**UCL Institute of Education Tutor Trust evaluation proposal**

## Team members

Dr. John Jerrim (Principal Investigator); Professor Dave Pratt (Mathematics process evaluation lead); Professor Gemma Moss (English process evaluation lead); Dr. Kirstine Hansen; Nikki Shure; Terry Ng-Knight; Dr. Cosette Crisan; Dr. Melissa Rodd

## Impact evaluation

*Summary*

The key elements of our evaluation of the Tutor Trust (TT) programme are as follows:

- Method = Randomised Controlled Trial (RCT) with an integrated Regression Discontinuity Design (RDD).
- Randomisation = School level (pending further discussions with TT and EEF).
- Number of participants = 100 schools and 1000 pupils
- Power = Minimum Detectable Effect Size (MDES) of approximately 0.25 (allowing for 15% attrition)
- Control condition = Active control. No financial incentives to be used.
- Allocation = Equal allocation between two arms.
- Testing = Key Stage 2 test scores used as the outcome measure. Key Stage 1 and Foundation Stage Profile used as the pre-tests. All pre and post test data drawn directly from the National Pupil Database.
- Primary outcomes = Impact upon average Key Stage 2 Maths and English scores as estimated by the RCT.
- Secondary outcomes = Impact upon KS2 Maths and English scores as estimated by the integrated RDD.

Further discussion around these choices can be found below.

*Level of randomisation*

Randomisation could occur at either the pupil, class or school level. We have discounted randomisation at the pupil level due to concerns over gaining ethical consent, possible 'contamination' between treatment and control groups, and potential teacher and parent objection to experimentation with individual children.

Class may offer more statistical power than school-level randomisation (assuming that 'streaming' of primary school pupils into different classes is minimal[1]). However, one-form entry schools would then not be eligible to participate in the study, and there would need to be acceptance of randomisation occurring within schools. We believe this may cause difficulties in terms of school recruitment, attrition post-randomisation, and possible non-compliance amongst parents.

Consequently, we suggest the trial should use school randomisation. Although not as statistically powerful as pupil or class level randomisation, we believe it is the most likely to be accepted by Tutor Trust, schools and parents. Moreover, our power calculation suggests that it will be possible to detect an effect size consistent with what has been found for other small group tuition programmes, so long as around 100 schools are recruited.

It is important to note that, despite our preference for school-level randomisation, our design could also be implemented with randomisation occurring at the class level. Consequently, the pros and cons of both should be discussed between the selected evaluators, EEF and TT before a final decision is made.

*Power calculation*

Table 1 reviews recent EEF evaluations of one-to-one and small group tuition programmes in England. Three of the four studies found an impact of around 0.25 standard deviations or below. Using this as a guide, we believe this trial should be powered to be able to detect an effect of at least 0.25 standard deviations.

**Table 1. Recent EEF studies of one-to-one or small group tuition programmes**

| | Evaluator | Effect size |
|---|---|---|
| Catch-up literacy | NFER | 0.12 |
| Catch-up numeracy | NFER | 0.21 |
| Switch-on reading | Gorard et al 2014 | 0.24 |
| One-to-one literacy coaching | NFER | 0.36 |

[1] Campbell (2014:10) suggests that around 17 percent of Key Stage 2 pupils are 'streamed'.

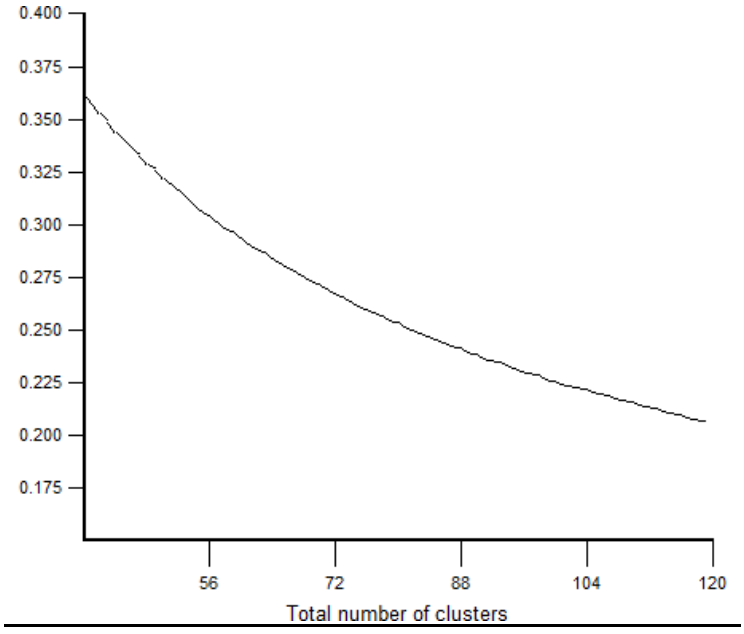
Figure 1 illustrates how the MDES varies with the number of clusters (schools) recruited into the trial. This graph has been produced under the following assumptions:

