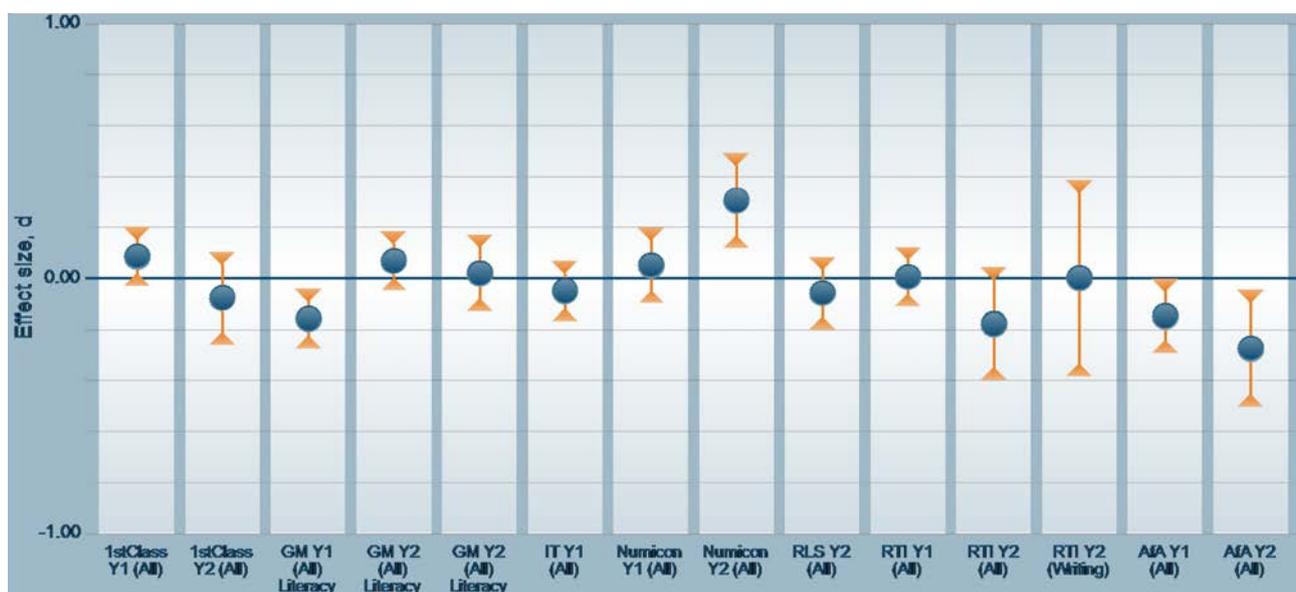
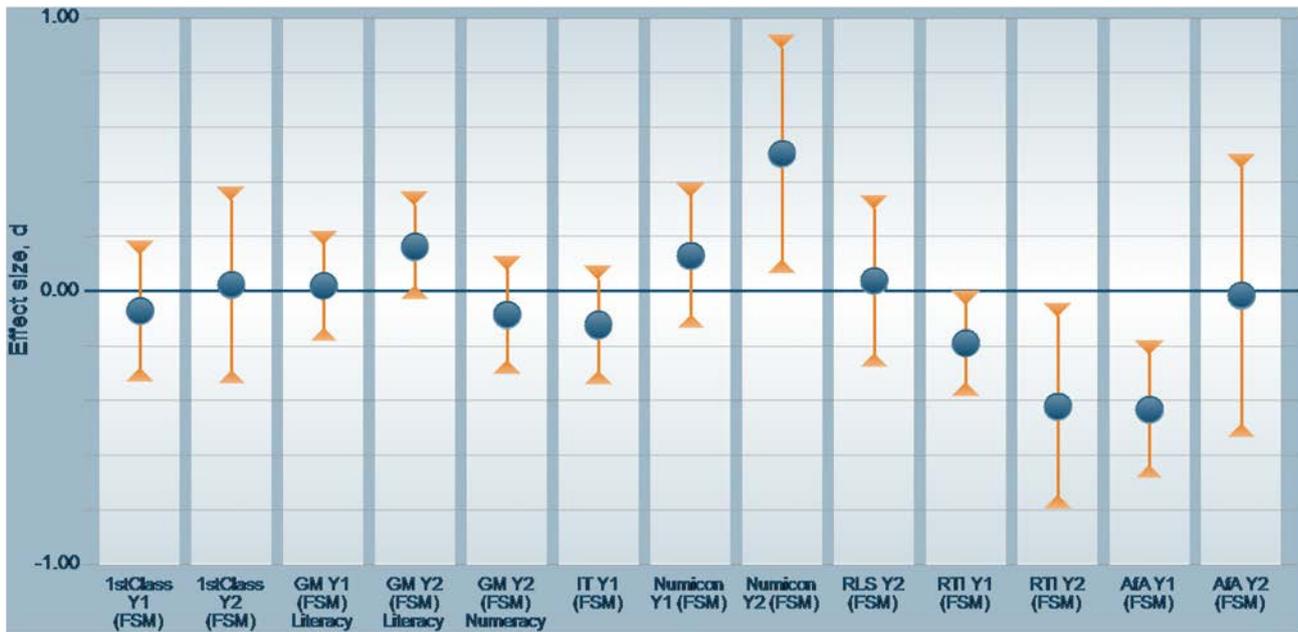


