Education Endowment Foundation

Hampshire Hundreds

Evaluation Report and Executive Summary

October 2014

Independent evaluators:

CENTRE *for* ECONOMIC PERFORMANCE

Centre for Economic Performance, London School of Economics

Education Research Group, London School of Economics

Professor Sandra McNally

The Education Endowment Foundation (EEF)

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

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

- Identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- Evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale;
- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

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

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

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

James Richardson
Senior Analyst
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
London
SW1P 4QP

p: 020 7802 1924
e: james.richardson@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the Evaluator

The project was independently evaluated by a team from the Centre for Economic Performance (CEP), London School of Economics, and the Education Research Group (ERG), Department of Social Policy, London School of Economics.

The impact evaluation was led by Professor Sandra McNally (Director of the Education and Skills Programme, CEP), with Ms Amy Challen and Dr Gill Wyness.

The process evaluation was led by Professor Anne West (Director of ERG), with Dr. Philip Noden.

Contact details:

Professor Sandra McNally
Centre for Economic Performance
London School of Economics
Houghton Street
London
WC2A 2AE

p: 020 7955 7673
e: s.mcnally1@lse.ac.uk

Dr Philip Noden,
Education Research Group,
London School of Economics,
Houghton Street
London
WC2A 2AE

e: p.noden@lse.ac.uk

Contents

Executive summary

The project

The Hampshire Hundreds project was a local authority led intervention which brought together lead teachers from Hampshire primary schools to provide them with evidence and support for effective teaching strategies to decrease the attainment gap between disadvantaged pupils[1] and their peers.

The intervention consisted of support to teachers to enable them to better understand the learning needs of their pupils, and to consider how to improve the quality of their teaching, in particular, their questioning and feedback.

The intervention

Participating schools were given a document that summarised some of the common characteristics of disadvantaged and vulnerable pupils (Appendix A). The schools were asked to identify pupils in Years 5 and 6 who showed those characteristics. They then nominated a cohort of these pupils for close supervision and tracking – the Hampshire Hundreds pupils.

The intervention was then facilitated by the project leader in Hampshire Local Authority and comprised three stages:

Stage 1 – Lead teachers attended meetings in clusters of approximately five schools, where they were shown a presentation on the nature of socio-economic disadvantage and the challenges to learning for such pupils.

Stage 2 – Lead teachers shared the presentation with colleagues in their schools and were asked to observe disadvantaged pupils in lessons using the characteristics of disadvantaged pupils resource.

Stage 3 – At a second meeting, lead teachers were invited to feed back on their observations and consider their school's responses to the needs of the target pupils. They were also provided with research material focusing on feedback and questioning (e.g Hattie, 2009; 2012). In particular, they were encouraged to consider how they could adapt Hattie's model of feedback and apply it in their lessons.

Schools were encouraged to design their own responses to the challenge of increasing the attainment of disadvantaged pupils, therefore approaches varied substantially across schools. Some teachers concluded that existing classroom processes did not work effectively for their vulnerable pupils and sought to make improvements. Some teachers used materials to enhance provision to all pupils, while others sought to specifically target Hampshire Hundreds cohort pupils.

The programme was delivered at a whole-class level in Years 5 and 6, so all pupils in the intervention group, including those not deemed disadvantaged, were taught by a teacher who had received the training and support. Although the intervention was aimed at disadvantaged (or 'vulnerable') pupils, all

[1] Pupils are deemed as disadvantaged if they possess one or more 'barriers to learning' – a standard set of criteria defined by Hampshire County Council. In brief, these are as follows: **1**. Being classified by any of the following criteria: eligible for free school meals (FSM), English spoken as a second language (EAL), in care (CiC) or from an ethnic minority. **2**. Having a barrier to learning, such as lacking equipment and resources for learning, or lacking self-esteem and self-confidence.

pupils in the intervention group, including those not deemed deprived, could effectively benefit from the intervention.

36 schools took part in the trial. 19 schools were allocated to the intervention group and 17 schools to a wait-list control group, which were offered the intervention at least two terms after the intervention schools.

The intervention lasted for approximately 8 months from December 2012 to July 2013.

Key conclusions
1. Hampshire Hundreds showed no significant impact on raising attainment for disadvantaged pupils during the evaluation period.
2. Schools were encouraged to respond to the training and support in a way that was most suitable for the context of the school.
3. This project illustrates that it is difficult to convert research evidence into effective action within schools. It also illustrates the importance of careful piloting of an intervention before attempting an impact evaluation.

What impact did it have?

The intervention showed an effect size on the disadvantaged pupils of 0.03. Similar effect sizes were found for children eligible for free school meals. This effect size is very small. However, the confidence intervals are very wide. As a result we are unable to say whether or not this intervention had an impact on months' progress. Additional analysis was also conducted using the Key Stage 2 assessments in Year 6. This also shows an indeterminate effect. Therefore it is not clear that the Hampshire Hundreds approach is an effective way of improving attainment outcomes for children.

Table 1 – Impact of Hampshire Hundreds Project

Group	No. of pupils	Effect size (95% confidence interval)*	Estimated months' progress	Is this finding statistically significant?*	Evidence strength**	Cost of approach***
Disadvantaged pupils	924	**0.03** [-0.09, 0.14]	+1[2]	No		£
Other pupils	1124	**-0.02** [-0.10, 0.08]	-1	No		£

[2] *Since this report was published, the conversion from effect size into months of additional progress has been slightly revised. If these results were reported using the new conversion, all headline measures would be reported as 0 months of additional progress rather than +1 or -1. See* ***here*** *for more details.*

Group	No. of pupils	Effect size (95% confidence interval)*	Estimated months' progress	Is this finding statistically significant?*	Evidence strength**	Cost of approach***
Free school meal pupils	436	**0.03** [-0.12, 0.17]	+1	No		

* Effect sizes with confidence intervals that pass through 0 are not 'statistically significant', suggesting that the difference may have occurred by chance.
** For more information about evidence ratings, see Appendix D in the main evaluation report. Evidence ratings are not provided for sub-group analyses, which will always be less secure than overall findings
*** For more information about cost ratings, see Appendix E in the main evaluation report.

Note: This table shows 'intention to treat' effects for subgroups of students, for the primary outcome measure (i.e. the combined InCAS reading and maths score). The outcome measure is standardised with mean 0 and standard deviation 1.

How secure is this finding?

The evaluation was set up as a randomised control trial to test the impact of the Hampshire Hundreds intervention in comparison to a 'business as usual' control group, with the local authority leading the training and overseeing the provision of the intervention. As the intervention was run by the local authority it is classified as an efficacy trial. Efficacy trials seek to test evaluations in the best possible conditions to see if they hold promise. They do not indicate the extent to which the intervention will be effective in all schools since the participating schools are selected from one area, and the programme is delivered by the developers.

Analysis was completed on an 'intention to treat' basis where schools were compared in the groups to which they were originally randomly assigned.

The primary outcome measure was disadvantaged pupils' progress in reading and maths (combined), as measured by InCAS (Interactive Computerised Assessment System) developed by the Centre for Evaluation and Monitoring (CEM) at Durham University.

37 schools were recruited into the trial. However, a high level of attrition occurred between the first and second round of testing, and resulted in only 14 treatment schools and 10 control schools completing the post-intervention testing. This significantly reduced the power of the trial. Both intervention and control schools dropped out (5 and 7 schools respectively). However, this does not appear to have led to bias in the experiment in that there did not appear to be any systematic differences between the schools that dropped out and those that remained in the treatment (see 'Pupil Characteristics' section on page 17).

The evaluation protocol estimated that it would only be possible to detect large effect sizes in a trial of this size. The extent of attrition made finding a statistically significant effect less likely. However, when we use administrative data to look at outcomes (where attrition is not an issue because we look at Key Stage 2 results), we also find results that are consistent with this analysis. This analysis is additional to that outlined in the evaluation protocol and the results are described below.

The intervention was conceived as an action research project in which participating schools would be free to shape their own delivery. The process evaluation suggested there was a mismatch between the action research approach and the evaluation design (as a randomised controlled trial) and these are discussed further in the concluding section of this report.

How much does it cost?

If not supported by the EEF, the school would have needed to pay for LA support time at a cost of £600 per day. An estimate suggests a time of 3.5 days per school (see 'Cost' section on page 26 for further detail) totalling £2100 for consultancy time per school (3.5 days per school @ £600 = £2100 per school).

The schools did not have to buy any additional resources but they did have to support the project work by additional time in school for staff to arrange testing and consider lesson observation evidence, to co-plan and collaborate on classroom application of resources, and to respond to data requests and other communications from the project management team. The amount of time each school allocated to this varied. A rough estimate would be a minimum of 5 days per school additional internal time. If we use the LA day rate for teacher supply, this is £190 per day or a minimum of £950 per school.

Introduction

Intervention

The intervention was facilitated by the project leader in Hampshire Local Authority. It focused on effective teaching strategies to help teachers improve teaching and outcomes for disadvantaged pupils (aged 9–11). The objective was to reduce the attainment gap between disadvantaged pupils and their peers. The intervention involved a facilitator working with staff from a number of schools, and providing supporting materials comprising the 'Hampshire Hundreds Handbook', to provide good quality teaching and support to disadvantaged pupils.

The intervention consisted of support to teachers to enable them to better understand the learning needs of their pupils, and to consider how to improve the quality of their teaching, in particular, their questioning and feedback.

Before the intervention began, teachers were provided with a grid (Appendix A), and for each child in their class they were asked to indicate if the child was eligible for free school meals, was in local authority care, came from an ethnic minority background, had English as an additional language, or had other characteristics which acted as a barrier to learning. They were also asked to say whether a series of statements was typical of the child. For each child, teachers were then asked to add together the number of positive responses to these questions (e.g. FSM, EAL, child thinking they do not fit in) to provide a total score, which was used to rank pupils within the class. This resulted in every pupil having a 'vulnerability score' of 0–20. On average, schools classified 44% of pupils (in Years 5 or 6) as having one or more disadvantage according to this index. In this report, we refer to pupils as 'disadvantaged' if they have at least one disadvantage according to this index (i.e. a score of between 1 and 20).

The intervention was then facilitated by the project leader in Hampshire Local Authority and comprised three stages:

Stage 1 – Lead teachers attended meetings in clusters of approximately five schools, where they were shown a presentation on the nature of socio-economic disadvantage and the challenges to learning for such pupils. They were also given a document that summarised some of the common characteristics for disadvantaged and vulnerable pupils (which the schools had previously used to assist them to identify their Hampshire Hundreds pupils).

Stage 2 – Lead teachers shared the presentation with colleagues in their schools and were asked to observe disadvantaged pupils in lessons using the characteristics of disadvantaged pupils resource.

Stage 3 – At a second meeting, lead teachers were invited to feed back on their observations and consider their school's responses to the needs of the target pupils. They were also provided with research material focusing on feedback and questioning (e.g Hattie, 2009; 2012). In particular, they were encouraged to consider how they could adapt Hattie's model of feedback and apply it in their lessons.

The programme was delivered to pupils within normal classes, so all pupils in the intervention group's classes (including those not deemed deprived) could receive the intervention. However, the intervention was aimed specifically at 'disadvantaged' pupils (and in particular those who were 'vulnerable learners').