- Intra-cluster correlation (ICC) of 0.13. (This figure is based upon the Key Stage 2 mathematics scores of children within the EEF Chess in Schools trial).
- Number of children per cluster (school) = 10 pupils
- 80 percent power for a 95 percent confidence interval
- 45 percent of Key Stage 2 test scores explained by Key Stage 1 and Foundation Stage profile scores. (This figure is based upon the Chess in Schools trial).

A total of 85 clusters (schools) would be needed to achieve a minimum detectable effect size of 0.25. Allowing for 15 percent attrition, this suggests that 100 schools should be recruited into the trial. (Note that conducting a specific pre-test is unlikely to substantially increase the power of the study over and above what we propose. For instance, under the assumption that a pre-test would explain 65 percent of the variance in Key Stage 2 test scores, the MDES would only decline from 0.25 to 0.23).

We do have some concerns regarding Tutor Trust's capacity to recruit and deliver the intervention in this number of schools within a single academic year (and within the specified geographic area). For instance, the geographic area identified (Leeds Local Education Authority) only has 224 primary schools in total (http://www.education.gov.uk/edubase/search.xhtml?page=). This suggests TT would need to recruit almost half of all primary schools in Leeds LEA into the study. Consequently, we believe that contingency plans should be put in place to guard against possible under-recruitment of schools. This should compose of two elements. First, the possibility of recruiting schools from another LEA should be considered. Second, the possibility that the RCT gets extended over two academic years should also be contemplated. For instance, the RCT would involve two cohorts: cohort A of 50 schools in 2016/17 and cohort B of 50 schools in 2017/18. These are of course decisions that need to be made in conjunction with EEF and TT.

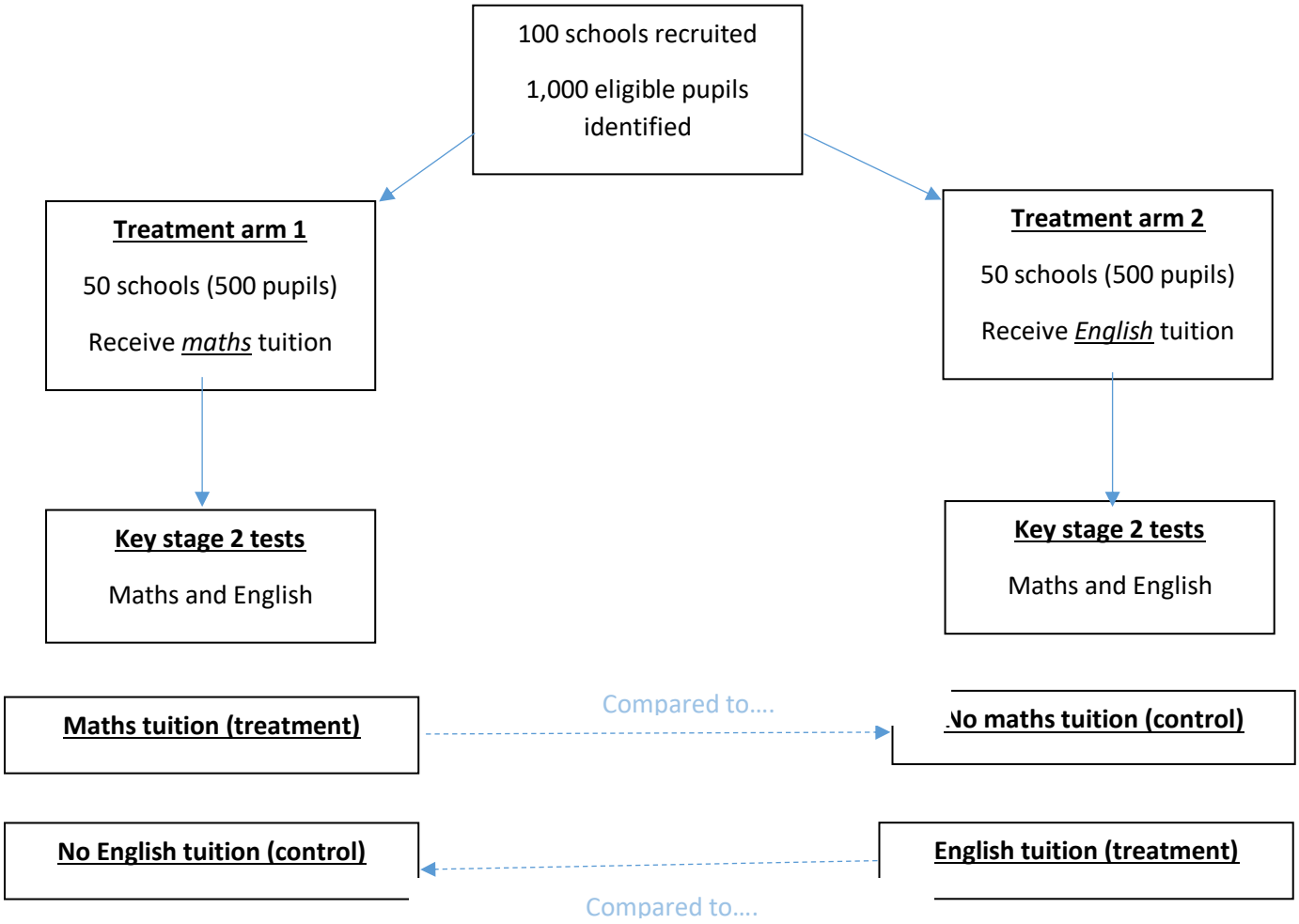
**Figure 1. Power calculation: MDES versus number of schools recruited**