Figure 3.1.2: Combined effect sizes for pupils eligible for FSM



3.2 Evaluation of the extent to which the pupils' attainment gap (compared to expected pupil progress) had been closed

GL Assessment standard age scores (SAS) are based on the student's raw score adjusted for age and placed on a scale that makes a comparison with a nationally representative sample of UK students of the same age. The national average score is 100. Using the GL Assessment SAS national average score as a baseline, figures 3.2.1 and 3.2.2 show how the attainment gaps between FSM pupils and other pupils changed over the period of the trials.

In the trial design used in year 1, schools selected whole classes of pupils for each trial. It was therefore possible to compare mean standard age scores for FSM pupils with mean standard age scores for pupils who were not eligible for FSM, in the same way that the government has reported the attainment gap for disadvantaged pupils at the end of key stage 2 and key stage 4¹. In the design used in year 2, schools only selected pupils that they felt were disadvantaged. In this case, the gap was calculated as the difference between the mean standard age scores for the FSM pupils and the standard mean of 100.

The AfA trial ran over the full two years of the scheme. In this trial, the intervention group began implementing AfA in the first year, while the control schools carried on as normal. In the second year, the control schools began implementing the first year of the AfA programme, while the intervention schools moved into the second year of the programme.

¹ [Link to social mobility indicators](#)

The bars in the figures show the change in the attainment gap. A positive score indicates that the attainment gap has been reduced for pupils eligible for FSM. A negative score indicates that the attainment gap has increased.

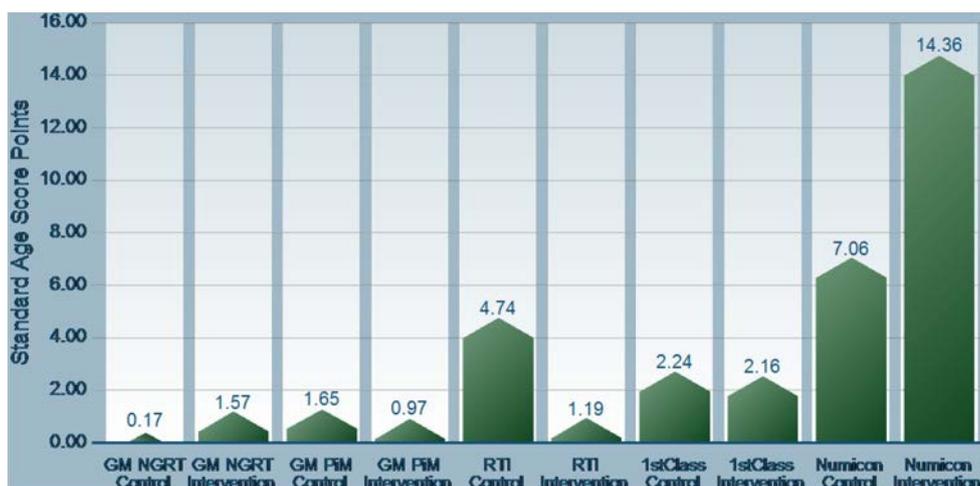
Figure 3.2.1: SAS points reduction in attainment gaps for the year 1 trials and RLS (trialled for the first time in year 2 of the programme) – control and intervention