Background evidence

Pupils from disadvantaged backgrounds continue to underperform relative to their peers in state schools in England.[3] In 2012/13, results indicate that only 40.9 per cent pupils receiving free school meals achieved five A*-C grade GCSEs including English and maths, compared to 67.8 per cent of all other pupils (Department for Education, 2014). Indeed, this is a long-running and well documented issue (Sutton Trust, 2009; Chowdry, et al, 2010)

There are a number of evidence-based ways of improving attainment. One promising approach involves using effective feedback and metacognitive strategies to increase the internal motivation of pupils. Intrinsic motivation has been shown to be a valid construct and is positively related to achievement, IQ and perception of competence, and inversely related to anxiety (Gottfried, 1990). Evidence shows that the demotivated and well-motivated groups of pupils can be identified using basic cognitive theories of motivation, and these theories can in turn be used to mediate such motivation – such as through effective teacher feedback (Lens and Decruyenaere, 1991).

Evaluation objectives

There were two main aims in the impact evaluation:

1. To assess whether the attainment (as measured by standardised test scores) of disadvantaged pupils in the intervention schools improved relative to that of similar pupils in the control schools;
2. To assess whether the attainment of other pupils (in the same class as the disadvantaged pupils) changed as a result of the intervention compared to similar pupils in the control schools.

The process evaluation element of the research had several aims:

- To provide a description and overview of the intervention and its implementation, providing a context in which the quantitative findings could be better understood;
- To give an account of variation in implementation between and within schools, and therefore of variation in the intervention being studied;
- To identify informants' views on the key mechanisms for change;
- To identify salient differences in modes of implementation including opportunities and challenges posed by different models.

Project team

The impact evaluation was led by Professor Sandra McNally (Director of the Education and Skills Programme, CEP), with Ms Amy Challen and Dr Gill Wyness. All researchers have previous experience of devising and implementing research experiments involving schools.

The process evaluation was led by Professor Anne West (Director of ERG), with Dr Philip Noden. They have extensive experience of conducting such research in schools. They have previously worked closely with CEP researchers in evaluating both the UK Resilience Programme (with Amy Challen) and Excellence in Cities (with Sandra McNally).

[3] It should of course also be noted that pupils from some minority ethnic groups outperform their white British counterparts.

Ethical review

The study was funded by the Education Endowment Foundation.

Eligible schools were provided with information about the project through a one-day conference and were required to sign up in writing, confirming that the school wished to be part of the project. The various consent forms for schools and parents are provided in Appendix C. The project developer took responsibility for this process.

Methodology

Design

The research design was a cluster randomised controlled trial, with the unit of allocation at the school level. This is because the intervention involved training teachers, and these teachers subsequently taught pupils in the whole class. It was deemed inappropriate to randomise classes, despite the additional power this would have provided, owing to concerns about sharing of best practice between class teachers (i.e. contamination of the treatment group in an experimental setting).

Pupils in treatment and control schools were tested in maths and reading before and after the intervention. Our main question was whether the attainment of pupils improved after the intervention in treatment relative to control schools.

The trial followed a wait-list control design, with control schools being offered the intervention at least two terms after the intervention schools. The wait-list design was intended to incentivise schools in the control group to complete the testing in the first year, despite not having received the treatment.

Hampshire Local Authority recruited schools to take part in the project. 36 out of a possible 65 primary schools were recruited from the launch conference. The recruitment of a relatively small number of schools made it more difficult to detect small effects (of statistical significance) in the randomised control trial (this is detailed in the evaluation protocol).

Eligibility

The aim was to recruit about 65 primary schools, with a reasonable number of deprived pupils, to take part in the project. The details of the recruitment strategy are described below. Participating schools were required to sign up in writing, confirming that the school wished to be part of the project. 39 schools signed up initially, with 2 dropping out before randomization took place.

The intervention aimed to change the way teachers delivered the standard school curriculum. Thus the intervention would not affect the curriculum being followed in the intervention schools, but rather how it was taught. Pupils in treatment schools (Years 5 and 6) received the intervention in the school year 2012-2013. The intervention was delivered to whole classes but was expected to differentially affect disadvantaged and non-disadvantaged pupils. All teachers who were scheduled to teach a Year 5 or Year 6 class in autumn 2012 or spring 2013 were involved in the project. Opt-out consent letters were sent to all parents of Years 5 and 6 pupils involved in the study (see Appendix C).

At the outset, the intervention group comprised 37 schools. However, two of these schools were effectively run by the same head teacher, and so these were treated as a single school and will be referred to as the same school throughout the remainder of this report.

Intervention

Having recruited schools to take part, the project was implemented in three stages. In the first stage, a cohort of disadvantaged pupils was identified in each participating school. This took place in June and July of 2012. The second stage involved meetings organised for clusters of about five schools which were attended by key teachers from participating schools. It was conceived as a one-year intervention but delays in pre-tests meant that the cluster meetings did not begin until November and December 2012. At these they were given a presentation on the characteristics of disadvantaged pupils and the

dimensions on which effective schools could take action. Participants were then invited to share that presentation with colleagues.

Participants were then set several tasks. Two tasks involved providing data to the project leader as the facilitator of the action research – to provide firstly attainment targets for the Hampshire Hundreds cohort and secondly Hampshire Hundreds' pupils' responses to an adapted version of the SEAL (Social and Emotional Aspects of Learning) questionnaire. As direct learning opportunities for the project participants, they were asked to arrange to observe disadvantaged pupils in class while another teacher was teaching the lesson. Finally, they were encouraged to read research reviews (including Desforges and Abouchaar, 2003).

At the second meeting, in the first half of the spring term, they were invited to feed back on their observations and, crucially, to consider their school's response to the needs of the target pupils. They were also provided with research material drawn primarily from the work of John Hattie, focusing on feedback and questioning (e.g Hattie, 2009; 2012). In particular, they were encouraged to consider how they could adapt Hattie's model of feedback and apply it in their lessons. In the third phase of the intervention, the project leader visited participating schools to find out how they had developed their practice regarding disadvantaged pupils.

It is important to note that the teaching resources were not specifically geared to disadvantaged pupils; rather, they focused on important aspects of pedagogy which could be specifically directed towards disadvantaged pupils. However, while the materials may be viewed as generic, the project leader intended that the experiences of target pupils be identifiably different from those of other class members.

However, as the intervention was conceived as an action research project in which participating schools would be free to shape their own responses, and also the resource materials would be used within a normal classroom context, the intervention was sometimes described as a 'whole class' intervention and sometimes as involving 'targeted interventions for a small number of disadvantaged pupils'.

From the outset, the project leader did not prescribe whether specific teaching interventions, such as higher level questioning, should be targeted on particular pupils within a whole class context, or whether participants should aim to improve teaching for all pupils (and therefore benefit target pupils), or whether schools should provide other specific additional support to target pupils.

In control schools, since there was a wait-list for the intervention, no form of intervention took place – that is, they undertook 'business as usual'.

Outcomes

Sample size

36 schools took part in the experiment with 12 schools withdrawing before the time of taking the second test (5 treatment schools and 7 control schools).

A power calculation was used to calculate likely effect sizes for a sample size of about 36 which was an achievable sample given practical and budgetary constraints. The following is an extract from the evaluation protocol:

We performed the power calculations for the trial using Optimal Design Software (Raudenbush et al., 2011), based on a cluster size of 20 'targeted' disadvantaged pupils per school (assuming 5 per class; 2 classes per year group; 2 year groups per school). We estimated the intra-class correlation coefficient (ICC) as 0.1-0.2 based on similar tests in similar contexts (e.g. Spybrook et al., 2008, find this is typically between 0.05 and 0.15 for US data sets on school achievement). Based on the baseline tests having a predictive power of 0.80 on the same tests one year later, and assuming high consent and retention

rates of pupils (>90%) and an ICC of 0.2, an effect size of about 0.25 SD is detectable with 80per cent power and 5per cent two-sided alpha with 42 clusters (schools). With an ICC of 0.15 we would need 38 schools. Thus the trial is only powered to detect effects of at least 0.25 SD.

Minimum detectable effect size (MDES)

We can perform a similar analysis based on parameters in our data. On average there were 40 disadvantaged students per school. The ICC was 0.03. The baseline test had a predictive power of about 0.82. We could use 24 clusters (schools) when estimating regressions using our primary outcome. For these parameters, the effect size would have had to be in the region of 0.20 SD to be able to estimate a statistically significant effect.

Randomisation

36 schools were randomised. Within this, 5 schools formed federations of schools, which were grouped together to make 2 clusters (1 containing 2 schools and 1 containing 3 schools). Schools were paired on the basis of Key Stage 2 test results in 2012, with the 2 clusters also paired. Randomisation was conducted within pairs of schools, using random number generators in STATA.[4] The randomisation was performed by the team at the Centre for Economic Performance. Once randomisation took place, in November 2012, the results of the randomisation were fed back to Hampshire Local Authority's project leader who then informed those schools that were selected to form the treatment group that they would be receiving the teacher training in the 2012-2013 school year. In addition, those schools selected to form the comparison group were informed by the project leader that they would be receiving the intervention in phase 2 (i.e. after the end of the trial, in the academic year 2013-14).

Analysis

We estimate the following equation for each group *g* (i.e the disadvantaged pupils; the other pupils within the class):

$$Y_{igst} = \beta_1 T_s + \beta_2 Y_{igst-1} + \varepsilon_{igst}$$

where Y is the test outcome for person i in group g and school s at time t (and Yigst-1 is their test outcome in the baseline), T is whether the school gets the treatment and ε is a random error term. All analyses took into account the clustered structure of the data (condition assignment at the school level). Analyses of secondary outcomes followed the same pattern. Standard errors were clustered at the school level.

It is important to look at the outcome of the programme on both targeted and non-targeted pupils because both groups' attainment may be affected by it. For example, the change in teaching style could also benefit non-targeted pupils. Alternatively, the focus on disadvantaged pupils could take the class teacher's attention away from other groups of students. The ranking of pupils within each class allowed us to look at any such differences in impact.

[4] Two of the schools were combined after randomisation because we subsequently discovered they had the same head teacher. They had both been assigned to a control group (but in different pairs).

Thus we estimated the 'intention to treat' effect on disadvantaged pupils (i.e. those with a non-zero vulnerability score) and other pupils. We also report regressions for those pupils by free school meal status, ability score (i.e. whether pupil scored above the median on the baseline developed ability score) and gender.

Process evaluation methods

Initial fieldwork for the process evaluation focused on the background to the project. This addressed an intervention at a secondary school on which the Hampshire Hundreds project was based (an intervention on which the project leader had worked previously). The main qualitative fieldwork then focused on 10 case study schools in the project group and two visits to schools in the control group. Face to face, semi-structured interviews with teachers comprised the main form of data collection, supplemented by observations and documentary evidence. In total, 18 visits were made to schools (including the 10 case study schools referred to in the process evaluation report) and 33 teacher interviews were carried out. Some interviews were recorded but interviews were not transcribed in full as the cost of full transcription was judged to outweigh the benefits. In addition, Philip Noden, who carried out all the qualitative fieldwork, observed the launch conference, a project cluster meeting, a project school's pupil observation session and two follow-up visits made to schools by the project leader.

Case study schools were selected to ensure variation in both the proportion of disadvantaged pupils attending schools and the quality of schools as judged by Ofsted, and to ensure that schools from each of the Hampshire Hundreds clusters were visited.

Fieldwork was spread out from the launch conference in May 2012 until the end of the intervention in July 2013. Data analysis and the draft process evaluation report were completed before the impact evaluation analysis was carried out.