*Allocation of schools and control condition:* The call for tenders suggested an 'inactive' control condition (i.e. business as usual) was being considered, in conjunction with a financial payment to encourage retention of control schools. Our advice is that an *active* control condition should be considered instead. First, the Department for Education generally advise against the use of financial incentives, citing them as ineffective and offering poor value for money. For instance, a payment of £500 to £1,000 per school in this trial would mean £25,000 to £50,000 added to the evaluation budget - with no guarantee that this would significantly improve the validity of the results. Second, we believe that there is a real danger an ethics committee could reject such a design; children at risk would have been identified, yet would not gain any direct benefit from being part of the trial. (Such ethical concerns are only likely to be heightened by the fact the EEF has previously found tuition programmes to have a positive impact upon children's test scores). Third, there could be substantial attrition from a trial with an inactive control, particularly amongst control schools who are not happy with their allocation. Finally, this type of intensive tuition programme has the potential to generate substantial 'Hawthorne effects' – particularly when a 'business as usual' control condition is used. In other words, a 'treatment effect' may be found simply due to the fact the treatment group is being closely observed and monitored, while the control group is not. Together, these factors could jeopardise the validity of results.

We therefore propose an 'active' control condition as an alternative. Our design is summarised in Figure 2. First, 100 schools are recruited into the trial. Half of these schools are then allocated to a 'maths tuition' treatment arm, and half to an 'English tuition' treatment arm. As we propose the intervention to be conducted in Year 6, children in both arms will sit Key Stage 2 mathematics and English tests at the end of the intervention period. Average Key Stage 2 *maths* scores of eligible children within treatment arm 1 (*maths* tuition) will then be compared to those of children in treatment arm 2 (*English* tuition). Hence our design will answer the question '*what is the impact of receiving maths tuition upon children's maths test scores, relative to receiving tuition in English.'* Moreover, we will be able to simultaneously answer the question *'what is the impact of receiving English tuition upon children's English test scores, relative to receiving tuition in mathematics*.'

**Figure 2. Trial design**




The following features of this design should be noted. First, it provides a robust estimate of the impact of *both* the TT mathematics and TT English programme (with equal statistical power). This hence meets one of the desirable criteria outlined in the tender document. Second, there is no need for a financial incentive, as all schools will receive some form of TT support. Third, as all children identified as in need of academic support will receive some form of tuition (either in maths or English) we believe this design will be more palatable to schools, teachers, parents and ethics committees. Finally, as the 'control groups' are active (they are also receiving tuition, but in a different subject area) Hawthorne effects are less likely to influence the results.