Figure 3.2.2: SAS points reduction in attainment gaps for AfA (years 1 and 2) – control and intervention



Figure 3.2.3: SAS points reduction in attainment gaps for the year 2 replicated trials – control and intervention



3.3 Limitations and important considerations when interpreting the results

All experimental research has limitations. With regard to the large-scale trials within the closing the gap: test and learn programme, these limitations were mainly the product of cluster randomisation and internal validity issues arising from the collaborative school-led approach which reduced the degree to which extraneous variables could be controlled for. There were also limitations with regard to some of the trials caused by the fact that, to some degree, the use of standardised literacy and numeracy tests (such as when used with growth mindsets and AfA) could be considered to have reduced the design's sensitivity in detecting changes caused by these interventions. Neither growth mindsets nor AfA claim to directly affect mathematics or literacy attainment. At the same time RLS is not a direct literacy approach but rather one which can be tailored for such use. Similarly, RTI is not a direct pedagogical intervention but rather a flexible targeting approach. This meant that although teachers had a choice between targeting the reading or writing skills of their pupils, it was not possible to match the assessment to reflect this level of teacher in-class usage.

It is also important to acknowledge the wide variation in the nature of the interventions which had in some cases very different focuses, breadth and emphases with regard to target pupils. In this respect, the fact that a teacher-identified sub-group of target pupils (in the year 1 trials) lacked reliability and therefore the main analysis reverted to the analysis of effects on the FSM sub-group may have affected some interventions such as RTI. It is also important to remember that some of the interventions had a long delivery history (such as NIP) and have had a far longer time to become established and improve through feedback and revision. In contrast, for example, RLS was a new training programme which was only piloted the previous year, with 20 of the participating schools.

4 Findings from the ‘early adopter’ strand

4.1 The effectiveness of teacher-led randomised controlled trials and other forms of experimental research

Across all of the micro-enquiries, the following was noted:

- 96 per cent yielded a positive effect and only 4 per cent a negative one
- the average effect size, when converted to d for meta-analysis, was $d = 0.53$
- 23 per cent of the studies yielded large positive effect sizes
- 55 per cent of the results were significant (despite the relatively small sample sizes).

The effectiveness of the research designs was almost certainly the result of the teachers using tighter, more controlled designs over a shorter period – designs which are more akin to laboratory psychological studies. In addition, the tests chosen (or developed) by the teachers were generally more closely related to the area of study. In most cases teachers chose things to test which were directly related to a specific school improvement challenge within their context and developed interventions which (based on the local knowledge and understanding) they believed to be likely to work.

Another indication of the effectiveness of the approach was the variety and creativity of the types of research design that teachers deployed and made use of. Across the 48 studies that were completed in time for presentation at the dissemination event, there were examples of just about all of the commonest forms of experimental research used in fields such as psychology.

Several implications emerge from the findings described above. The approach has much potential as a means to develop teachers’ scientific literacy so that they can more easily engage in and develop an understanding of this type of evidence as the number of RCTs increases.

A second implication and potential of the approach becomes clear when one considers the strong role of context within the education evidence debate. Teacher-led RCTs have much potential as a means of mediating and exploring the effects of prior large-scale studies in the teacher’s own context, so that external solutions can be appraised according to local circumstance and priorities for spending. By extension, such approaches could also become a powerful way for schools to pilot changes in pedagogy before they are rolled out across a whole school or teaching school alliance (TSA).