Impact evaluation

Timeline

April–September 2012

- CEP, ERG, Hampshire Council: Planned a detailed methodology and protocol. Hampshire devised a method of indexing pupils according to disadvantage, and circulated a template to CEP.
- Hampshire Council: Recruitment of schools and development of handbook.
- Schools: Indexed pupils in each class according to their deprivation in line with Hampshire template (Appendix A).
- CEP, Hampshire Council: Hampshire provided CEP with data on students in all participating schools prior to randomisation. This included the pupil names, their Unique Pupil Numbers (for matching with the National Pupil Database) and the index constructed as part of the ranking process. CEP entered, cleaned and checked the index data. CEP liaised with the Department for Education to match these pupils to the National Pupil Database.
- ERG: Visited Quilley School of Engineering.
- ERG: Observation of launch conference in May 2012.

September–November 2012

- CEP, Hampshire Council: Organised for tests to take place in schools Software problems led to delays in some schools so data collection planned for September was not completed until half-term.

November 2012

- CEP: After the baseline tests had been conducted in all schools, evaluators at CEP randomised schools into treatment and control groups. They then informed the Hampshire developer about their status.

December 2012–July 2013

- Hampshire Council: Worked with schools to apply strategies.
- ERG: Visited case study schools.
- CEP, Hampshire Council: Organised for final test to take place in schools (before schools broke up for the summer holidays). Test data was uploaded to the CEM website. Hampshire LA was provided with a list of schools that had uploaded data. However this list contained all schools who had ever uploaded their data (in either the first or second round of testing), leading to the mistaken impression that all 36 schools had uploaded their data.

July 2013 – December 2013

- CEP liaised with testing company to download test data. At this point only 18 schools had uploaded their data for the second round of testing. Hampshire urged the other schools to perform the second wave of testing. Several schools agreed to re-test, though others were unwilling or unable to do so, and this was further compounded by technical problems with the CEM software which meant that some schools who had completed their testing found themselves unable to upload their test data, or that it was lost.
- By end of December 2013, 24 of the original 36 schools had uploaded their second test data. The others were considered to have withdrawn from the project. Thus, while some of the schools completed their post-intervention testing in the summer term (as planned), for other schools the tests took place in the autumn term of the following academic year (2013/14).

January 2014

- CEP, ERG: Quantitative data analysis and report writing.

Impact evaluation results

The primary outcome measure for the evaluation was the change in attainment of disadvantaged pupils on the InCAS standardised computer-adaptive test (combined reading and maths). Secondary outcome measures assessed impact on mathematics and reading separately, the impact on other pupils, and heterogeneity of impact by pupil characteristics (subgroup analysis). We also looked at outcomes using Key Stage 2 tests in English and Maths (pupils in Year 6 only) although an analysis using these outcomes was not specified in the original protocol.

The InCAS tests were used because it is a well-established computer-adaptive assessment tool for schools that can be administered at any time. The test provider, CEM (University of Durham), is one of the UK's leading test providers, with a wealth of experience in testing pupils at a young age. Being computer adaptive the tests differentiate particularly for pupils at the lower end of the attainment spectrum which was important for this evaluation. The testing was intended to take place at the beginning and end of the same school year for all students in Years 5 and 6. Other forms of tests did not have this flexibility with regard to timing and administrative tests are taken only by students in Year 6. The InCAS tests also contain more measures of academic attainment than the older PIPS tests (for instance, they include spelling tests as well as reading tests).

The test results were standardized by InCAS at a national level, with a mean score of 100 and a standard deviation of 15. More details of the tests and their content can be found in the InCAS Feedback Guide (see http://www.cem.org/incas/feedback, or contact enquiries@cem.dur.ac.uk).

An issue that was not anticipated at the time of project design was that primary schools found it difficult to administer the tests due to technical barriers (having insufficient IT) within the schools. This delayed the start of the project in schools and may have been a reason why many schools did not proceed with the second round of testing.

Participants

Recruitment strategy

Schools were recruited into the project by Hampshire Local Authority. 78 primary schools in Hampshire (those with reasonable numbers of deprived pupils) were invited by the developer to send staff to a one-day conference about promoting the attainment of deprived pupils, which also introduced the project and explained that some schools would be chosen to receive the intervention in the first year, and some in subsequent years. In total, 65 schools expressed an interest in participating and staff from 62 schools attended the conference. Participants were provided with information about the project; if interested, they were asked to sign up in writing, confirming that the school wished to be part of the project (see Appendix C for consent letter). 39 schools signed up initially, with 2 dropping out before randomization took place. The main reasons for not signing up appeared to be not having sufficient numbers of FSM pupils or not wishing to increase teachers' workload. In addition, some were put off by not knowing exactly the size of the commitment they would be signing up to.

Attrition

2 schools dropped out after baseline testing but before randomisation, so 37 schools remained (though, as previously mentioned 2 of these schools were effectively run by the same head teacher, and so were treated as a single school, bringing the total to 36 schools). These were randomised into treatment and control groups by CEP. As discussed above, between the first and second survey, a further 12 schools dropped out of the project leaving a final sample of 24 schools. The most plausible reason for drop-out is that the testing process was onerous and the intervention was not intensive. Even within schools where the tests were completed, the percentage of pupils who did not undertake the second test was 20 per cent and 25 per cent for maths and reading respectively. A smaller percentage did not undertake

the baseline test (6 per cent and 7 per cent for reading and maths respectively). In cases where we did not have the test result for the second wave, we used baseline results and included missing variable dummies for affected pupils.

We know the Key Stage 2 results for students in Year 6. This allows us to estimate the intention to treat effect for all Year 6 students in the original sample. This was not part of the evaluation protocol. The analysis showed that our conclusions were unchanged if we estimated regressions using this sample.

Full details of the pupil vulnerability questionnaire can be found in Appendix A.

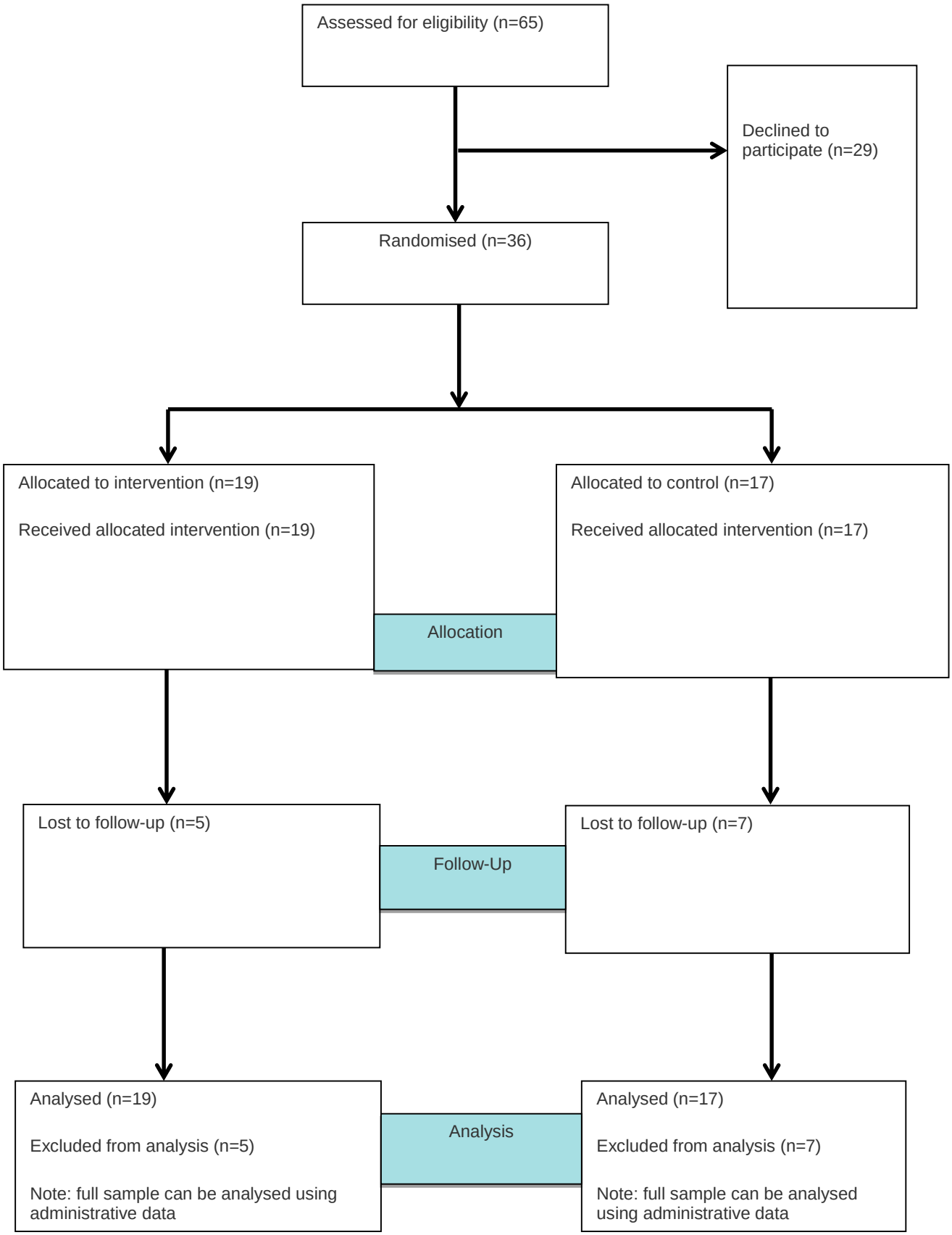
All pupils were effectively indexed in terms of the total sum of barriers to learning (which is defined as the total number of barriers each child is reported to have by their class teacher). This could vary between 0 and 20. On average, schools classified 44 per cent of pupils (in Years 5 or 6) as having one or more disadvantage according to this index.

Using this information we were interested in estimating the impact of the treatment on the disadvantaged sub-population of students and other students in their classes.

The project's launch conference took place in May 2012. Randomisation of schools into treatment and control groups took place in November 2012. Hampshire Local Authority worked with the schools from December 2012 until July 2013. The first round of testing was carried out in September and October 2012. Post-programme testing was carried out between July and December 2013.

Hampshire Local Authority was responsible for the delivery of the programme, with a team led by a Teaching & Learning Adviser, Hampshire Inspection and Advisory Service (HIAS).

Figure 1: Participant flow diagram



School characteristics

In Table 1, we provide summary statistics for schools involved in this study. They are all mixed primary or junior schools. Schools have similar characteristics by treatment status.

Table 2 Means of treatment and comparison schools (all 36 schools)

	Treatment	Control	Difference, t–c
Number of FSM pupils	23.15	22.77	0.377
	(9.31)	(5.80)	(2.63)
Pupil numbers	229.05	267.53	-38.48
	(69.93)	(95.77)	(28.40)
% Boys	50.9	51.6	-0.77
	(3.32)	(2.57)	(1.04)
OfSTED rating: outstanding	5.3%	5.9%	
OfSTED rating: good	52.6%	35.3%	
OfSTED rating: satisfactory	42.1%	47%	
OfSTED rating: unavailable	--	12%	

Note: Means and standard deviations (in brackets)

Pupil characteristics

In Tables 2 and 3 we provide summary statistics of pupil characteristics, Key Stage 1 test results (i.e. taken at age 7), and baseline standardised InCAS test scores of our initial sample (of 36 schools), testing for differences between the treatment and control groups at baseline (with standard errors clustered at school level). We do this for the full sample of 36 schools (Table 2) and for the sample of schools that completed both rounds of testing and are therefore viable for the analysis (Table 3).