However, it is also important to note the limitations of this design. First, TT will only be able to provide Maths or English tuition within any given school (and not both). Second, any indirect impact the intervention has upon pupils' attainment (e.g. through increased enjoyment of learning or higher aspirations due to interacting with university students) will not be captured. This may or may not be important, depending upon Tutor Trust's theory of change. However, we do note that the previous EEF qualitative evaluation of the programme thought such factors may play a role. Consequently, we have also integrated an RDD within our evaluation design, allowing us to also estimate the impact of the programme against an *inactive* control (thus helping to overcome the second limitation given above). Further details are provided below.

Finally, it should also be noted that the broad design outlined above could be adjusted, following discussion with the EEF and the programme developer, to align with their preferences. For instance, 'arm 1' could be altered so that it

includes both maths and English but taught with *low intensity*, while 'arm 2' covers the same material but is taught with *higher intensity*. We will therefore be flexible in the specifics of our design, and be willing to adjust the details following further conversations with EEF and TT.

*Year group:* The call for tenders suggested that Tutor Trust have a preference to focus upon children preparing for Key Stage 2 examinations. This suggests participants should be pupils in Year 6. We agree that this is a good practical choice, given that teachers are likely to be more receptive to tutors coming into their school as children are preparing for important national exams. It also lends itself to use of high quality data contained within the National Pupil Database (NPD). Consequently, we strongly encourage the study to focus on pupils in Year 6. Note that pupils entering Year 6 in September 2016 will sit their Key Stage 2 exams in June 2017, with data likely to be available by November 2017.

*Identification of eligible pupils:* As a tuition programme, only a sub-set of eligible pupils will receive the intervention within each school. Our power calculations suggest that this needs to be approximately ten pupils per school; any less than this will lead to non-trivial increases in the minimum detectable effect size.

The tender document noted that schools are usually free to select the pupils to receive TT support. We therefore wish to maintain some element of teacher selection of pupils within our evaluation. Our preferred method of doing so is as follows. Prior to randomisation, all schools will be asked to provide a list of UPNs for each pupil within the school. In this spreadsheet, schools will also be asked to provide responses to the following six questions for every child:

Please provide a score between 0 and 10 to the following questions for each child within you school:

- Is this child likely to meet the expected national standard in Key Stage 2 mathematics?
  (0 very unlikely to 10 extremely likely)
- Is this child likely to meet the expected national standard in Key Stage 2 English?
  (0 very unlikely to 10 extremely likely)
- How would you rate this child's maths skills relative to other Year 6 pupils in your school?
  (0 = Far below average; 5 = Average; 10 = Far above average)
- How would you rate this child's English skills relative to other Year 6 pupils in your school?
  (0 = Far below average; 5 = Average; 10 = Far above average)
- Do you think this child would benefit from extra small group support in maths?
  (0 = Of little benefit; 10 = Of great benefit)
- Do you think this child would benefit from extra small group support in English?
  (0 = Of little benefit; 10 = Of great benefit)

Using schools responses to these questions, we will conduct a principal components analysis to create a scale. This scale will then be used to rank all children within each school in order of priority of receiving the TT intervention. All children with a certain ranking (e.g. within the bottom 10) will then be considered eligible to receive TT support. This selection of pupils will be binding, and cannot then be altered by schools. All these children will then receive TT support in *either* mathematics or English, depending upon which arm of the trial their school is assigned.

We will expect Tutor Trust to gain at least opt-out consent from children's parents to conduct the intervention, and for schools to provide consent enabling our use of the NPD data. If parents do opt their child out of the programme, the child ranked next in the list will replace them. In other words, this methodology has an in-built reserve list.

As an alternative to asking the questions above, schools could provide a list of around 20 pupils who they believe would benefit most from the intervention. They would then be asked to rank these pupils from 1 to 20, in order of priority of receiving the intervention. Which method is chosen could be jointly decided between the evaluator, EEF and TT.

*Testing and outcome measures:* All data used in this evaluation will be drawn from the National Pupil Database (NPD). This has the advantage of being inexpensive, minimising burden upon schools, being externally marked, high stakes, high quality, and will mean there is almost zero attrition. Key Stage 1 and Foundation Stage Profile scores will be used as the pre-test measures. (Although separate maths and English pre-tests could be conducted, we believe the cost and burden placed upon schools does not make this worthwhile with respect to the relatively small increase it offers in statistical power).

One issue TT and EEF should keep in mind is that Key Stage 2 tests will change in 2016. A decision will have to be made as to whether IRT-scaled or raw scores will be used as the outcome. Our preference would be for the former, but are happy to agree to either, so long as this is pre-specified in the study protocol. Moreover, there has been suggestions that children's responses to each test question will be provided within the NPD. If so, we will offer additional formative analysis of each question for the benefit of TT. This will help TT understand which test questions their programme has the biggest impact upon. (See the 'value-added' section below for further details).