Finally, not only are teacher-led RCTs desirable professionally, they could also be very cost effective, particularly if teachers can collaborate across schools to build larger sample sizes. For example, one study led by Kyra TSA pooled 11 classes within 10 primary schools, enabling a sample size of 231 pupils.

5 Conclusions

5.1 Summary of programme findings from all the areas of delivery and engagement (including qualitative data from teacher surveys during the life of the programme)

Pooling the evidence from the trials, OUDE research, focus groups and engagement activities, a number of overall conclusions can be drawn:

- the programme has clearly demonstrated the capacity of schools to engage in research through large-scale multi-arm trials and micro-RCTs; this also increased engagement with research and discussion of research findings
- contrary to assumptions made at the start of the programme, schools were not resistant to the use of control groups, engagement in statistical research or to the use of RCTs in general
- teachers can take a more active role in the delivery of RCTs. However, this form of approach requires investment in training and careful control of the communication structure and engagement protocols to ensure that individual trial sites do not become too distant from any middle-tier process used to build scale
- teachers can design and implement teacher-led RCTs focused on local school improvement questions and the evaluation of interventions to determine their effectiveness in individual contexts, paralleling approaches in healthcare
- surprisingly, despite their relatively small scale, teacher-led RCTs frequently produce statistically significant findings – almost certainly the result of shorter treatment windows, tighter controls and the ability to use research designs such as within-subject and matched-pair designs – designs which inherently increase the power to detect an effect
- the teaching profession already has a deep and available resource of individuals with experience of quantitative methods and their application outside of education research (such as science teachers and psychology teachers). The challenge going forward will be to connect these people with one another and provide them with models of how such approaches can be taken forward in school-led education research projects
- the results can be seen as emphasising the challenge of diagnosis within education – in that it is hard to find general things which improve outcomes across the board (except for NIP). Schools therefore need to think carefully about how their adoption and embedding of research-based approaches will stand more chance of success ('be above average'). Effective use of large-scale evidence is likely to require good diagnosis and targeting – and training of school leaders to understand how to do this

- micro-trials build capacity for research knowledge and engagement with research (the content of research findings) as well as engagement in research (trials design and implementation) which will in turn build capacity for the critique of the robustness of research findings. The evidence from this programme suggests that further development would be a valuable policy proposition

5.2 Policy implications and recommendations

Drawing on the above, the following policy implications and recommendations can be made:

- collaborative large-scale trial approaches such as the one applied in the closing the gap: test and learn programme have much potential with regard to reducing the costs of large-scale trials and thus the evidence base. Such trials in the future should, however, implement the learning from this programme with regard to those areas which need tighter national administration to avoid internal validity issues (for example, data labelling (if schools are given responsibility for this)) and consistency in relation to intervention delivery
- ‘toolkit’ summaries of research (such as meta-analyses of effect size) are at a broad level effective to some degree, but these need to be tested at a micro level. Such an approach, if co-ordinated nationally, could build an evidence base to support implementation and understand variation related to context
- finding no effect increases the range of choices that can be made, as this suggests that the impact on learning outcomes is the same. This point needs to be emphasised in the development of scientific literacy as policy develops in this area
- as the volume of trial evidence increases, the need to develop teachers’ scientific literacy and understanding of experimental research methods, in order to interpret those findings accurately and appropriately, is becoming increasingly clear
- there is potential for a micro-enquiry approach in initial teacher training, supported by good research design and analysis skills, as this would also address trainees’ scientific skills about enquiry and statistical capability. With regard to the use of inferential tests, effect sizes and confidence intervals, such training could also improve the quality and efficacy of local school improvement processes



National College for
Teaching & Leadership

© Crown copyright 2016

Reference: DFE-RR500a

ISBN: 978-1-78105-560-1

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/3 or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at:
teaching.schools@education.gsi.gov.uk or www.gov.uk/nctl

This document is available for download at www.gov.uk/government/publications