Table 2: Means of treatment and comparison pupils (all 36 schools)

	Treatment	Control	Difference, t–c (for outcome variables, this is the 'effect size' at baseline)
Fsm	0.231	0.208	0.023
	(0.422)	(0.406)	(0.031)
Male	0.507	0.513	-0.006
	(0.5)	(0.5)	(0.018)
non-white	0.106	0.06	0.046
	(0.307)	(0.237)	(0.031)
index of disadvantage	2.471	2.588	-0.116
	(3.822)	(4.025)	(0.44)
KS1: reading	15.498	15.321	0.177
	(4.22)	(4.371)	(0.379)
KS1: writing	14.037	14.063	-0.026
	(3.52)	(3.885)	(0.394)
KS1: maths	15.495	15.291	0.204
	(3.534)	(3.7)	(0.356)
KS1: science	15.776	15.585	0.191
	(3.323)	(3.653)	(0.372)
InCAS reading	93.784	94.219	-0.435
	(15.972)	(16.567)	(1.324)
InCAS maths	90.685	91.127	-0.442
	(16.564)	(18.051)	(1.366)
InCAS developed ability	100.131	100.298	-0.168
	(16.456)	(16.431)	(1.184)
InCAS reading and maths	-0.017	0.017	-0.033
(restandardized to mean 0, sd=1)	(0.966)	(1.034)	(0.084)
InCAS reading and maths: vulnerable students	-0.365	-0.325	-0.039
	(0.903)	(1.013)	(0.071)
InCAS reading and maths: FSM students	-0.327	-0.480	0.153
	(0.945)	(1.020)	(0.107)

Notes: Means and standard deviations (in brackets) are calculated at pupil level. In column 3 (difference between treatment and control pupils), standard errors are in parenthesis. These are clustered at school level.
Index of disadvantage varies from 0-20 (with 20 being most disadvantaged).
KS1 points scores taken from National Pupil Database
InCAS scores taken at baseline and are standardised at national level, with mean 100 and standard deviation 15.
*significant at 10% level **significant at 5% level ***significant at 1% level (where no stars, there are no significant differences)

As can be seen in Table 2, the initial sample was well matched in terms of the characteristics of treatment and control schools at baseline. There were no significant differences in characteristics, suggesting the randomisation was successful.

Looking at Table 3, which comprises the set of schools which carried out both rounds of testing, it is interesting to see that the sample of treatment and control schools were also well balanced. This shows that attrition did not lead to any imbalance in the observable characteristics of treatment and control schools at baseline.

Table 3: Means of treatment and comparison pupils (final sample of 24 schools)

	Treatment	Control	difference (t-c) [for outcome variables, this is the 'effect size' at baseline]
Fsm	0.215	0.235	-0.02
	(0.411)	(0.424)	(0.022)
Male	0.51	0.519	-0.01
	(0.5)	(0.5)	(0.019)
non-white	0.123	0.065	0.059
	(0.329)	(0.246)	(0.039)
index of disadvantage	2.501	3.203	-0.702
	(3.847)	(4.321)	(0.551)
KS1: reading	15.759	15.024	0.735
	(4.204)	(4.553)	(0.447)
KS1: writing	14.294	13.767	0.528
	(3.523)	(3.959)	(0.438)
KS1: maths	15.715	15.076	0.639
	(3.526)	(3.861)	(0.404)
KS1: science	15.846	15.33	0.516
	(3.324)	(3.791)	(0.448)
InCAS reading	94.953	92.942	2.011
	(15.671)	(17.002)	(1.217)*
InCAS maths	91.217	90.016	1.201
	(16.634)	(17.951)	(1.067)
InCAS developed ability	100.707	99.828	0.879
	(16.767)	(16.642)	(1.497)
InCAS reading and maths (restandardized to mean 0, sd=1)	0.034	-0.054	0.089
	(0.961)	(1.050)	(0.071)
InCAS reading and maths: vulnerable students	-0.328	-0.317	-0.039
	(0.903)	(1.037)	(0.071)
InCAS reading and maths: FSM students	-0.313	-0.529	0.217
	(0.961)	(1.045)	(0.138)

Notes: As for Table 2.

In Tables 4 and 5 we illustrate the 'index of vulnerability' score as specified by Hampshire LA – again, of all 36 randomised schools (Table 3) and then of the sample of 24 schools that undertook both the baseline and post-intervention testing. In each case we show the mean disadvantage (i.e. the proportion of pupils fitting into each category) and whether this differed for the treatment and control schools (again with standard errors clustered at school level). Again, the pupils appeared to be well balanced between the treatment and control schools for both the full sample and the sample of 24 schools completing the final tests.

Table 4: Index of deprivation (all 36 schools)

	Treatment	Control	Difference, t–c
Lack equipment and resources for learning	0.108	0.12	-0.012
	(0.311)	(0.325)	(0.029)
Rarely take part in school trip activities	0.031	0.032	0
	(0.175)	(0.176)	(0.011)
Lack regular counsel, feedback, support, praise	0.101	0.11	-0.009
	(0.301)	(0.313)	(0.029)
Lack resilience and the emotional support that they need	0.145	0.177	-0.032
	(0.352)	(0.382)	(0.036)
Hide their emotions or true feelings	0.15	0.183	-0.033
	(0.357)	(0.387)	(0.036)
Have limited language repertoire	0.119	0.093	0.027
	(0.324)	(0.29)	(0.024)
Are less tolerant of passive approaches to teaching and learning	0.117	0.113	0.005
	(0.322)	(0.316)	(0.027)
Mistrust authority and are influenced by any negative attitudes	0.083	0.103	-0.019
	(0.276)	(0.304)	(0.022)
Place a ‘lid’ on their aspirations	0.14	0.12	0.02
	(0.348)	(0.325)	(0.027)
Areconcerned about the here and now	0.123	0.101	0.022
	(0.328)	(0.301)	(0.023)
Display symptoms of physical hardship	0.103	0.111	-0.008
	(0.304)	(0.314)	(0.02)
Think they do not fit in	0.066	0.084	-0.017
	(0.249)	(0.277)	(0.018)
Seem to be resentful or alienated	0.066	0.071	-0.005
	(0.249)	(0.257)	(0.017)
Think they are misunderstood	0.079	0.118	-0.039
	(0.27)	(0.323)	(0.024)
Lack self-esteem and confidence	0.219	0.237	-0.018
	(0.414)	(0.425)	(0.036)
Are reluctant to ask for help	0.138	0.172	-0.034
	(0.345)	(0.377)	(0.031)
Are evasive and slow to put themselves ‘on the line’	0.145	0.137	0.008
	(0.352)	(0.344)	(0.033)
Feel that they have little to contribute to the school	0.068	0.063	0.005
	(0.251)	(0.243)	(0.015)
Try too hard to be like everyone else	0.049	0.081	-0.032
	(0.216)	(0.273)	(0.019)
Rebel or try to be deliberately different as a means of defence	0.072	0.07	0.002
	(0.259)	(0.256)	(0.018)

Notes: As for Table 2.

Table 5: Index of deprivation (final sample of 24 schools)

	Treatment	Control	Difference, t–c
Lack equipment and resources for learning	0.101	0.156	-0.056
	(0.301)	(0.363)	(0.042)
Rarely take part in school trip activities	0.03	0.039	-0.009
	(0.17)	(0.193)	(0.015)
Lack regular counsel, feedback, support, praise	0.107	0.147	-0.041
	(0.309)	(0.355)	(0.041)
Lack resilience and the emotional support that they need	0.145	0.21	-0.065
	(0.352)	(0.407)	(0.05)
Hide their emotions or true feelings	0.151	0.234	-0.082
	(0.359)	(0.423)	(0.047)*
Have limited language repertoire	0.112	0.128	-0.016
	(0.316)	(0.334)	(0.029)
Are less tolerant of passive approaches to teaching and learning	0.131	0.153	-0.022
	(0.338)	(0.36)	(0.032)
Mistrust authority and are influenced by any negative attitudes	0.081	0.13	-0.049
	(0.273)	(0.337)	(0.029)*
Place a ‘lid’ on their aspirations	0.145	0.146	-0.002
	(0.352)	(0.354)	(0.033)
Are concerned about the here and now	0.123	0.12	0.002
	(0.328)	(0.325)	(0.026)
Display symptoms of physical hardship	0.104	0.132	-0.028
	(0.305)	(0.338)	(0.026)
Think they do not fit in	0.068	0.101	-0.033
	(0.251)	(0.301)	(0.022)
Seem to be resentful or alienated	0.067	0.085	-0.018
	(0.25)	(0.279)	(0.023)
Think they are misunderstood	0.078	0.152	-0.074
	(0.268)	(0.359)	(0.028)**
Lack self-esteem and confidence	0.232	0.276	-0.044
	(0.422)	(0.447)	(0.045)
Are reluctant to ask for help	0.15	0.204	-0.054
	(0.357)	(0.403)	(0.042)
Are evasive and slow to put themselves ‘on the line’	0.16	0.167	-0.007
	(0.367)	(0.373)	(0.043)
Feel that they have little to contribute to the school	0.074	0.073	0.002
	(0.263)	(0.26)	(0.019)
Try too hard to be like everyone else	0.051	0.101	-0.05
	(0.22)	(0.301)	(0.027)
Rebel or try to be deliberately different as a means of defence	0.064	0.088	-0.024
	(0.245)	(0.284)	(0.017)

Notes: As for Table 2.

Outcomes and analysis

We first report the intention to treat effect. For ease of interpretation, the outcome variables have been re-standardised to have a mean of zero and standard deviation of 1. Standard errors have been clustered at the school level. The intra-class correlation in baseline test scores is 0.03.

As discussed above, we estimated regressions only for schools where tests were conducted at baseline and for the post-test. However, for pupils in these schools who did not undertake the post-test, we used their baseline scores and included dummy variables for the fact that the score was missing for the post-test. We have also estimated regressions where these pupils were dropped from the analysis. In general this made little difference to estimates, apart from widening the standard error.

Results in Tables 6a and 6b report the intention to treat effect for pupils who were classified as disadvantaged and for other pupils in their classes (within Years 5 and 6). In both cases, the effect is imprecisely estimated and does not allow us to rule out positive or negative effects on the treatment group. However, the 95 per cent confidence interval would suggest that medium to large effects are unlikely (for example, the effect is within the range of -0.086 to 0.138 standard deviations for pupils who are classified as vulnerable). It makes little difference to estimates if we include control variables. This is shown in Appendix B (after including additional controls, the point estimate is also close to zero and the confidence interval is in the range -0.125 to 0.105). In the case of other (non-vulnerable) pupils, estimates are sensitive to including controls, with the point estimate becoming negative and significant at the 5 per cent level (although small in magnitude at -0.06 standard deviations).

It makes little difference to exclude the students for whom test scores are missing at Wave 2 (about 25 per cent. This just widens the confidence interval around small point estimates.

A similar story is revealed if we estimate regressions separately for Maths and Reading scores. We report results in Appendix B.

Table 6a: Maths and reading scores, intention to treat effect for disadvantaged pupils

(ie pupils with a non-zero score on the vulnerability index)

	Effect size	Standard error	95% CI
Treatment	0.026	0.054	[-0.086, 0.138]
Baseline score	0.909	0.015	[0.879, 0.939]

Notes: number of observations=924. Standard errors clustered by school (24). This regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
The dependent variable has mean zero and standard deviation of 1.