In principle, the NPD could then be used to follow children in the trial throughout their educational career. For example, in June 2022 the children involved in the trial will sit their GCSE exams, and in 2024 they will complete A-Levels. The NPD will allow monitoring of these long-term outcomes - although we recognise this is out of scope of this project.

*Primary outcomes: RCT analysis:* Following recruitment of schools, we will assess external validity by analysing the pattern of school response among sampled schools. To check whether randomization has indeed balanced observed potential confounding factors, we will be able to compare characteristics of children in the two treatment arms through the information included in the NPD (e.g. FSM status, gender, Foundation Stage Profile and KS1 scores). We will also be able to include these variables into our analysis to (i) further reduce the possibility of confounding and (ii) increase statistical power.

Our main analysis, which will be the foundation of the EEF report, will be based upon the RCT design outlined above. Our design means we have a treatment group (e.g. all children ranked 1 to 10 within schools assigned to 'mathematics tuition' treatment) and a randomly equivalent counterfactual group (e.g. all children ranked 1 to 10 in schools assigned to 'English tuition' treatment).

Key Stage 2 maths and English test scores will be the dual primary outcome. The RCT analysis for Mathematics will be based upon the following OLS regression model:

$$KS2\_Math_{ij} = \alpha + \beta.T_{ij} + \delta.KS1_{ij} + \gamma.D_{ij} + \varepsilon_{ij}$$

Where:

KS2_Math = Children's scores on the Key Stage 2 Mathematics test

T = Equal 0 for the English tuition ('control') group and 1 for the maths tuition ('treatment') group.

KS1 = Children's scores on the Key Stage 1 tests (teacher reported)

D = A set of basic demographic controls (e.g. gender, FSM)

I = Child i

J = School j

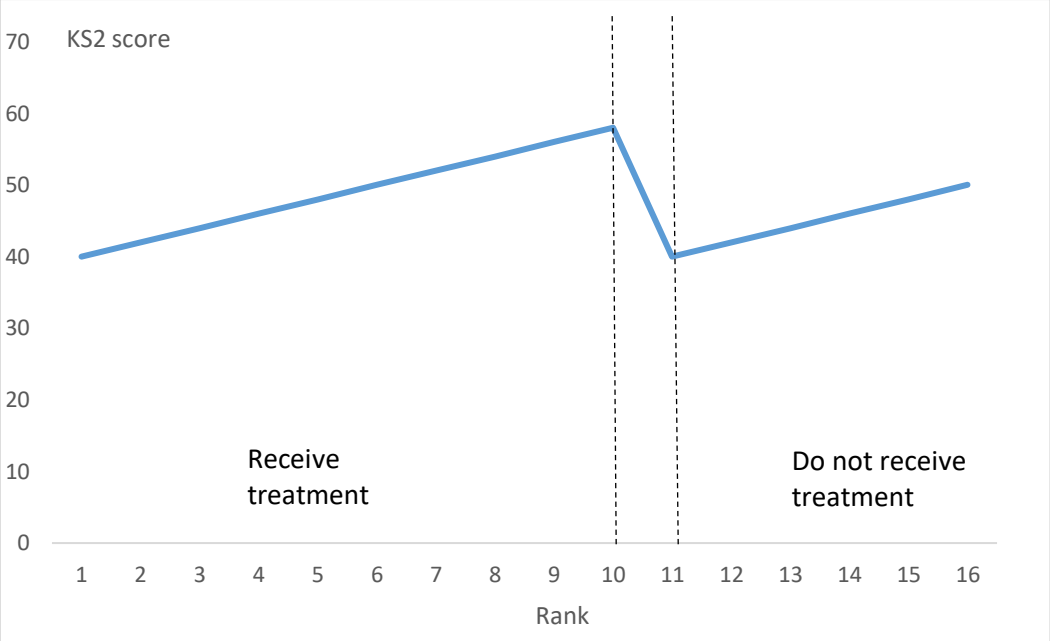
We will begin by estimating this model including all of the selected children from across the 100 schools. The model will then be re-estimated for FSM pupils only. Huber-White adjustments will be made to all standard errors to account for the clustering of children within schools. The parameter of interest from this model will be β – the impact of receiving mathematics tuition (rather than English tuition) upon children's mathematics test scores. (The example above focuses on the impact for mathematics. A separate model, following a similar process, will also be estimated for English).