Table 6b: Maths and reading scores, intention to treat effect for other pupils

(ie pupils with a zero score on the vulnerability index)

	Effect size	Standard error	95% CI
Treatment	-0.012	0.044	[-0.103, 0.080]
Baseline score	0.926	0.010	[0.906, 0.946]

Notes: number of observations=1124. Standard errors clustered by school (24). This regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
The dependent variable has mean zero and standard deviation of 1.

In Table 7, we report results for the treatment dummy variable when the same regressions have been estimated for separate subgroups: according to gender, ethnicity (i.e. whether non-white), whether eligible to receive free school meals, and whether the person is above the median score for developed ability (a measure of IQ) in the baseline. In every case, the point estimate is zero with a wide confidence interval. Once again, we are unable to say whether the treatment has a positive or negative effect, although we can rule out medium to large positive effects.

Table 7: Maths and reading scores, intention to treat effect for separate subgroups

	Effect size on treatment dummy	Standard error	95% CI
(1) Males	0.009	0.051	[-0.096, 0.115]
(2) Females	0.010	0.044	[-0.081, 0.101]
(3) FSM	0.025	0.069	[-0.118, 0.169]
(4) Non-FSM	0.003	0.043	[-0.086, 0.091]
(5) White	0.004	0.045	[-0.090, 0.098]
(6) Non-white	0.013	0.041	[-0.071, 0.098]
(7) Above median ability	0.021	0.045	[-0.073, 0.115]
(8) Below median ability	0.013	0.047	[-0.085, 0.112]

Notes: Each row comes from a separate regression. Standard errors clustered by school (24). Regressions control for baseline score, whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading. The dependent variable has mean zero and standard deviation of 1.

Finally, we estimate the same regressions for the sample of students in Year 6 for whom we have Key Stage 2 scores for reading and maths. These are normalised in the national sample to have a mean score of zero and a standard deviation of 1. In this case, all 36 schools are included in the analysis. We estimate regressions for all students (regardless of whether or not they are designated as vulnerable pupils). Results are reported in Tables 8a and 8b for reading and maths respectively. Again, our point estimates are close to zero with a large standard deviation. We cannot rule out positive or negative effects. This is true if we include detailed controls (including KS1 attainment, Appendix B, Tables 4a, 4b) or if we estimate regressions for disadvantaged and other students separately, or if we estimate regressions for different subgroups (Appendix B, Tables 5a, 5b).

Table 8a: KS2 Reading (standardised score), intention to treat effect

	Effect size	Standard error	95% CI
Treatment	-0.038	0.085	[-0.211, 0.135]
Baseline score	0.750	0.031	[0.687, 0.813]

Notes: number of observations=1347. Standard errors clustered by school (36). This regression also controls for a missing variable dummy for the combined baseline score of maths and reading.

Table 8b: KS2 Maths (standardised score), intention to treat effect

	Effect size	Standard error	95% CI
Treatment	-0.006	0.084	[-0.177, 0.164]
Baseline score	0.787	0.024	[0.739, 0.836]

Notes: number of observations=1351. Standard errors clustered by school (36). This regression also controls for a missing variable dummy for the combined baseline score of maths and reading.

Cost

The intervention consisted of support to teachers: to enable them to better understand the learning needs of their pupils, to consider research evidence and training materials, and to adapt their practice. If not supported by the EEF, the school would have needed to pay for LA support time at a cost of £600 per day. Based on a rough estimate, the time involved is as follows:

- 1 day for initial planning and set-up
- Half a day working with key staff and establishing some baseline tests
- Half a day with the key staff in school who had undertaken the baseline tests, submission of data and classroom observations on pupils, listening to their findings and planning next steps
- Half a day after a two-month trial of the resource materials and the actions agreed in the last meeting to follow-up.
- At the end of this the schools were invited to join together to discuss findings – a further day.

So in total – 3.5 days per school @ £600 = £2100 for consultancy time per school.

In addition to this, all schools attended an initial conference to explain the key expectations of the project and its intended outcomes. This was free to schools but would have cost each school £200 to attend if it had not been covered by general project funding. Each school would have had to fund supply cover for 2 members of teaching staff who attended (one day in each case) – a further £380 per school.

The schools did not have to buy any additional resources but they did have to support the project work by additional time in school for the colleagues involved to consider the test and lesson observation evidence, to co-plan and collaborate on classroom application of resources and findings and time to respond to data requests and other communications from the project management team. The amount of time each school allocated to this varied. The more successful schools put a lot of time into the activities and provided ample opportunity for staff to experiment and reflect. A rough estimate would be a minimum of 5 days per school additional internal time. To calculate the cost of this it might be helpful to use the LA day rate for teacher supply costs of £190 per day – so 5 x £190 = £950 per school.

If we add together all these estimated costs, the notional cost per school is £3,630 per treatment school and £580 per control school. The total cost is then (£3,630 x 19) + (£580 x 17) = £59,830.

Process evaluation

A premise for the project, which is well supported by research evidence, was that pupils are not equally well prepared for school (in part because of pupils' social circumstances) and classroom practice may systematically reinforce that disadvantage (eg George et al., 2007; Strand, 1999). The project leader's view was that vulnerable children do not ordinarily answer questions in class and so fall behind whereas those from advantaged backgrounds are familiar with the type of questioning.

The project leader therefore aimed to ensure that teachers made a greater effort to involve disadvantaged pupils and engage them in learning. Fundamentally the project aimed to renew the pupils' relationship with the teacher and refocus that relationship on learning.

We might thus schematically think of the project as responding to the fact that different groups of pupils enter school with unequal educational resources. The schooling process may then fail to compensate for these differences and thus an equal schooling process may lead to unequal outcomes. Alternatively, the schooling process may have unequal effects on different groups of learners and potentially exacerbate pre-existing differences.

Pupil selection and baseline testing

Before the intervention began it was necessary to select pupils for the project, carry out baseline testing and then select an intervention and control group of schools. Although these processes were not part of the project intervention, they were not inconsequential and so are included in the discussion of the process evaluation findings.

Primary schools in Hampshire were invited to send representatives to a launch conference in May 2012 and most of those attending signed up to participate in the project.

Schools then took different approaches to the identification of a target cohort of pupils. In particular, some schools identified only FSM eligible pupils to be part of the cohort (FSM pupils being the target group for EEF interventions). At some schools all FSM pupils were included as Hampshire Hundreds pupils. In other schools teachers were guided by the results of the Pupil Vulnerability Questionnaire (Appendix A). Teachers were aware that FSM, as an indicator of low income, is not a measure of educational vulnerability. Teachers frequently observed that pupils belonging to particular disadvantaged groups were not necessarily the most educationally vulnerable.

Indeed, in one school a larger proportion of FSM pupils achieved the Year 6 benchmark than non-FSM pupils in the previous year, and at another school there was no difference between the two groups on this measure. In short, in some schools, teachers did not perceive the Hampshire Hundreds cohort of pupils to be the pupils most in need of educational support and in some schools the attainment of FSM pupils matched or exceeded that of non-FSM pupils (on the government's headline attainment measure).

The testing process

The testing process was more complex than expected, requiring the downloading of software to individual computers (rather than being web-based), and took much longer than had been anticipated. Substantial problems arose in schools using a particular operating system for their IT systems.

At some schools the testing process was relatively unproblematic, for example described by one teacher as:

Time-consuming but not a disaster for the school… (Teacher, school 4)

At most schools, however, the testing process was described in much more negative terms as, for example, a 'nightmare' and 'hugely disruptive'. Problems related in particular to timetabling the use of school ICT facilities and to login details not working for substantial numbers of pupils. In addition, difficulties were reported relating to the test being timed and so, having begun the test, being unable to interrupt; it not being possible for teachers to try the test so they could describe the process to their pupils; headphones not working; and the test taking different lengths of time for different pupils and so teachers were faced with pupils who unexpectedly had nothing to do. Owing to delays in the baseline testing process, randomisation did not take place until November, 24 weeks after the launch conference and several months behind the original schedule.

The intervention itself began just before Christmas 2012. The project leader saw the intervention as an action research project. As such, it was intended that participants (teachers) would take ownership of the implementation of the project, identifying their own responses, suitable to their own schools, to meet the educational needs of disadvantaged learners.

However, participants in the project often saw this flexibility as a lack of clarity about the direction of the project and its intended stages. This was one of the strongest messages communicated by participants.[5]

Implementation

Pupil observations

As part of Stage 1 of the project, teachers were encouraged to carry out pupil observations.

The opportunity for teachers to observe their pupils could be instructive. Pupil observations were carried out in most, although not all, of the schools visited.

One teacher reported undertaking a very informative observation. She observed a Hampshire Hundreds cohort pupil with English as an additional language. She described watching the pupil during a science lesson during which the teacher referred to a 'sieve'. This was an unfamiliar word to the pupil and so he had asked a classmate twice to explain what the word meant but was 'shooshed' by the classmate each time he asked. The pupil then gave up trying to find out and sat quietly. This observation led the teacher to explain very clearly to the pupil that it was always acceptable to ask the teacher to explain a word that he did not understand.

While observations were instructive in some of the project schools, they were not undertaken in all schools and in some cases were unproductive. One teacher explained that she had only had 10 minutes to observe the target children and during that time had seen no behaviour which would differentiate the Hampshire Hundreds children from other members of the class. She explained:

[5] The original plan was for the intervention to be supported by a Hampshire Hundreds Toolkit which was to be distributed at the launch conference. This would have identified the anticipated focus areas for the intervention. However, it was essential for the evaluation to have a wait-list control group which was uncontaminated by the intervention and so a 'toolkit' was not distributed as planned. Of course the evaluation did not prevent a toolkit from being provided to participants after randomisation. However, under the revised plan the project leader chose to provide teachers with materials supporting the project in several stages rather than at the outset. Thus the evaluation did necessitate a revision of the intervention plan which in turn led to a less clearly communicated intervention.

> When those children were highlighted and we were looking at these characteristics it was quite easy to put them all in one group and say – this is what you might see in them. It was quite easy to label and say you may see these characteristics in these children. And actually we didn't. (Teacher)

She went on to suggest that the picture of educational vulnerability presented in relation to pupils who had received free school meals was 'a bit of a stereotype'.

Nevertheless, in some schools the process of identifying vulnerable learners and observing them in class had prompted what were seen as constructive discussions which attempted to get beyond seeing FSM as a cause of educational underachievement and instead to look at the specific barriers to learning faced by some disadvantaged pupils.

In short, some teachers observed the normal classroom process as being less effective for disadvantaged learners. However, this conclusion was not drawn at all of the schools.

Feedback and questioning

Following the second Hampshire Hundreds cluster meeting participants were encouraged to consider how they could apply a model of feedback to their classroom practice. The model was drawn from the work of John Hattie, differentiating between levels of feedback. Once again this was not taken on at all participating schools. For example, some felt they had already made substantial progress in developing effective feedback at the school.

At one school the materials were used to enhance the quality of marking carried out at the school (which was already seen as exemplary by the local authority). A teacher used the resource pack on feedback and produced materials to assist teachers in further enhancing their feedback, converting the materials into ready-to-use, teacher-friendly, language. Teachers at the school had then explained to pupils that in future they would be expecting more from the children when a piece of work had been marked and thus they might comment, 'you have successfully completed this piece of writing. What, from the success criteria, have you done well?' or for a mathematics task 'can you repeat this process using decimals?' Thus the project had, in the view of these teachers, contributed to an on-going process of school improvement.

Pupil conferencing

Participants were encouraged to initiate discussions with their target pupils, in the first instance taking feedback as a topic for conversation. In some schools such pupil 'conferencing' had been carried out and additional release time provided to teachers to facilitate this. Some teachers were able to report on the additional knowledge about vulnerable pupils that they gained through this process. For example, in some instances pupils had told their teachers of their reticence to ask questions and so one class teacher reported that she was consciously moving around the classroom during the lesson to ensure that she could ask particular children if they had any questions. In another a pupil had explained that he did not want to put his hand up to answer questions because he did not want appear to be 'sucking up' to the teacher.

In addition, in one school Hampshire Hundreds children had met as a group, along with their class teacher, for additional discussions during the lunch break. The teacher felt this had made them feel a little bit more special and enabled her to initiate conversations which she imagined for many other pupils might take place routinely at home.

In another school the project had contributed to an experiment with random questioning, choosing children to answer questions by drawing lolly sticks (with pupils' names on them) from a pot. This was introduced to ensure that teachers' questions were not always directed to a small number of pupils who might always put up their hands to contribute an answer. While the teacher was very enthusiastic about the effect of this innovation, such a practice was already in operation in several of the other project

schools (including, for example, putting three sticks into the pot with the same person's name on them to increase the chance of particular pupils being asked questions).

Conversely, in response to pupil observations the use of random questioning was reduced in one school because a teacher had seen how frustrated one disadvantaged pupil became when he was unable to show that he knew the answer to a question.

Individual school responses

Overall there was substantial variation between case study schools not only in how they responded to the project but also in the level of activity generated in response to the project.

In a school judged as outstanding by Ofsted, the project was described as having directed the focus onto vulnerable children, given a structure to review and implement change, and provided a network through which to share practice. The project was also seen to have bred other ideas with the launch of a summer school for FSM and vulnerable children for the last four days of the summer holidays so they were able to meet their teacher. As part of the school's response to the project, key questions were identified on lesson plans for use with target pupils, and Hampshire Hundreds children were named on lesson plans.

Another school, which had responded the most vigorously to the project's objectives, had introduced the most active targeting of pupils, with relentless targeting of questions on Hampshire Hundreds pupils and intensive work to develop parental input into their children's education. Unlike at other schools, parents of these target pupils were invited to school to discuss the project.

In contrast, among the least responsive schools were two schools which had reason to believe they were already effective in meeting the needs of disadvantaged pupils with, for example, positive value added scores, relatively high levels of progress among disadvantaged pupils and 'green' responses for all relevant categories reported in raise-online. At both schools teachers reported that they were always on the lookout for new techniques to improve the attainment of disadvantaged pupils and had mistakenly expected the Hampshire Hundreds to provide an array of activities which they could try.

Fidelity

It is difficult to assess the fidelity of implementation because the project leader set out to enable schools to develop different responses to improve the achievement of Hampshire Hundreds pupils. It is however possible to make some general observations relating to programme fidelity.

The project sought to deliver elements of the support previously delivered to another school in Hampshire which had seen dramatic improvements in attainment. However, the intervention differed for the Hampshire Hundreds primary schools compared with the pilot secondary school in several key ways. Perhaps most notably, the project leader had much less intensive involvement with the Hampshire Hundreds project schools than with the original school. Whereas she had been able to establish a very highly valued relationship with senior leaders at the original school developed over numerous meetings, this was not possible across the large number of Hampshire Hundreds schools. The project leader's background was in secondary school teaching rather than primary schools and this may have had an additional effect. Notably in one school where specific training had been provided to staff by the project leader, an interviewee spoke of how constructive meetings had been. However other interviewees commented on having had insufficient contact with the project leader. It is also important to note that the Hampshire Hundreds project sought to replicate only some aspects of the pilot intervention. For example, the academic characteristics of pupils were not included as selection criteria for inclusion in the cohort whereas in the original school pupils were also selected on the basis of unfulfilled academic potential. Thus there were several notable differences between the Hampshire Hundreds Project and the pilot project.

It is also important to note that schools varied substantially in their manner of implementation of several of the project activities. For example, one teacher reported that the school had been unable to arrange for teachers to observe classes so had instead had a discussion of the things they believed they might have seen. Indeed, overall schools varied noticeably in the extent to which the project was resourced within schools. For example, some schools arranged lesson cover to enable pupil observations or conferencing to take place while, for others, the latter might take place during assembly or lunchtime. One respondent commented that had she known what was required to support the project it would have been useful to set aside in-service training days to enable staff to undertake discussions relating to the project.

Additionality

In the pilot school the intervention was presented as a central feature of a narrative of school improvement. However, in almost all the Hampshire Hundreds primary schools visited, the project was presented as, at most, contributing to improvements already being made within the school rather than as a driving force. That is, at no school was the project presented as a central feature of a story of school improvement.

Control group activity

As intended, the control group did not receive the Hampshire Hundreds intervention during 2012/13. Visits to control group schools provided examples of highly individualised approaches to teaching and learning, including support for disadvantaged learners. The process evaluation did not indicate any resentfulness or rivalry between the intervention and control groups.

Formative findings

The intervention comprised an action research project in which the project leader acted as facilitator and in which variation in school responses was encouraged because the project leader wished schools to take ownership of their own plans for action. The start of the intervention was delayed owing to the requirements of the project evaluation. In addition, the intervention was not communicated clearly from the outset in order to ensure the control group of schools was not contaminated. As an action research project, the intervention then began with further data collection and, owing to the desire to accommodate heterogeneous responses, a clear plan for the intervention was not communicated. In combination, this led to project participants feeling there was a lack of clarity and direction in the project. Thus there were various interactions between the evaluation design and the project itself from which valuable lessons may be learned.

In drawing together the variation in the responses of schools, their orientation to the selection of pupils for the project, and how far they incorporated the project into a story of school improvement, it is possible to summarise schematically the pathways taken by the case study schools and a plausible model is shown in Figure 2.

Schools varied in how far they took an active or relatively inactive response to the project. Schools also varied in how far they took ownership of the intervention's target group or viewed the selected group as a legitimate target group. That is, in some schools a group of pupils had been identified who teachers felt were in need of particular consideration and support. In other schools a cohort was selected which was not felt to contain the most vulnerable pupils.

In their response to the activities initiated by the Hampshire Hundreds project, schools also varied considerably. Some participants observed that classroom processes did not work effectively for their

vulnerable pupils whereas other participants did not identify such shortcomings in their existing processes.

Participants were also provided with some stimulus materials relating to pupil feedback and questioning. As generic interventions such materials could be used to enhance provision to all pupils or attempts could be made to specifically target Hampshire Hundreds cohort pupils through, for example, targeted questioning.

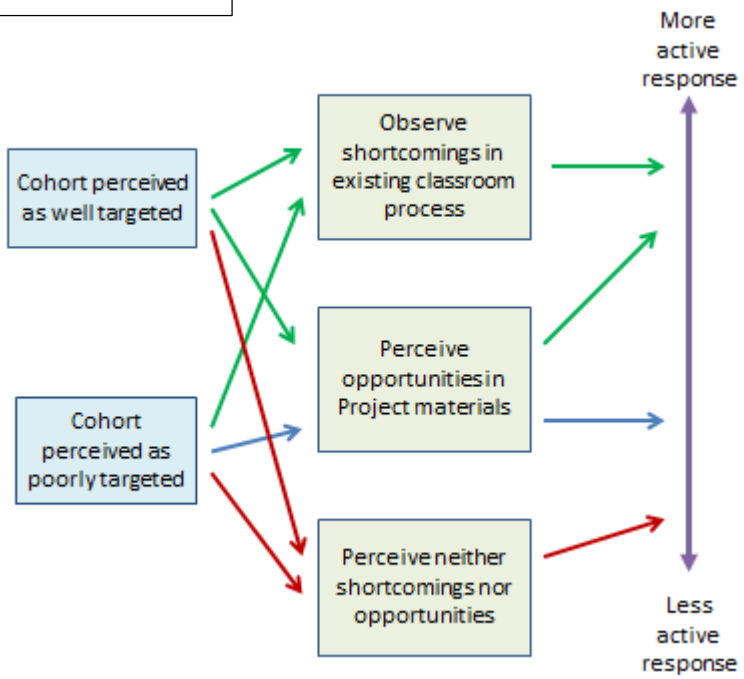
Drawing together the responses of case study schools to these selection, observation and intervention phases it is possible to provide a plausible account of the circumstances which led to a relatively active or less active response.

The green lines shown in Figure 2 represent schools which made an active response to the project. For schools which undertook pupil observations and where it was noted that school processes did not work as effectively for disadvantaged pupils, action was taken to make that process more effective. At such schools active responses were made regardless of whether the school felt ownership of the selection process identifying the cohort for the Hampshire Hundreds project.

The other green route through Figure 2 was traced by a school at which the cohort had been carefully selected as a group in need of intervention. The school's active response comprised striving to target improvements in pupil questioning specifically on the target cohort.

At schools following the blue pathway through Figure 2, participants did not perceive a need to modify their existing practice following observations (because either no observations were carried out or no shortcomings were identified). In addition they did not feel ownership of the identification of the target cohort. The materials provided by the Hampshire Hundreds project were therefore incorporated into generalised 'whole school' processes. Finally, the red lines were traced by schools which neither identified shortcomings in their practice through pupil observations, nor saw the project materials as offering an opportunity to improve lesson planning or pupil feedback. For these schools, regardless of whether they felt the selected cohort was in particular need of intervention, an inactive response was made to the project.

Figure 3: Diagram description



Conclusion

Limitations

It is important to note that the conceptualisation of the project as an action research project did not fit well with the randomised controlled trial design for the project evaluation in several ways. First, to increase the chances of detecting an effect, the evaluation required the number of schools participating to be increased. This consequently reduced the amount of time which the project leader could devote to each individual school. Second, the need for the control group not to be contaminated by the intervention meant that a 'Hampshire Hundreds toolkit' was not distributed at the launch conference as originally planned. The project leader instead modified the project plan and released a 'Hampshire Hundreds Handbook' in stages during the year. This undoubtedly contributed to participants' uncertainty about the process and endpoint of the intervention. In addition, the lack of clarity about the intervention, as described at the launch conference, may have contributed to some potential participants being put off taking part. Third, in order for the evaluation to compare the impact of the intervention on different groups of pupils, a questionnaire was developed to identify vulnerable learners. However, this made visible (and problematic) the inadequacy of FSM eligibility as an indicator of educational vulnerability at the individual level. Fourth, the baseline testing process was demanding and caused delay which created a loss of momentum for the project. Twenty-four weeks elapsed between the launch conference and randomisation to identify an intervention group and control group. The delay caused by the testing process, and the time devoted to it by schools, was the largest problem encountered in the implementation of the project and resulted in the attrition described earlier in the report. Fifth, as an action research project the first stage of the intervention was then to collect more data which further blurred the distinction between the intervention and its evaluation. Sixth, the randomised controlled trial design for the evaluation may also have contributed to participants' inaccurate expectations of a tightly conceived intervention.

Interpretation

The project leader of the Hampshire Hundreds sought to implement innovative ways of raising the attainment of disadvantaged pupils in primary schools. Although these methods were grounded in research evidence, the impact evaluation did not identify a significant effect on attainment. The confidence intervals are very wide, in part due to the small sample size and to attrition. Nevertheless they indicate that medium to large sized effects can be ruled out.

Future research

While there may be no reason in principle to exclude action research projects from evaluation through randomised controlled trials, this project illustrates the need to carefully consider the practical implementation of such an evaluation and in particular its possible effects on the intervention.