Secondary outcomes: RDD analysis: The design we propose has an additional advantage. As well as allowing us to estimate the 'average treatment effect' of the intervention through the RCT, we can also estimate the impact upon the 'marginal' child (i.e. the child on the boundary of receiving the intervention or not) through an embedded Regression Discontinuity Design (RDD). Analysis of this RDD will form our 'secondary outcome'.

The graphs below illustrate how we have also embedded an RDD into our design. Running along the horizontal axis is the rank ordering of children to receive the intervention. The vertical axis then provides average Key Stage 2 test scores for children assigned each rank. Within this example, we assume children ranked 1 to 10 within each school received the treatment, while those ranked 11 and below did not. There is consequently a sharp discontinuity in the probability of receiving the treatment between children ranked 10$^{th}$ (where the probability equals 1) and children ranked 11$^{th}$ (where the probability equals 0). If the intervention does indeed have a positive impact upon the marginal child, then we should observe a pattern similar to Figure 3. Specifically, within treatment schools, there should be an appreciable difference in KS2 scores between children on one side of the discontinuity (e.g. ranked 9$^{th}$ or 10$^{th}$) versus the other (e.g. ranked 11$^{th}$ or 12$^{th}$). It is this 'kink' in Figure 3 which provides the estimate of the causal effect of the programme upon the marginal child.

This additional analysis is likely to offer a number of supplementary benefits to both EEF and TT. First, it means we are able to provide additional information on the impact of the intervention; we will not only estimate the impact on average (via the RCT), but also upon the marginal child (via the RDD). Second, recall that the RCT will have an '*active*' control condition. In contrast, children ranked just below the discontinuity within each school (e.g. 11$^{th}$ or 12$^{th}$) will not be in receipt of any intervention. Hence the RDD will provide evidence on the impact of the TT programme relative to an '*inactive*' control. Third, by integrating an RDD within an RCT, we will be testing a new type of evaluation design that could be used in future EEF trials. Fourth, we will be happy to present this design to other evaluators at the annual EEF conference, and to other interested groups.

**Figure 3. Hypothetical pattern of KS2 scores observed within <u>treatment</u> schools**




<u>Costs:</u> Information on costs will mainly be provided by the developer. We will ask TT a series of structured questions, asking for exact details on *specific* costs. This will include financial costs (e.g. tutor pay, resources and travel costs) and time (e.g. recruiting and training the extra tutors), with the latter then converted into a monetary value. Further data may then also be collected within the qualitative research within schools.

**<u>Outputs and the 'value-added' of our proposal</u>**

As per a standard EEF grant, our proposal includes all data analysis, along with writing and disseminating the final report. However, we will also provide the following additional outputs for the benefit of Tutor Trust and the EEF:

1. *Additional analysis of KS2 data on behalf of TT and participating schools:* Key Stage 2 tests are changing as of June 2016. The Department for Education recently suggested that children's scores on each individual test question may be deposited within the NPD. If this is indeed the case, we will provide additional formative analysis to TT about the impact of their programme. Specifically, we will provide an analysis of each individual test question, focusing on the percentage of children who provided the correct response in treatment and control schools. This will help TT better understand which questions their intervention has had an impact upon. Moreover, we will provide data back to participating schools, illustrating how their children perform on each Key Stage 2 test question. (This is of course conditional upon question-level data indeed becoming available).
2. *Presentation of our design at the EEF annual conference*: We believe that our design of integrating an RDD within an RCT is innovative, and has not been used within an educational RCT before. At the same time, we recognise it is important for best practise to be shared amongst evaluators. Hence we would happily present our proposed approach within a future EEF conference, were it of interest.

**Process Evaluation focused on Mathematics tuition**

We note that the Tutor Trust programme covers two academic areas: English and Mathematics. Due to space constraints, in this proposal we only provide details on the process evaluation in Mathematics. However, if our bid is successful, Professor Gemma Moss from the University of Bristol (formally UCL-IoE) has indicated her interest in conducting a process evaluation of the English programme. All costs and methodological details below therefore refer to the process evaluation of the mathematics element of the Tutor Trust programme only.

The process evaluation will aim to:

(i) understand how the intervention was implemented, including the extent to which it followed faithfully the intended approach;
(ii) identify successful tuition and what factors seemed to be important in determining that success;
(iii) identify less successful tuition and the corresponding factors;
(iv) clarify what steps might be taken that would support success in future interventions. A mixed methods approach, taking care not to place an unacceptable level of burden on pupils, teachers, schools, tutors or project managers, will be used across three phases.

Phase One

The purpose of Phase One will be to understand the nature of the mathematics intervention.

To this end, the evaluation team will visit 2 primary schools in Manchester, which are already adopting the Tutor Trust tuition programme. During the visit, we will discuss with stakeholders such as headteachers, teachers and their assistants about their implementation of the programme and its perceived impacts.

These visits will take place in June/July 2016 if practical or otherwise September/October 2016.

Phase Two

The purpose of Phase Two will be to assess the existing baseline conditions in schools that might influence the process and outcomes of the intervention.

The evaluation team will create an online questionnaires based on the aims of the Tutor Trust programme, issues arising from Phase One, and from our extensive knowledge about teaching and learning mathematics in primary schools. We will collect this survey data from all 100 schools directly involved in the mathematics intervention or acting as controls through the English intervention. An initial analysis will reveal interesting issues such as exceptional cases, which we will seek to clarify through follow up visits to collect interview data from headteachers, teachers and their assistants, and if appropriate from students. We anticipate at least two such visits. We will also collect data from the Tutor Trust project team. In particular we will need data about the qualifications, experience and understanding of the programme of the tutors involved in the teaching. The purpose of this aspect of the data collection is to explore the issues they have experienced in implementing the programme in the pilot schools. A key question for this stage is: how can the intervention and the delivery of the intervention be improved?

The Phase Two data will be collected early in, or even prior to the start of, the intervention to act as a baseline for our further investigations (Sept/Oct 2016).

Phase Three

Phase Three will collect further data which, by comparison with that already collected in the first two phases, will enable us to assess the main questions:

(i) how was the intervention implemented in practice, including the extent to which it followed faithfully the intended approach;
(ii) what does successful tuition look like and what factors seem to be important in determining that success;
(iii) what does less successful tuition look like and what are the corresponding factors;
(iv) what steps might be taken that would support success in future interventions.

A variety of instruments and methods will be used. We will need data about the participation of tutors and tutees to check fidelity to the intervention programme (collected June 2017). We will collect data from:

a) pupil focus groups attended by a self-selected subset of the 10 pupils in each of six schools (May/June, 2017);
b) interviews with the classroom teachers of pupils involved in the intervention in six schools (May, June 2017);
c) observations of one tuition as it happens, followed by an interview of the tutor, in each of six schools (May, June);
d) a questionnaire, based on that used in Phase One but tuned to the issues in the intervention, completed by the 100 intervention and control schools (June/July 2017).

**Budget:** The Table below provides an outline of provisional costs and are subject to change. In particular, it should be noted that the process evaluation costs only cover the mathematics component of the Tutor Trust programme. Any costs for a process evaluation of the English tuition (to be led by Gemma Moss) would be additional.

Impact evaluation

| | 2016 | 2017 | 2018 | Total days | total £ |
|---|---|---|---|---|---|
| John Jerrim | 2 | 4 | 12 | 18 | 10,300 |
| Kirstine Hansen | 2 | 0 | 6 | 8 | 5,000 |
| Terry Ng-Knight | 8 | 0 | 4 | 12 | 5,500 |
| Nikki Shure | 8 | 0 | 4 | 12 | 5,500 |
| Administration | 1 | 1 | 1 | 3 | 1,300 |
| | | | | | 27,600 |

Mathematics process evaluation

All costs for the process evaluation will be incurred during 2016/17.

| | |
|---|---|
| Dave Pratt and process evaluation team: 27.5 days @ £630 | 17, 325 |
| Train journeys: 16 trips to Manchester and Leeds @ £200 | 3,200 |
| Transcription: 6/0.5+6x0.5+6x1 = 12 hours @ £80 per hour | 960 |
| Total Cost for mathematics process evaluation | 21, 485 |

Total provisional cost = £49,085 (excluding VAT, if applicable).

## The Project Team

***John Jerrim*** is a Reader in Education and Social Statistics. He has led on a number of EEF evaluations, including Maths Mastery (secondary school) and Chess in Schools, and has contributed to a number of others. He will be Principal Investigator and manage the impact evaluation.

***Kirstine Hansen*** is a Reader in Social Policy and is Head of Teaching in the UCL Institute of Education Department of Quantitative Social Science. Her research interests include: child development, life course trajectories of disadvantaged children. She will lead the data analysis on the project.

***Nikki Shure*** is a Research Associate at the UCL Institute of Education and D.Phil. candidate in the Department of Economics, University of Oxford. Her research interests are education policy evaluation, early education, non-cognitive traits and educational outcomes. She will conduct the data analysis and assist with writing the report.

***Terry Ng-Knight*** is a Research Fellow in the Department of Quantitative Social Sciences. His research interests include school transitions and child development. He will conduct the data analysis and assist with writing the report.

**Dave *Pratt*** is Professor in Mathematics Education at UCL-IoE. His interests include the relationship between technology and mathematical thinking, statistical thinking and the micro-evolution of mathematical knowledge. He led the EEF process evaluation of the Maths Mastery RCT. He will lead the mathematics process evaluation.

***Cosette Crisan*** is a lecturer in Mathematics Education at the UCL Institute of Education. Her areas of interest include the development of the professional identities of the non-specialist maths teachers. She was part of the team who contributed to the EEF process evaluation of the Maths Mastery secondary school RCT.

***Melissa Rodd*** is senior lecturer in mathematics education at the UCL Institute of Education. Her research interests center around mathematical thinking, affective responses to mathematics, and participation in mathematics, using qualitative and philosophical methodologies. She will be involved with the mathematics process evaluation.

***Gemma Moss*** is Professor of Education at the University of Bristol. She has been a member of British Educational Research Association (BERA) Council since 2010, and before that was convenor of the BERA special interest group in Educational Research and Educational Policymaking. She will lead the English process evaluation.

## Appendix A. Timeline for impact evaluation

The impact evaluation we propose will have the following steps:

Step 1: Define the population of interest (December 2016)

The IoE will work together with Tutor Trust (TT) to define the population of interest. This population is likely to be restricted to certain geographic areas (e.g. Leeds and possibly surrounding LEAs) and schools with particular characteristics (e.g. a moderate to high proportion of children eligible for FSM). Using this definition, the IoE team will create a 'population list' of all schools who will be eligible to participate in the RCT.

Step 2. School recruitment by TT (January to June 2016)

From the list of eligible schools developed in step 1, TT will asked to recruit 100 schools to participate in the trial. To maximise external validity, these schools should ideally be a random sample. However, we appreciate that practicalities may mean this is not possible.

Step 3. School information provided (January to June 2016)

In order to participate in the trial, schools will be required to provide the evaluation team with the following pieces of information *prior* to randomisation:

(a) A letter of consent, allowing the evaluators Tier 1 access to information from the NPD
(b) A spreadsheet of all pupils within their schools. This will include some combination of UPN, pupil name and Date of Birth. (Exact details to be decided after further discussions with EEF and pending the latest advice from the Department for Education).
(c) *Critically*, within this spreadsheet, schools will also need to provide for each child their responses to the six questions given above.

No school will be randomised until all pieces of information are received.

Step 4. Children who are eligible for treatment are identified (1st week of June 2016)

At this point, IoE will create a scale based upon schools responses to the six eligibility questions. They will then rank children within each school, with the bottom X (e.g. 10) deemed eligible to receive TT support. It is critical that this decision is made at this point, and is not communicated to schools before they have completed step 3 above. This selection of pupils will be binding.

Step 5. Random assignment of schools (2nd week June 2016)

School-level randomisation will be used, with 50 schools in the 'mathematics tuition' treatment group and 50 schools in the 'English tuition' treatment group. Prior to randomisation, we will stratify schools into ten equal groups based upon historical Key Stage 2 achievement and the proportion of children eligible for FSM. Within each of these ten strata, five schools will be randomly allocated to the mathematics tuition treatment arm, and the other five schools to the English tuition treatment arm. The IoE team used a similar approach for the Chess in Schools trial, which resulted in very good balance between treatment and control groups.

Step 6. TT to inform schools of schools allocation (3rd week June 2016)

The IoE will inform TT of the allocation to treatment arms by Friday 10th June 2016. TT will be able to inform schools the following week.

Step 7. Intervention in schools (September 2016 to July 2017)

The intervention will start when Year 6 pupils return in September 2016. The intervention will run until July 2017.

Step 8. Key Stage 2 exams sat (June 2017)

Children will sit their Key Stage 2 examinations in June 2017.

Step 9. Key Stage 2 data received by the evaluation team (October/November 2017)

The impact evaluation team will receive the Key Stage 2 data in October/November 2017. They will begin their analysis of this data.

10. RCT and RDD analysis and reporting (October/November 2017 to April 2018)

The final report will be published in April 2018. This is under the assumption that the RCT is run over one academic year, and the NPD is made available to the evaluators in October 2017.