References

Cohen, J., (1988), Statistical Power Analysis for the Behavioral Science. Hillsdale, N.J: Lawerence Erlbaum. 2nd edition.

Department for Education (2014) GCSE and equivalent attainment by pupil characteristics in England, 2012/13, SFR 05/2014.

Desforges, C. and A. Abouchaar (2003) The impact of parental involvement, parental support and family education on pupil achievements and adjustment: a literature review, DfES RR433.

Duflo, E., R. Glennerster, M. Kremer, (2007) Using Randomization in Development Economics Research: A Toolkit, CEPR Discussion Papers 6059

http://ideas.repec.org/h/eee/devchp/5-61.html

Finn, J. D. and C.M.Achilles (1990) Answers and Questions about Class Size: A Statewide Experiment, *American Educational Research Journal*, 27, 3, 557-577.

George, A., Hansen, K. and I. Schoon (2007) Child behaviour and cognitive development, in Hansen K. and H. Joshi (eds) *Millenium Cohort Study Second Survey: a user's guide to initial findings,* London: Centre for Longitudinal Studies, Institute of Education, University of London

Gottfried, A, E. (1990) Academic intrinsic motivation in young elementary school children, *Journal of Educational Psychology*, 82(3), 525-538.

Hattie, J. (2009) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement,* Oxford: Routledge.

Hattie, J. (2012) *Visible learning for teachers: Maximizing impact on learning,* Oxford: Routledge.

Krueger, A. and D. Whitmore (2001) The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star, *Economic Journal,* 111, 34-63.

Lens, W.and M. Decruyenaere (1991) Motivation and demotivation in secondary education: Student characteristics. *Learning and Instruction*, 1, 145–159.

Raudenbush, S.W., Spybrook, J., Congdon, R., Liu, X., Martinez, A., Bloom, H.and C. Hill (2011) *Optimal Design Plus Empirical Evidence* (Version 3.0)

Spybrook, J., Raudenbush, S.W., Liu, X., Congdon, R. and A. Martinez (2008) Optimal Design for Longitudinal and Multilevel Research: Documentation for the 'Optimal Design' Software, http://sitemaker.umich.edu/group-based/optimal_design_software

Strand, S. (1999) Ethnic group, sex and economic disadvantage: associations with pupils' educational progress from baseline to the end of Key Stage 1*, British Educational Research Journal*, 25 (2), 179–202.

Appendix A: Worked example of pupil ranking table ('Deprivation index')

Pupils initials and year group	AB Y5	CD Y5	EF Y5	GH Y5	IJ Y5	KL Y5	MN Y5
Are they FSM, EAL, CiC, from an ethnic minority or have other barriers to learning?	FSM	CiC	None	Other	FSM, EAL	EM	FSM, EM
Do they?							
Lack equipment and resources for learning	√				√		√
Rarely take part in school trip activities							
Lack regular counsel, feedback, support, and praise for the personal skills they acquire					√		
Lack resilience and the emotional support that they need to overcome difficulties	√				√	√	√
Hide their emotions or true feelings	√	√			√		√
Have limited language repertoire and rely almost entirely on informal modes of talk	√				√		
Be less tolerant of passive approaches to teaching and learning	√						
Mistrust authority and are influenced by any negative attitudes to school displayed by others					√		
Place a 'lid' on their aspirations and adopt restricted and unambitious view of their future					√		√
Be concerned about the here and now of their experience in school and are unconvinced by the 'it will be good for you in the long term' type of argument					√	√	√
Display symptoms of physical hardship (e.g. listless through lack of sleep or an adequate breakfast, fail with homework	√				√		√

through lack of a suitable work space or materials)							
In addition, do they?							
Think they do not fit in		√			√	√	
Seem to be resentful or alienated				√			√
Think they are misunderstood				√	√		√
Lack self esteem and confidence							√
Be reluctant to ask for help	√	√		√	√	√	√
Be evasive and slow to put themselves 'on the line'				√	√		√
Feel that they have little to contribute to the school or the school has little to offer them		√			√	√	√
Try too hard to be like everyone else; pretend to be what they are not, in order to fit in					√		
Rebel or be deliberately different as a means of defence				√	√		√
Total	7	4	0	5	16	5	13
Group ranking **(Please allocate a rank order for each pupil based on the total number of ticks – highest total = 1)**	3	5	6	4	1	4	2

Appendix B

Table B.1a: Maths and reading scores, intention to treat effect for vulnerable pupils

(i.e. pupils with a zero score on the vulnerability index)

	Coefficient	Standard Error	95% CI
Treatment	-0.010	0.056	[-0.125, 0.105]
Baseline score	0.884	0.019***	[0.844, 0.923]
Male	-0.037	0.027	[-0.093, 0.020]
FSM	-0.004	0.031	[-0.068, 0.061]
Non-white	0.125	0.042***	[0.039, 0.213]
Above median ability	0.102	0.029***	[0.041, 0.162]

Notes: This is the same regression as Table 6a, with additional controls (for which we show coefficients and standard errors). In addition, the regression also controls for dummy variables indicating the pair grouping of the school (16 pairs) and a missing variable dummy for FSM status and developed ability at baseline. As for Table 6a, this regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.

The dependent variable has mean zero and standard deviation of 1. Number of observations=924. Standard errors clustered by school (24).

*significant at 10% level **significant at 5% level ***significant at 1% level (where no stars, there are no significant differences)

Table B.1b: Maths and reading scores, intention to treat effect for other pupils

(i.e. pupils with a zero score on the vulnerability index)

	Coefficient	Standard Error	95% CI
Treatment	-0.062	0.029**	[-0.123, 0.002]
Baseline score	0.894	0.015***	[0.862, 0.926]
Male	-0.045	0.025**	[-0.096, 0.007]
FSM	-0.010	0.036	[-0.085, 0.065]
Non-white	0.116	0.027***	[0.060, 0.172]
Above median ability	0.118	0.037***	[0.041, 0.197]

Notes: This is the same regression as Table 6a, with additional controls (for which we show coefficients and standard errors). The regression also controls for dummy variables indicating the pair grouping of the school (16 pairs) and a missing variable dummy for FSM status and developed ability at baseline. As for Table 6a, this regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.

The dependent variable has mean zero and standard deviation of 1. Number of observations=1124. Standard errors clustered by school (24).

*significant at 10% level **significant at 5% level ***significant at 1% level (where no stars, there are no significant differences)

Table B.2a: Reading score, intention to treat effect for vulnerable pupils targeted subgroup

(i.e. pupils with a non-zero score on the vulnerability index)

	Coefficient	Standard Error	95% CI
Treatment	0.047	0.063	[-0.083, 0.177]
Baseline score	0.882	0.021	[0.837, 0.926]

Notes: number of observations=901. Standard errors clustered by school (24). This regression also controls for whether whether wave 2 reading score is missing; and a missing variable dummy for the baseline score for reading.
The dependent variable has mean zero and standard deviation of 1.

Table B.2b: Reading scores, intention to treat effect for other pupils

(i.e. pupils with a zero score on the vulnerability index)

	Coefficient	Standard Error	95% CI
Treatment	0.054	0.081	[-0.114, 0.223]
Baseline score	0.848	0.025	[0.796, 0.900]

Notes: number of observations=1099. Standard errors clustered by school (24). This regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
The dependent variable has mean zero and standard deviation of 1.

Table B.3a: Maths scores, intention to treat effect for vulnerable pupils

(i.e. pupils with a non-zero score on the vulnerability index)

	Coefficient	Standard Error	95% CI
Treatment	0.019	0.047	[-0.077, 0.117]
Baseline score	0.896	0.017	[0.861, 0.930]

Notes: number of observations=863. Standard errors clustered by school (24). This regression also controls for whether whether wave 2 reading score is missing; and a missing variable dummy for the baseline score for reading.
The dependent variable has mean zero and standard deviation of 1.

Table B.3b: Maths scores, intention to treat effect for other pupils

(i.e. pupils with a zero score on the vulnerability index)

	Coefficient	Standard Error	95% CI
Treatment	0.014	0.046	[-0.082, 0.109]
Baseline score	0.894	0.014	[0.865, 0.923]

Notes: number of observations=1068. Standard errors clustered by school (24). This regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
The dependent variable has mean zero and standard deviation of 1.

Table B.4a: KS2 Reading (standardised score), intention to treat effect

	Effect size	Standard error	95% CI
Treatment	-0.065	0.052	[-0.170, 0.041]
Baseline score	0.438	0.033***	[0.370, 0.505]
KS1 average points score	0.100	0.012***	[0.075, 0.124]
Male	-0.057	0.028**	[-0.114, -0.001]
FSM	-0.009	0.042	[-0.096, 0.078]
Non-white	0.081	0.047*	[-0.016, 0.177]
Above median ability	0.291	0.051***	[0.188, 0.395]

Notes: number of observations=1347. Standard errors clustered by school (36).
This is the same regression as Table 8a, with additional controls (for which we show coefficients and standard errors). The regression also controls for dummy variables indicating the pair grouping of the school (16 pairs) and a missing variable dummy for FSM status and developed ability at baseline. This regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
*significant at 10% level **significant at 5% level ***significant at 1% level (where no stars, there are no significant differences)

Table B.4b: KS2 Maths (standardised score), intention to treat effect

	Effect size	Standard error	95% CI
Treatment	-0.049	0.044	[-0.137, 0.040]
Baseline score	0.462	0.030***	[0.400, 0.523]
KS1 average points score	0.116	0.010***	[0.095, 0.137]

Male	0.242	0.03***	[0.181, 0.303]
FSM	-0.057	0.051	[-0.162, 0.047]
Non-white	0.159	0.054***	[0.050, 0.268]
Above median ability	0.206	0.029***	[0.147, 0.263]

Notes: number of observations=1351. Standard errors clustered by school (36).
This is the same regression as Table 8b, with additional controls (for which we show coefficients and standard errors). The regression also controls for dummy variables indicating the pair grouping of the school (16 pairs) and a missing variable dummy for FSM status and developed ability at baseline. This regression also controls for whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
*significant at 10% level **significant at 5% level ***significant at 1% level (where no stars, there are no significant differences)

Table B.5a. KS2 Reading (standardised score), intention to treat effect for separate subgroups

	Effect size on treatment dummy	Standard error	95% CI
(1) Pupils classified as disadvantaged	-0.097	0.105	[-0.31, 0.115]
(2) All other pupils	0.013	0.936	[-0.177, 0.204]
(3) Males	-0.087	0.088	[-0.265, 0.091]
(4) Females	0.004	0.094	[-0.188, 0.197]
(5) FSM	-0.054	0.119	[-0.295, 0.186]
(6) Non-FSM	-0.029	0.084	[-0.20, 0.142]
(7) White	0.004	0.045	[-0.090, 0.098]
(8) Non-white	-0.115	0.124	[-0.372, 0.142]
(9) Above median ability	-0.035	0.087	[-0.212, 0.142]
(10) Below median ability	0.018	0.099	[-0.183, 0.220]

Notes: Each row comes from a separate regression. Standard errors clustered by school (36).
Regressions control for baseline score, whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
The dependent variable has mean zero and standard deviation of 1.

Table B.5b: KS2 Maths (standardised score), intention to treat effect for separate subgroups

	Effect size on treatment dummy	Standard error	95% CI
(1) Pupils classified as disadvantaged	0.026	0.119	[-0.215, 0.266]
(2) All other pupils	-0.029	0.084	[-0.201, 0.142]
(3) Males	-0.016	0.075	[-0.168, 0.137]
(4) Females	0.008	0.104	[-0.202, 0.218]
(5) FSM	-0.063	0.082	[-0.229, 0.103]
(6) Non-FSM	0.015	0.092	[-0.173, 0.203]
(7) White	-0.024	0.086	[-0.200, 0.152]
(8) Non-white	0.087	0.096	[-0.112, 0.286]
(9) Above median ability	0.013	0.096	[-0.182, 0.207]
(10) Below median ability	-0.011	0.098	[-0.211, 0.188]

Notes: Each row comes from a separate regression. Standard errors clustered by school (36).
Regressions control for baseline score, whether wave 2 maths score is missing; whether wave 2 reading score is missing; and a missing variable dummy for the combined baseline score of maths and reading.
The dependent variable has mean zero and standard deviation of 1.

Appendix C.1: Letter to headteachers

Dear

Hampshire Hundreds Research Project 2012-14

I am writing to you to update you on developments relating to the Hampshire Hundreds research project. Firstly, can I thank you for returning your data and pupil analyses at the end of the summer term, this has enabled a number of things to take place. Allison Head has put all of the pupil data into a large database; this has two parts – a general database for over 3000 pupils in Years 5 & 6 with their pupil characteristics and a second database that represents the identified cohort for the research intervention – this amount to 800 pupils that you have identified. As you can appreciate this has been some task but gives us some very useful information. All schools who have agreed to participate in the project and have returned their data will receive a small amount of funding. The actual amount is being agreed at our next Hampshire Hundreds Leadership Group meeting and details will follow in due course. It is hoped to send this funding out to schools before the end of October.

Over the summer I have analysed your responses to the FSM audit of practice and prepared a report based on my findings. This was an interesting exercise and has enabled me to consider other avenues of support that we can offer to schools. I wish to share the findings of this report with you all over the coming weeks and will be arranging meetings with Headteachers and project leaders shortly.

If you recall, one of the conditions of the research grant was to support the work of the external evaluators, the Centre of Economic Performance at the London School of Economics. One of the requirements is it that all Year 5 and Year 6 pupils in the primary schools are tested using the University of Durham CEM centre's InCAS test at the start of Year 1 and at the end. The object of this exercise is to see if the phase one schools improve the attainment and attitudes of pupils following the intervention. InCAS assesses maths, literacy, developed ability and attitudes through a computer-delivered assessment. The tests are computer-adaptive and quickly work out the level of the ability of each pupil completing the test and pose questions accordingly. The maths and literacy modules take 30 minutes each, the picture vocabulary, non-verbal ability and attitudes test takes 30 minutes also. The Education Endowment Foundation is funding the tests and we would like all pupils to take all of the tests but we suggest you break them up into 3 modules over a two-week period.

To make this process as easy as possible we have enlisted the support of the Hampshire IT team at Fareham. They have been working through the summer in consultation with the CEM centre to prepare the software for download and instructions for conducting the tests. We have also uploaded individual pupil data from our central school's database so that you can easily transfer this on to your systems. Attached with this letter are the step-by-step instructions of the processes involved. We have also trained a number of people in our IT offices to help you with any technical queries that might arise and have completed a dummy run in one primary school to ensure that we can support all your needs; our aim in so doing is to make this process as straightforward as possible. Contact details of IT help are included in the guidance notes. Each school has a unique user name and password; your details are attached.

We cannot proceed with the research work until all pupils in the Hampshire Hundreds primary schools have sat the tests and the results returned to the University of Durham. We now entreat you to download this software and prepare to run the tests in your school over the next 3 weeks. **We must have the results of all the tests from all of the schools submitted to Durham University by Friday 28th September**. If schools do not do this within this time frame then everyone will be held up. Following receipt and processing by Durham, the data sets will be sent to the LSE who will apply their randomisation procedures and allocate schools in to phase one and phase two cohorts.

We think it is important that pupils and parents are told about the tests and that the school has been invited to participate in the research project. To this end I attach a letter that I propose you share with parents; it gives them the right to exclude their child from the test if they choose to do so, but also

explains the significance of the work for their children and Hampshire schools. I am also happy to send this to you electronically if you wish to adapt it yourself. If a parent raises a query you may find the following weblink helpful: http://www.cemcentre.org/incas/introduction.

So in conclusion, the next steps in the Hampshire Hundreds timeline are as follows:

- Read the instructions for downloading the InCAS software on to your school system, and set up the tests.
- Send a letter to all parents of all pupils in Years 5 and 6 telling them about your school's involvement in the project and the nature of the test, offering the opportunity to withdraw their child if they wish.
- Run the tests at a time to suit your curriculum (the three modules do not have to be completed at the same time but all pupils need to sit each individual test at the same time)
- Ensure that all tests are completed within the three-week timeframe.
- **Save the results and send to the University of Durham as instructed,** using your unique school log in and password, **no later than 28th September.**

In due course, you will then receive a further communication from me detailing whether you are in phase one of the research or phase two, with a request to visit each phase one school as soon as possible.

I do hope this information is clear; if not please feel free to email me with any individual queries. I know that one or two of you have shared with us individual requirements regarding participation and these are noted on our database.

I look forward to working with you in the very near future,

Yours sincerely

Lead Adviser – Hampshire Hundreds Project

Enc: Hampshire IT – InCAS supplementary user guide
Allocated school username and password
Letter template for parents of pupils sitting the InCAS tests

Appendix C.2: Letter to parents

Dear Parent/Carer

Hampshire Hundreds Research Project 2012-14

Your child's school has been invited to take part in a large scale educational research project led by Hampshire County Council but funded by the Education Endowment Foundation, a national charity.

Children in school years 5 and 6 are the focus of this work and over coming months teachers in your child's school will receive support and guidance to enable them to understand the learning needs of all children better and to improve pupil achievement. There are over 40 primary schools across Hampshire involved in this pilot project and it is hoped that the findings of the research will eventually be rolled out to all Hampshire maintained schools.

To measure the effectiveness of the research and its outcomes, the London School of Economics has been appointed to externally evaluate the work. As part of the evaluation process, all pupils need to sit a series of short tests at the start of the research and at the end, to see if there is an impact. The tests are taken individually by all pupils on a computer as part of your child's normal lesson time over a two-week period; there is a test for literacy, mathematical ability and a more general test for wider abilities and attitude. The tests are pupil adaptive which means they quickly work out the ability of each child and then pose questions accordingly, making the tests stress-free, fun and easy to follow. Each test module lasts approximately 30 minutes. The results of the tests will not be shared with the pupils but will be shared with the school at the end of the research.

If you feel you do not wish your child to sit the tests, then you must tell your Headteacher as soon as possible so that they can be withdrawn. **If you wish to discuss this further – please contact your Headteacher directly**. If you are happy for your child to be involved with the rest of the class, then you do not have to respond to this letter.

Hampshire County Council is the only local authority in the country to have received a research grant from the Education Endowment Foundation for such work and is very fortunate to have this opportunity to improve learning and outcomes for pupils. We sincerely hope that you will support this endeavour and recognise the value to your child, your child's school and the local community.

Yours faithfully

Helen Fenton
Lead Adviser – Hampshire Hundreds Research Project

Appendix C.3: Parent consent form

Dear Parents,

????? School is very pleased to be working in collaboration with The University of Winchester on the evaluation of a county-wide initiative which has focused on supporting and improving children's learning and progress.

In order to do this one of the trainee teachers currently on the BA Primary Education programme at the University will be allocated to the school and undertake some discreet school-based research which will also contribute to the trainees' course programme. Your child's class teacher and I have agreed to support this research, which will involve working with the children in your child's class during the Spring term on Mondays between January and March. This will consist of observing children learning, some short group interviews and reviewing work.

Full details and information about the procedures used to collect and protect data are provided in the accompanying information sheet. All information, data and final reports will be fully anonymized, remain confidential and used only by the school and to contribute to the trainees' coursework.

As with all our work with trainees and visitors this will be monitored by the school and university staff and follow the university's procedures for any school-based research. Trainee teachers are checked by the DBS (Disclosure and Barring Service) and so are fully cleared to work with children in school.

If you require any additional information please contact me at the school where I will be able to help or direct you to the appropriate contact at the University .

I hope that after reading this information, you will feel able to support this research and give consent for your child's participation. If so I would be grateful if you would please complete and return the attached form to your child's class teacher by ???.

With thanks

Yours sincerely,

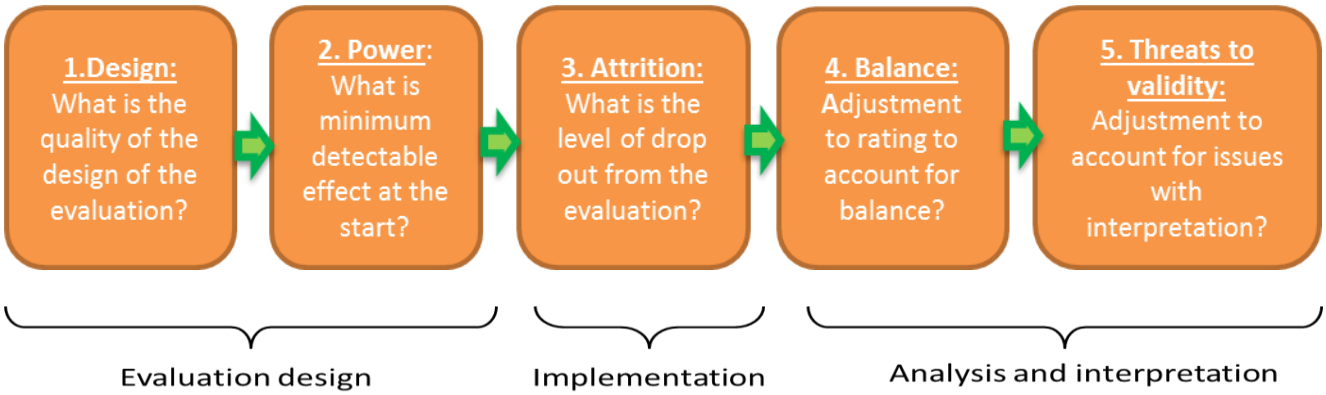
Head teacher

Dear ???

I give consent for my child …… (class ……) to participate in the University of Winchester research project between January and March 2014.

Signed…… Printed name……..
Parent/Carer

Appendix D: EEF security rating summary – Hampshire Hundreds



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

The final security rating for this trial is 2 🔒. This means that findings are of moderate security.

The trial was designed as a cluster randomised efficacy trial with the intention of recruiting 65 schools. Only 36 were achieved providing a minimum detectable effect of 0.18 at randomisation meaning the trial could still have achieved a maximum of 5 🔒. However, the trial experienced high attrition with 33% of schools dropping out. This did not result in significant imbalance at the baseline on pupil characteristics although there was some evidence of a small difference in prior attainment that was controlled for in the analysis. The main threat to the validity and interpretation of the results stems from the intervention being poorly defined and a large variation in delivery between schools. Also, although tests were digital and therefore marked blind, the schools were responsible for their delivery and not blind to treatment allocation.

Appendix E: Cost rating

Cost ratings are based on the approximate cost per pupil of implementing the intervention over one year. Cost ratings are awarded using the following criteria.

Cost	Description
£	*Very low:* less than £80 per pupil per year.
£ £	*Low:* up to about £170 per pupil per year.
£ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

This document is available for download at **www.educationendowmentfoundation.org.uk**.

Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk