

## 1. Evaluation Summary

<b>Age range</b>	Year 2 (then 3) and Year 5 (then 6) (2-year intervention)
<b>Number of pupils</b>	Approximately 13,000 pupils
<b>Number of schools</b>	100 schools
<b>Design</b>	Cluster-randomised controlled trial with school-level randomisation
<b>Primary Outcomes</b>	Pupil level: Year 3 (Reading) and Year 6 (English and Mathematics)
<b>Secondary Outcomes</b>	Pupil level: Pupil engagement School level: Improved TA deployment; Change in teacher/TA practice
<b>Type of trial</b>	Efficacy

## 1. Background

### 1.1 INTERVENTION

**BRIEF NAME:** Maximising the Impact of Teaching Assistants (MITA)

#### **WHY (RATIONALE, THEORY AND/OR GOAL OF ESSENTIAL ELEMENTS OF THE INTERVENTION)**

The aim of this trial is to assess the effectiveness of a whole school intervention that aims to better deploy teaching assistants (TAs) and improve pupil outcomes through: (i) higher quality TA-pupil interactions; (ii) improved classroom management and lesson planning; and (iii) allowing classroom teachers to work more with lower-achieving pupils. This is one of the first times a trial will test a whole school intervention aiming at improving how schools, teachers and TAs can improve the use of TAs in everyday classrooms. The trial will address common issues with existing practice and replacing this with more a strategic, coherent and collaborative approach.

The EEF's extensive coverage of TA deployment, culminating in the TA Toolkit,<sup>1</sup> serves as the backdrop to this trial. Despite support staff being widely used and viewed positively in terms of their impact on teaching, evidence about the actual effectiveness (or otherwise) of TAs is scarce (Blatchford et al., 2009). The genesis of MITA is the Deployment and Impact of Support Staff (DISS) study, which found that the presence of support staff (TAs) facilitated individual attention and overall teaching as well as pupil engagement, but at the same time, decreased pupils' contact with teachers (Blatchford et al., 2007). An unintended consequence of these arrangements was that pupils receiving the highest amounts of TA support (lower attaining pupils and those with SEND) achieved less well academically, compared with pupils who had little or no support (Blatchford et al, 2012). Although there is evidence that TAs have a positive impact on teachers' workload and stress level, it was not until the DISS research when evidence on TAs impact on pupils for a long sustained period of time could be gathered. So far there have been two streams of research focused on the impact of TAs on learning outcomes: i) those assessing curriculum changes or 'catch-up' interventions, and ii) focused on TA deployment. It is clear from the evidence that the first type of interventions have a positive impact, although only after TA had been properly trained to implement change (Alborz et al, 2009).

This intervention consists of three levels of support: 1) training delivered to Senior Leaders in schools (two leaders from each school, including the headteacher) held in school 'clusters' throughout the year; 2) School visits from a National Leader of Education (NLE) (a practicing Senior leader) linked to the London Leadership Strategy ('LLS') who will provide support in identifying gaps in current practice and developing and implementing a change management plan; and 3) School training for all teachers and

---

<sup>1</sup> [http://www.dyslexiaoutreach.co.uk/wp-content/uploads/2015/01/Making\\_best\\_use\\_of\\_TAs\\_printable.pdf](http://www.dyslexiaoutreach.co.uk/wp-content/uploads/2015/01/Making_best_use_of_TAs_printable.pdf)  
[accessed 20.12.2016]

TAs on the 'scaffolding framework' focused on effective interactions with pupils.<sup>2</sup> The NLE consultant will aim to provide continuous support to school staff between training sessions and this will also promote engagement with other elements of the intervention.

This protocol has been developed by RAND Europe/University of Cambridge, and shared and refined with UCL Institute of Education's intervention team, whose comments and input have informed its content.

### **SIGNIFICANCE (POLICY CONTEXT, RATIONALE AND EVIDENCE FOR EQUIPOISE)**

This trial starts from the research done first on the DISS project, followed by two other studies: the Effective Deployment of Teaching Assistants (EDTA) and the Making a Statement (MAST) (Webster et al, 2015). Both the DISS and MAST studies shed light on the impact of TA support on pupil learning, namely: i) TAs have a predominantly pedagogical role, with direct interaction with pupils with SEND, which can produce an unintended separation of these pupils from their teacher and the curriculum, ii) the MAST study showed that pupils with SEND have a lower quality learning experience compared to their peers, and iii) TAs' quality of preparation is poor.

In 2014, the UCL Institute of Education implemented a two-term school improvement programme called MITA (Webster et al, 2015), based on a book, which in turn was the product of the project. This programme aimed to support school leaders by rethinking on and fostering change of the use of teaching assistants.

The DISS and MAST studies had an impact on policy making informing the Lamb Inquiry and the 2014 SEND Code of Practice, and researchers at UCL-IOE have worked with the EEF to produce a toolkit and recommendations for school leaders and teachers on how to better deploy and use teaching assistants.

The documented evidence emerging from the previous research and the culmination of it into the EEF's extensive coverage of TA deployment through the TA toolkit serves as a starting point to produce robust evidence on the actual effectiveness of TAs, which is still scarce (Blatchford et al, 2009). This lack of evidence coupled with the fact that the proportion of TAs on the total school workforce has increased in the last 20 years by more than three times makes a strong case for robust evaluation of a scaled-up intervention.

Two different views will inform the evaluation approach: the implementation science and realist. The first view will produce evidence on 'dosage' or 'exposure', fidelity/adherence and implementation quality, whilst the second will address questions related to what works for whom and under what conditions.

### **WHO (RECIPIENTS OF THE INTERVENTION)**

In this trial, the outcome analysis will focus on Year 3 and Year 6 pupils, but the intervention will be implemented across the whole school through improving school leaders' strategic decision-making and changing teacher and TA behaviour. The three groups who are targeted for the intervention are: headteachers/school leaders, teachers and TAs. Schools will receive training and will develop an action plan during school year 2017/18 (intervention phase), and will implement it in 2018/19 (follow-up phase) when pupils are in Year 3 and Year 6. Baseline assessment will be undertaken at when these cohorts are at the end of Year 1 and Year 4 respectively, and then at the end of the follow-up year (2018/19).

### **WHAT (PHYSICAL OR INFORMATIONAL MATERIALS USED IN THE INTERVENTION)**

The trial will make use of the guides developed by the implementation team:

- The Teaching Assistant's Guide to Effective Interaction (Bosanquet et al, 2015);
- Maximising the Impact of Teaching Assistants (2nd edition) (Webster et al, 2015),

---

<sup>2</sup> This framework is designed to support TAs to scaffold pupils 'learning and foster independent learning where pupils are expected to self-scaffold while TAs observes their progress, intervening only when pupils show they are unable to proceed (Bosanquet et al, 2015).

- Making Best Use of Teaching Assistants guidance report (Sharples et al, 2015).

These guides have a number of resources such as a self-evaluation assessment, lesson observation schedules, online staff survey, action planning template and visioning exercise, to name a few. Some of these resources were produced in response to EEF guidance that has seven recommendations on how to improve the deployment of TAs. As above, the intervention will also comprise of external consultants (NLEs) visiting and providing support and producing reports summarising key areas for action.

#### **WHAT: PROCEDURES, ACTIVITIES AND/OR PROCESSES USED IN THE INTERVENTION**

The first set of activities involve the Senior Leadership Team (SLT) (composed of two leaders from each school, including the headteacher) attending four local half-day training sessions (delivered in geographical clusters) throughout the year, introducing the framework, process and strategies for the better deployment and effective use of TAs.

These training sessions will lead into 'gap tasks' that the schools will complete. The first gap task (between sessions 1 and 2) consist of 'auditing and visioning': critically reviewing current practice and developing a vision of what school leaders would like to see at the end of the process and changes they expect to see along the way (September 2020). Schools will use the resources listed above to complete the self-evaluation components of the audit. The NLE will additionally conduct an independent review of practice (including talking to staff, doing observations). The results will be given to school leaders to feed into the overall audit. The second gap task (between sessions 2 and 3) focuses on developing a detailed action plan and early piloting of new strategies. Again, the NLE will attend each school to support SLTs.

In total, there will be three school visits (one per term). As well as providing the essential tailored challenge and support to SLT, NLEs will keep a record of the planning, implementation and change process using a predetermined list of items to check (NLE checklist).

In preparation for implementing the action plan and changing TA deployment and practice, there will be two half-day training sessions for TAs (one early in the Spring term and a second one four weeks later). These will emphasise what effective interactions with pupils look like (e.g. promoting pupil independence; applying the scaffolding framework) with 'gap tasks' for TAs to practise. On the same days as the TA training is taking place, teachers will also receive two hours of training on how to plan and organise classrooms effectively to make the best use of TAs' new skills and knowledge. A senior leader involved in the project will be asked to attend the teacher and TA training, so that the training can be fully embedded and supported as implementation proceeds.

#### **WHO (INTERVENTION PROVIDERS/IMPLEMENTERS)**

The intervention was developed by a group of researchers at University College London-Institute of Education (UCL-IOE), the University of East London (UEL) and London Leadership Strategy (LLS) who will jointly implement the intervention.

#### **HOW (MODE OF DELIVERY)**

UCL-IOE will deliver the training sessions for the SLTs in geographical clusters; the University of East London will lead on training teachers and TAs in each school; and the London Leadership Strategy will deliver the in-school support and coaching via the NLEs.

#### **WHERE (LOCATION OF THE INTERVENTION)**

The intervention will take place in participating schools in four geographical areas:

- Cluster 1: West Midlands (principally around Birmingham and Wolverhampton)
- Cluster 2: Portsmouth, Havant, Fareham and Hampshire
- Cluster 3: Barking & Dagenham, Redbridge, Havering, Thurrock and London
- Cluster 4: Suffolk

### **WHEN AND HOW MUCH (DURATION AND DOSAGE OF THE INTERVENTION)**

MITA is a whole school intervention lasting one academic year. Given that it involves large-scale organisational change it is felt by the developers that pupil outcomes will not be affected sufficiently until the final term of the year in which changes are implemented in the schools (the Summer term of 2018-19).

The intervention consists of 12 hours training for SLT and three support sessions from the NLE in-school, six hours' training for TAs and four hours' training for teachers. These training/support sessions will be delivered across the course of the 2017-18 school year. In the following year, 2018/19, substantive changes developed during the training will be implemented by the schools.

A planned output from the intervention and a measure of both engagement and the likely success of the approach is a detailed and achievable action plan. Schools will develop this plan during the 2017/18 academic year, and implement it in the following academic year (2018/19). Outcomes will be measured in the final term of the 2018/19 academic year. In the context of MITA, 'dosage' relates to the level of engagement in relation to each component of the intervention: (i) attendance at MITA SLT sessions; (ii) meetings with the NLE; (iii) TA/Teacher attendance at training; and (iv) completion of gap tasks (see Logic Model in **FIGURE 1** below). Dosage and fidelity will be monitored during the implementation phase.

### **TAILORING (ADAPTATION OF THE INTERVENTION)**

Although all schools will, on paper, receive the same training and be supported by a NLE from the same organisation, it is likely there will be some tailoring depending on the starting point for each school. This starting point will be recorded by the NLE and passed on to the evaluation team. This will be best captured via the numerical outputs of the data from the components of the audit, to which the NLE can add qualitative detail and additional context. Each school's vision and action plan, resulting from the first year of the intervention, will be adapted to each school's context and current situation.

The training offered to TAs is standardised, though feedback from the NLE can be used to adapt the overall delivery approach. The training for TAs is a standardised package, but the teacher input can be adjusted for session 2, depending on what SLT decide needs to be developed in more detail – hence, the value of a member of SLT being present at the training.

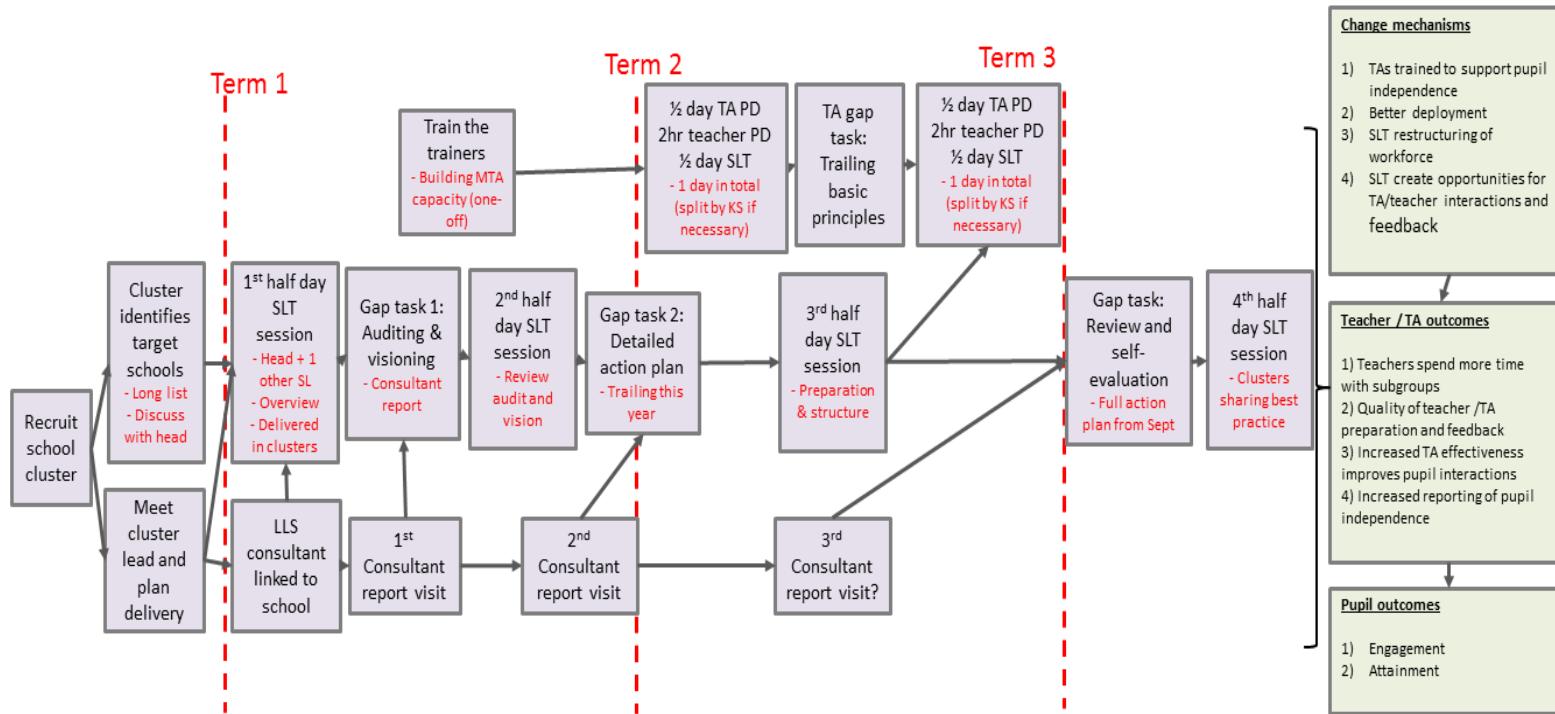
### **HOW WELL (PLANNED): STRATEGIES TO MAXIMISE EFFECTIVE IMPLEMENTATION**

The intervention includes staggered training sessions throughout the year as well as regular visits by a consultant (NLE) in each school to support SLT, maintain their engagement and ensure progress of the intervention. The intervention and evaluation teams will develop a communication and "keeping in touch" strategy to keep schools engaged throughout the project, especially for the second year when schools will no longer be receiving training. This engagement will include regular newsletters; furthermore schools in the control condition will be invited to attend free SEN training in summer 2018 and to bring Year 6 pupils to a special widening participation day at UCL in summer 2019. The control schools will also receive a small financial incentive at the end of the evaluation.

### **LOGIC MODEL**

**FIGURE 1** sets out the intervention logic model developed during trial setup meetings between the Education Endowment Foundation (EEF), UCL-IOE / UEL / LLS and the evaluation team (RAND Europe and the University of Cambridge).

**FIGURE 1: MITA LOGIC MODEL**



## 2. Methods

### 2.1 RESEARCH HYPOTHESES

This project will test several hypotheses relating to the impact and delivery of MITA. Specifically that MITA:

1. has a positive effect on pupils' attainment, specifically:
  - a) Better reading outcomes (vs controls) for Year 3 pupils.
  - b) Better reading and mathematics outcomes (vs controls) for Year 6 pupils.
2. results in improved *deployment* of the school TA workforce.
3. results in change of practices, specifically:
  - a) Practices aimed at improved interactions between TAs and pupils.
  - b) Practices aimed at fostering pupil independence.
4. has a positive effect on *pupils' engagement*.

### 2.2 DESIGN

The main trial will consist of a stratified, two-arm, cluster-randomised controlled trial (cRCT), with at least 100 schools recruited from four geographical areas. Schools will be assigned to either treatment (MITA) or control (business as usual). All schools signing up will have a 50% chance to be assigned to the treatment group.

Given the readily available EEF TA resources, it will be important to collect high quality data on what control schools are doing, using a usual practice survey<sup>3</sup>. The intention of this survey is, in part, to uncover the extent to which control schools may be implementing elements of the MITA intervention in their school, thus making it harder to detect an impact of the intervention in the treatment schools.

The project lifecycle is divided into **planning, intervention and follow-up** phases. During the **planning** phase (September 2016 - June 2017) the evaluation team will complete this protocol and prepare key documents for school recruitment (Memorandum of Understanding, opt-out letters) as well as data collection tools (surveys, observation protocols). Meanwhile, the delivery team will recruit schools through events run at each cluster area. At the end of the planning phase, eligible schools will be randomly allocated into two groups, and implementation will begin. During the **intervention** phase (September 2017 – July 2018), schools in the treatment group will take part in the MITA programme. Schools in the control group, instead, will: i) receive newsletters from IoE about the progress of the MITA trial, ii) receive SEND training (not related to MITA) delivered by IOE in summer 2018; and iii) receive £250 in autumn 2017. The **follow-up** phase will see schools in the treatment group implementing changes over the academic year 2018/19 (see Section 2.5 below for further details on outcome measures by stage). During this phase, control schools will: i) be invited to bring Year 6 pupils to a widening participation day at UCL (summer 2019), and ii) receive £500 upon completion of the evaluation process. (Meaning that in total, control schools can receive up to £750 for participating.)

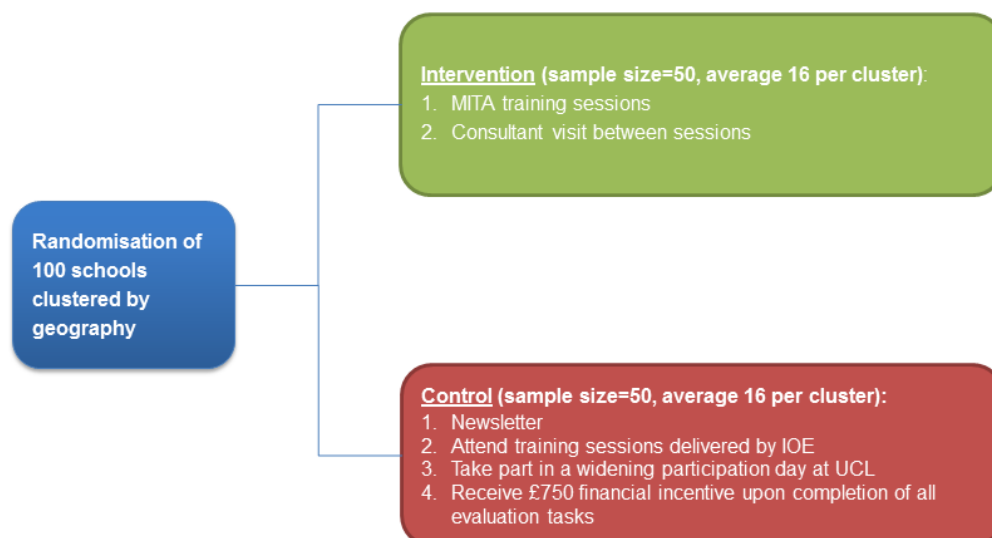
### 2.3 RANDOMISATION PLAN

There is only one intervention being randomised in this trial at the school level: training and support to improve deployment of TAs. The primary focus for the trial – and the basis on which power calculations have been conducted – is the effect of better deployment of TAs on pupil outcomes. We propose a stratified randomisation at school level (Figure 2), incorporating the fact that schools will be clustered in geographical areas. All schools will be randomised into one of two arms, with the school geography / cluster membership being a key stratification variable.

---

<sup>3</sup> A baseline survey exploring usual practice will be delivered in both intervention and control schools, see

FIGURE 2: RANDOMISATION DESIGN



Simple randomisation only balances confounding factors and yields unbiased estimates *on average* (Morgan and Rubin, 2012). For any particular experiment, covariate imbalance may still remain after randomisation in terms of both observed and unobserved characteristics. Imbalance of important covariates among treatment groups may influence the analysis and interpretation of the results. Although *ex post* adjustment can be made for imbalance, this is less efficient than achieving *ex ante* balance, and cannot be used in cases where all individuals with a given characteristic are allocated to one treatment group (Bruhn and McKenzie, 2009).

One approach to address imbalance is stratification. In the simplest form, strata are formed by creating groups based on sampling variables and other measures pertinent to pupil achievement. Within each stratum, we would then conduct simple randomisation using a random number generator (e.g. available in Microsoft Excel, <https://www.random.org/> or Stata). Each school within a stratum will be assigned a random number, with schools then sorted by the random number and the first half in each strata being allocated to treatment. Trial analysis will incorporate stratum variables (and will assess the effects of including and excluding these measures on analysis).

In this study, there are likely to be a range of variables that we would ideally stratify the randomisation on. The exact set of stratification variables will vary depending on the availability of data, but we will try to factor in salient characteristics that are related with teachers’ and students’ performance. We will stratify the sample based on the primary variables of geographic area and average prior attainment (the latter measured at the school level in terms of KS2 results). Table 1 below sets out the primary variables, what level of the study they relate to and the justification for actively seeking balance rather than leaving this to the randomisation.<sup>4</sup>

**TABLE 1: LIST OF STRATIFYING VARIABLES**

Measure	Level	Justification
Geographic areas	School	Reducing the chance of imbalance in terms of geographical location.
Prior attainment <sup>5</sup>	School	Reducing the chance of imbalance in terms of pupil attainment.

<sup>4</sup> A more recent proposal, and one that has gained traction in development economics in particular, is that of *re-randomisation* (Morgan and Rubin, 2012, Rubin, 2012, Worrall, 2010). Based on Morgan and Rubin (2012), and only if necessary, we will adopt a hybrid approach of randomising units to assess the chance of imbalance with a given set of covariates and if re-randomisation is necessary re-run the randomisations until a pre-specified degree of balance is achieved. The steps in this process are given in the Appendix. There are two re-randomisation methods that are commonly used in previous literature (Bruhn and McKenzie, 2009). The first is if a random allocation shows any difference in means between treatment and control group at the five percent level or lower of statistical significance (or another level if desired), re-randomisation is deemed to be required. The second method runs 1,000 randomisations at the same time and picks the assignment with the smallest *t* statistics.

<sup>5</sup> Based on 3 years of historical data to avoid any sort of anomalies that may be present from year to year.



Randomisation will be undertaken by a member of the evaluation team independently of the developers. Further details on stratification are explained in Appendix 1.

## 2.4 PARTICIPANTS

Given the limited training resource, school recruitment will be on a ‘first come first served basis’ – meaning that schools will be accepted on the trial once they have completed the required paperwork and prerequisite tasks (provision of pupil UPNs, list of teachers and TAs, signing of MoU.) All schools who fulfil the inclusion/exclusion criteria and who volunteer for the trial will be eligible for the intervention. School eligibility criteria for the intervention are listed below:

TABLE 2: ELIGIBILITY CRITERIA

MITA eligibility criteria
Primarily two or three form entry primary or junior school (one and four form entry schools will also be allowed, but we expect a small minority of schools will be in this category)
No prior engagement in MITA and/or Maximising the Practice of Teaching Assistants MPTA training sessions
No substantial prior action(s) taken following recommendations from the EEF TA guidance or MITA or MPTA handbooks
Not in special measures <b>OR</b> facing imminent leadership changes

Schools with high proportion of SEND and FSM pupils will be prioritised if possible. Local knowledge about leadership transition will be collected via cluster contacts (e.g. LA leads; Teaching School Alliance leads).

The recruitment of schools will be done via a set of events held in each of the three geographical clusters, where the implementation team leads (IOE-UCL) will present the project and share expression of interest (EOI) information. EOI information will also be available online, with a video presentation of one of the sign-up events: <http://maximisingtas.co.uk/projects/mita-eeef-project.php>. The team will follow up with schools who express interest and send a Memorandum of Understanding to be signed and sent to IOE-UCL. This in turn will trigger the sending of the items required to complete the project prerequisites. This staged approach spares schools receiving too much information in the first instance.

### 2.4.1 SCHOOLS

The study will be conducted in schools from four cluster areas:

- Cluster 1: West Midlands (principally around Birmingham and Wolverhampton).
- Cluster 2: Portsmouth, Havant, Fareham and Hampshire.
- Cluster 3: Barking & Dagenham, Redbridge, Havering, Thurrock and London.
- Cluster 4: Suffolk.

We expect the average number of schools per cluster to be about 30 (of which approximately 15 will be in the treatment group and 15 in the control group), but that there may be variation in the size of the four regional clusters. Participating schools will be asked to sign a Memorandum of Understanding that will outline the roles and responsibilities of all stakeholders involved and clearly set out the requirements for schools.

### 2.4.1 NATIONAL LEADERS OF EDUCATION (NLE) CONSULTANTS

The process of recruiting 12 NLE consultants will begin in December 2016, before the intervention starts. NLE consultants will initially be profiled and recruited from the existing pool of Special Educational Needs and Disability (SEND) Leaders already attached to LLS, and who will be familiar with the process of peer support and coaching in schools that will be used in the MITA intervention. NLEs attached to clusters where the trial is running will be invited to apply via an additional open recruitment process. Their local knowledge of schools will be an asset, plus it allows for local capacity building. The materials for the NLE training will be developed between January and April, and the

training sessions will take place in May and July 2017. Consultants will be assigned to schools taking into consideration their relationship with schools and how far away they live from a given school/cluster.

### 2.4.3 TEACHERS AND TEACHING ASSISTANTS

Given that this is a whole school intervention, all teachers and TAs (including those with different role titles, but ostensibly working in pupil/classroom support roles) in treatment schools, across all years, will be eligible for and receive the intervention.

### 2.4.4 PUPILS

All Year 2 and Year 5 pupils are eligible for inclusion in the trial (see Table below). Parents will be given the opportunity to opt out from the evaluation via a standardised EEF opt-out consent letter that will be distributed by participating schools once schools have agreed to take part in the study.

## 2.5 OUTCOME MEASURES

The **primary outcome** will be pupil attainment in Reading (Year 3 pupils) and Maths and Reading (Year 6 pupils). There are two **secondary outcomes** at two different levels: i) pupil engagement (pupil level), and ii) change in teacher/TAs practice (teacher/TA level). As outlined in the Logic Model above (**FIGURE 1**), it is expected that a better deployment of TAs will contribute to improved outcomes in these areas.

Table below shows the cohorts along the project lifecycle and when they will be tested.

**TABLE 3: OVERVIEW OF PUPIL YEAR GROUPS AND TESTING**

	<b>Planning year (2016/17)</b>	<b>Intervention year (2017/18)</b>	<b>Follow-up year (2018/19)</b>
	<b>Will be in _____ at baseline</b>	<b>Will be in _____ during the intervention</b>	<b>Will be in _____ during outcome testing</b>
<b>Year group</b>	Year 1 (end of)	Year 2	Year 3
	Year 4* <sup>6</sup>	Year 5	Year 6*

\*For Year 4s (who later will be in Year 6) KS1 and KS2 data will be used respectively as baseline and follow up.

### ***Pupil attainment***

Table 4 below shows the primary outcome measures for both cohorts at baseline and at the end of the implementation phase. A third party provider (ACER) will implement the external examination for the younger cohort while Key Stage 1 and 2 results will be used for the older cohort. ACER staff, who are blind to allocation, will invigilate and mark the tests.

ACER will undertake a data processing stage, recoding and checking for integrity with a report documenting recode syntax files, and will also produce a test calibration using ACER ConQuest as well as reviewing psychometric properties. Further, ACER will verify that the non-response rate is below 5%.

---

<sup>6</sup> KS1 results will be 2 years old, but the main purpose of this baseline data is to diminish variation at our analysis stage, not for the purpose of before and after analysis.

TABLE 4: SUMMARY OF ATTAINMENT MEASUREMENT

2016/17 (Planning year)	2017/18 (Intervention year)	2018/19 (Follow-up year)
<i>Baseline</i>		<i>Post-implementation</i>
Reading end of Year 1 (June 2017) - ACER, ELMs test	→	Reading end of Year 3 (June 2019) – ACER, ELMs test
Maths/Reading Year 4 (Key Stage 1)	→	Maths/Reading Year 6 (Key Stage 2)

### ***Pupil engagement test***

Pupil engagement will be measured at the end of the follow-up year (Term 3 during the 2018/19 school year). A previously validated measure, the “Engagement vs. Disaffection with Learning: Student-report” (EvDLS) which has previously been used with the age groups in this trial will be used for this purpose (Skinner et al. 2008).

### ***Change in practice measures***

With good fidelity of implementation, the expectation is to see **change** on the following aspects of practice:

- **Deployment of TAs:** teachers spending *more* time (and TAs *less* time) with lower-attaining and SEND pupils (e.g., those within the bottom quarter of attainment in the class).
- **Increased quality of TA-pupil interactions:** better TAs knowledge and use of effective interactions (e.g. pupils being given more ‘wait time’; greater use of prompts and clues; reduced use of correction; and more utilisation of self-scaffolding strategies).
- **Increased quality of preparations:** quality of TAs’ pre-lesson preparation and TAs’ feedback to teachers, improvements in opportunities for and quality of teacher-TA liaison.

This information will be collected through the teacher and TA surveys. Information about TA practice will also be gathered through classroom observations and audio recordings made in a small number of lessons. Staff surveys and interviews, classroom observations and audio recording will be used for triangulation to map change in practice across time.

### **2.5.1 BASELINE MEASURES**

Key Stage 1 (KS1) test results will be used as baseline scores in Mathematics and Reading for those pupils who will be in Year 4 at baseline and in Year 6 during the post-intervention phase of the project (the older age group). While KS1 tests take place in the final term of Year 2, and so the results will be two years old when these pupils are in Year 4, they provide the most reliable, efficient and equitable way of obtaining baseline data on attainment for this group. The tests have the advantage of being consistent across the school population and comparable with KS2 results, which we will use as the attainment measure post-implementation. KS1 tests also have the advantage of being taken by all pupils, and using them avoids the need for additional independent testing as part of the evaluation, which would have added cost, increased disruption in the schools, and placed unnecessary stress on pupils.

For the younger group, those in Year 1 during the planning phase (2016/17) and who will be in Year 3 for outcome testing, an external Reading test developed by ACER will be used. The RAND/Cambridge evaluation team chose the test, in consultation with the EEF and UCL Institute of Education, based on the evidence emerging from the pilot in terms of its effectiveness and measurement reliability.<sup>7</sup>

The **Essential Learning Metrics (ELMs) Reading Comprehension** test was designed for pupils from Year 2 to Year 10 and reports achievement in reading comprehension on a continuous scale. The scale

---

<sup>7</sup> Piloting performed by ACER in April 2017. The piloting was to assess the suitability of the Essential Learning Metrics (ELMs) Reading comprehension for baseline testing for pupils at the end of Year 1. ELMs is intended for Year 2 pupils so the piloting was used to check for floor/ceiling effects. Results from the piloting are included in Appendix 2, along with the write up of the testing provided by ACER.

is represented in both scale scores and described bands of achievement, allowing progress to be monitored and mapped over time across the years of schooling. This test covers a range of text types (narrative, persuasive, informational) to assess a pupil's ability in retrieving information, interpreting explicit information, and interpreting by making inferences and reflecting on texts. ELMs is online, standardised, can be administered to an individual or group (limited by number of computers), is not adaptive, and is untimed (but 40-50 minutes are recommended). ELMs is designed with reference to the National Curriculum. All items in the tests have been reviewed by English teachers in England for linguistic, cultural or curriculum bias, and all items are multiple selected responses that minimise marking bias. A paper version of the test will be used.

## **2.6 SAMPLE SIZE AND POWER CALCULATIONS**

The target number of schools is a minimum of 100, based on the capacity of the intervention team to deliver training at scale. We assume that there are on average 33 pupils per class and that with 100 schools, two year groups per school and two form entry (e.g. around 130 pupils per school in four classes), we would be collecting data on at least 13,000 pupils for this evaluation.

Based on EEF guidelines, the amount of variation explained by covariates is assumed to be 0.53 for level 1 (pupils) and 0.00 for level 2 (schools). We also assume an alpha of 5% and an intended 80% power to detect effects. Power and minimum detectable effect size (MDES) calculations were performed using the PowerUp tool (Dong and Maynard, 2013). We use two-level clustered designs and base our calculations on a range of values for the Intra-class correlation ICC (starting with an ICC of 13% as per EEF guidelines). Using the parameters above and with equal allocation to treatment and control the MDES is 0.207, but this is reduced if the ICC falls. With only 100 schools, the study will not be powered for sub-group analysis, such as different types of SEND.

Table 5 sets out the MDES calculations for our analysis. We have used cut offs of an MDES  $\leq 0.20$  and  $> 0.20$  to colour the table below. Anything below or equal to 0.20 is green, anything above 0.20 is orange. The collaborators on this project have some control over both participation rate and ICC – for example the evaluation team can bring the ICC down if we include correlates of achievement at pupil level (e.g. sex, ethnicity).

If one assumes that the two year groups could *not* be combined for analysis then the MDES would be 0.213 for a two-form entry year-group and 100 schools and n=66 pupils per school on average.

**TABLE 5: MDES CALCULATIONS BASED ON 100 PARTICIPATING SCHOOLS**

Average number of year / classes per school	ICC		
	0.13	0.10	0.08
<b>1 (n=33)</b>	0.214	0.190	0.173
<b>1.5 (n=98 pupils)</b>	0.207	0.183	0.164
<b>2 (n=130 pupils)</b>	0.207	0.182	0.163

Note: ICC is assumed between-school variability

## **2.7 ANALYSIS PLAN**

The starting point for the outcome analysis would be intention-to-treat (ITT). This method compares outcome means for the treatment and comparison groups, and subjects are analysed according to their randomised group allocation. The ITT approach is inherently conservative as it captures the averaged effect of *offering* the intervention, regardless of whether or not the participants comply with the assignment.

Problems of dropout/non-attendance may be an issue for this trial depending on how motivated school staff are. Our main concern is that new teachers come in or that schools and/or teachers/TAs and/or consultants drop out at some point during the trial. Both risks are mitigated by this being a whole school intervention, although it will be important to assess the levels of staff turnover in both arms of the trial.

The **primary outcome** is pupil-level test scores. We will use a two-level multilevel model to account for clustering of data due to the influence of unmeasured variables. This approach assumes that the schools in the study are a random sample of all schools and the multilevel modelling framework can flexibly handle complex variation within/between schools (Snijders and Bosker, 2012).

The main analysis consists of the model for outcomes of pupils nested in schools, which is:

$$Y_{ij} = \beta_0 + MITA_j\tau + X_{ij}\beta_1 + u_j + e_{ij} \quad (1)$$

where  $Y_{ij}$  is the achievement of student  $i$  in school  $j$ ;  $MITA_j$  is a binary indicator of the school assignment to intervention [1] or control [0];  $X_{ij}$  represents characteristics at pupil level (pupil  $i$  in school  $j$ ), such as average baseline scores (KS1 prior attainment results in Maths, Reading, Writing and Phonics);  $u_j$  are referred to as school-level residuals ( $u_j \sim i.i.d N(0, \sigma_u^2)$ ) and  $e_{ij}$  are individual-level residuals ( $e_{ij} \sim i.i.d N(0, \sigma_e^2)$ ). Equation (1) is known as a 'random intercepts' model because  $\beta_{0j} = \beta_0 + u_j$  is interpreted as the school-specific intercept for school  $j$  and  $\beta_{0j} \sim i.i.d N(\beta_0, \sigma_u^2)$  is random (as in it can take any value). The total residual variance can be partitioned into two components: the between-school variance  $\sigma_u^2$  and the within-school variance  $\sigma_e^2$ . Our target parameter  $\tau$  is the average effect of the intervention on pupil outcomes. In order for data from the year groups to be pooled, we will convert all scores into standardised metrics (z-scores). This means that results in the analysis will be in terms of standard deviation changes in the outcome.<sup>8</sup> For Year 6 pupils, analyses will be adjusted to account for multiple outcome testing.

We will run the analysis by sub-groups when possible taking into account that this could impact negatively the significance of coefficients. SEND feature will be incorporated into the regression analysis as a homogeneous category through a binary variable [1] if SEND, [0] otherwise. Nevertheless, we will report mean results by sub-categories of SEND. Further, FSM will be incorporated into the analysis as done with SEND.

The **secondary outcomes** are pupil engagement (at pupil level), improved TA deployment and change in teacher/TA practice (at school level). Pupil engagement will be measured following a similar specification to equation (1) above, whilst the last two secondary outcomes will only be available at school level and captured through staff surveys, interviews, classroom observations and audio recordings of TA-pupil interactions.

## **2.8 MISSING DATA AND ATTRITION**

Missing data can arise from item non-response or attrition of participants and / or schools. Our use of administrative data for pupil baseline data and outcome data for older pupils should reduce missingness arising from both item non-response and attrition. Should some data be missing, our analysis would always begin with an intention-to-treat approach as set out above, and then move from there depending on the extent of missingness and where it occurs in the data.

We will explore attrition across trial arms as a basic step to assess bias (Higgins et al., 2011) to assess whether there are systematic differences between those who drop out and those who do not. We can also use random slope models if heterogeneity is an issue and thus whether these factors should be included in analysis – we would model missingness at follow-up as a function of baseline covariates, including treatment. For item non-response, the extent of missingness may in part determine the

---

<sup>8</sup> In the event that re-randomisation is required, standard errors of the estimates following re-randomisation are considered to be too conservative. As a result, and only if necessary, we will use randomisation tests to obtain the accurate p-values and conduct hypothesis testing. The randomisation test simulates test statistics for each randomisation following the re-randomisation criteria used in the experiment, and uses this correct p-value for hypothesis testing.

analytical approach. For less than 5% missingness overall a complete-case analysis might suffice (i.e. assuming data are MCAR), but our default would be to check results using approaches that account for missingness but that rely on the weaker MAR assumption. Our preference would be to use Full-Information Maximum Likelihood (FIML) over multiple-imputation because FIML can be estimated in a single model and simulation studies show that it can reduce bias as well as MI (for a discussion of FIML vs MI see Allison, 2012).<sup>9</sup>

## **2.9 IMPLEMENTATION AND PROCESS EVALUATION METHODS**

Following the EEF guidelines on process evaluation (Humphrey et al., 2015), and informed by both implementation science and realist approaches to evaluation, we will conduct a process evaluation to better understand the complexities of MITA. As a whole school intervention based on changing how TAs are deployed across the school and in classrooms, and how TAs interact with pupils, MITA presents a number of evaluation challenges, not least for the process evaluation. The primary issues for the process evaluation are to capture fidelity of implementation throughout the school.

Implementation science emphasises thinking about ‘dosage’ or ‘exposure’, fidelity/adherence and implementation quality. The realist approach emphasises thinking about what works, for whom, and under what conditions. The purpose of the process evaluation will thus be to address the following questions:

- Was the intervention implemented with fidelity in the intervention schools?
- What factors and initial conditions appear to explain variation in fidelity of implementation?
- What appear to be the necessary conditions for success of the intervention?
- What were the barriers to delivery?

The process evaluation aims to examine the mechanisms of the intervention and inform the interpretation of findings from the quantitative analysis.

For the purpose of this evaluation, a small sample of teachers and TAs will be targeted for interviews, classroom observations and audio recording of interactions with the TAs. The observations will capture changes in pupil participation and types of TA talk; the audio recordings changes in TA practice/talk; and the interviews will offer insight into the rationale for using types of high or low intervention strategies.

Classroom observations will be focused on a small number of teachers, TAs and pupils. The selection of lessons for these observations will be made on a pupil cohort basis. A selection of up to 3 cohorts of pupils in each school will be made (one younger cohort at the end of Year 1 at baseline, and two older cohorts at the end of Year 4 at baseline). These cohorts will be followed through time at baseline (in 5 intervention schools and 5 control schools); at the end of the intervention year (in the same 5 intervention schools), and at the end of the follow-up year (in the 5 intervention schools and the 5 control schools) (see Table 6).

During these classroom observations, TAs will be asked to carry or wear an audio recording device supplied by IoE (intervention and control schools). The purpose of these recordings will be to supplement and enhance the observations by providing more in-depth data on the nature and quality of the interactions between the TAs and the pupils during the lesson. A small subsample of recordings from intervention and control schools will be selected for analysis.

The staff surveys will be administered to the Headteacher, SLT and all teachers and TAs in the school. The intention with these surveys is to measure school-level changes in practice over time. As such, high response rates will be sought.

Table 6 below shows the methods we propose to use for the process evaluation at each stage of the intervention.

---

<sup>9</sup> For missingness on outcome variables only then standard statistical packages such as Stata use ML for estimating parameters so FIML would not be necessary (Allison, 2012).

**TABLE 6: DATA COLLECTION ACTIVITY**

Activity↓	Year→	Planning Year: 16/17			Intervention year: 17/18			Follow-up year 18/19		
	Term →	T1	T2	T3	T1	T2	T3	T1	T2	T3
Documentary review (change plans developed with LLS consultant)			X	X						
Baseline testing of pupils (KS1 for Y4; ACER ELMs test for YR1)				X <sub>I/C</sub>						
Outcome testing of pupils (KS2 for Y6; ACER ELMs test for YR3)										X <sub>I/C</sub>
Staff surveys (Heads/SLT, teachers and TAs)				X <sub>I/C</sub>			X <sub>I/C</sub>			X <sub>I/C</sub>
Staff interviews – 3-5 interviews per school (5 intervention schools; 5 control schools)					X <sub>I/C</sub>		X <sub>I</sub>			X <sub>I/C</sub>
Classroom observations (5 intervention schools; 5 control schools)					X <sub>I/C</sub>		X <sub>I</sub>			X <sub>I/C</sub>
Audio recordings of TA-pupil interactions					X <sub>I/C</sub>		X <sub>I</sub>			X <sub>I/C</sub>
Training sessions observations in 5 schools (SLT, teacher and TA training)					X <sub>I</sub>	X <sub>I</sub>	X <sub>I</sub>			
Review of consultancy visits reports in 5 schools (2 visits per school)					X <sub>I</sub>	X <sub>I</sub>	X <sub>I</sub>			
Pupil engagement assessment										X <sub>I/C</sub>

*Note: I= only intervention, I/C=both intervention and control*

In addition to these measures, data relevant to the fidelity of implementation will be collected as part of the intervention and shared with the evaluation team (Table 7 in section 2.9.1).

### 2.9.1 PLANNING YEAR (2016/17)

#### ***Theory of change (TOC) workshop***

The evaluation will start with a TOC workshop in which the intervention team and evaluation team further develop the intervention logic. In the TOC workshop it will be important to specify in detail the mechanisms by which the intervention is expected to affect outcomes. The development of the intervention logic will draw on the logic model developed during the two initial set-up meetings held with both teams and the EEF. The TOC will describe how, and under what circumstances, a better deployment of TAs implemented through training sessions, gap tasks and the support from consultants will yield better pupil outcomes (attainment, engagement and independence).

In particular, there are a number of crucial elements of the intervention logic:

- Gap tasks and action plans;
- TA and teacher training sessions
- Strategic input for SLT;
- NLE consultancy visits schools and reporting on progress made;
- Moderator effects:
  - SLT buy-in and ability to ensure school staff buy-in
  - School staff trained and engaged to implement changes to classroom practice
  - Effective project coordination centrally and within each cluster
  - Adequate workforce and good quality training, embedded into school-wide practice
  - Staff turnover (especially TAs leaving)

In addition, it will be important in the TOC workshop to also develop a detailed understanding of: 1) the intervention logic; 2) assumptions for success (and risks threatening this); 3) measurements required by the evaluation at each phase

#### ***Baseline staff survey***

We will implement surveys for school staff (headteachers/SLT, teachers and TAs) across both intervention and control schools as outlined above in Table 6.

The surveys will collect data on usual practices, attitudes and perceptions. For example, they will include:

- Basic demographic information (if not available by other means): gender, years of experience in the school and in total;
- Deployment of TA (reported time teachers and TAs spend with lower-attaining and SEND pupils);
- Reported quality of TA-pupil interactions (such as more utilisation of self-scaffolding strategies);
- Reported practice in support of pupil independence;
- Quality of pre-lesson preparations (reported quality of TAs' pre-lesson preparation; opportunities for and quality of teacher-TA liaison);
- Proportion of TAs trained and attending sessions at school;
- Information on costs associated with intervention
- Facets of teaching quality (HT/SLT and teacher survey only)

The type of questions will be tailored to each type of respondent (headteacher/SLT, teachers and TAs) in each group (treatment and control).

#### ***School visits: Classroom observations, audio recordings and staff interviews***

In order to examine the implementation of the intervention in more depth, we will conduct school visits in five treatment and five control schools at different points in time throughout the intervention (Table 6). The first of these visits will take place at baseline in both treatment and control schools. In practice, these visits will need to take place in September 2017, before the intervention in schools gets underway, to allow the visits to be scheduled after the schools have been recruited and randomised in July 2017. Schools will be selected purposively in such a way as to get a range in terms of prior attainment and proportion of FSM pupils.



To examine the quality of practice, an observation protocol has been developed based on the anticipated changes in practice the MITA intervention is designed to precipitate. Classroom observations will focus on the activities of teachers and TAs, and the frequencies with which pupils spend interacting with the teacher and/or TA. Observation alone tends to be insufficient to the task of obtaining high quality data on what is happening at the talk-level. Therefore, data on TAs' interactions with pupils will be captured via audio recordings of TAs and pupils during the observed lessons. These recordings will be analysed in relation to the scaffolding framework that forms the basis of MPTA training. A coding protocol for the audio recordings of the TA-pupil interactions will be developed by the developer team (UCL-IOE and UEL). The evaluation team will collect the audio recordings and anonymise them before giving them to the developer team for blind transcription and coding. Transcripts will be coded using the categories from the scaffolding framework introduced to TAs (self-scaffolding, prompting, clueing, modelling and correcting). Sub-categories will be used to differentiate between types within these broad categories (for example, different types of prompt). The coding will be carried out by two coders experienced in analysing interactional data, and with strong understanding of the coding categories. Both coders will code small subsample of transcripts independently of one another in order to provide an inter-rater reliability measure.

At baseline, observations and audio recordings will be made in both intervention and control schools. Additionally, three to five interviews per school will take place in the same five treatment and five control schools (interviews in control schools may take place over the phone). These interviews will include the headteacher and/or SLT, teachers and TAs and will explore in more depth the usual practices surrounding TA deployment in the schools.

Finally, NLE consultants will indicate in their checklist if they observe a focus on workforce restructuring, which will be reviewed as part of the analysis of consultancy visits (see below).

#### **INTERVENTION YEAR (2017-18)**

##### ***Observations of training***

During the intervention phase in the first year of the intervention, the evaluation team will observe the training sessions in five schools (intervention schools only, the same selected for the school visits):

- Four ½ day sessions for headteachers/SLTs,
- Two ½ day sessions for TAs,
- Two 2hs sessions for teachers,

##### ***Review of consultancy visits***

The evaluation team will review three NLE consultancy visits per school in each of the same five intervention schools. Each consultancy visit will produce a report where consultants will follow a predetermined format to record and share a summary of each visit; action points for the SLT; and a standard checklist. This checklist will have some open-ended questions and will be collected by IOE along with the summary reports, thus producing a dataset ready for analysis. This file will be analysed by the evaluation team who will triangulate with the other data collection tools. The aim of this tool is to capture data related to implementation fidelity, which will feed into the process evaluation.

##### ***Staff surveys***

We will deliver the second staff survey across all intervention and control schools<sup>10</sup> at the end of the academic year.

##### ***School visits: Classroom observations and staff interviews***

The second visits will take place only in intervention schools in the final term of the 2017/18 year and will focus on identifying any potential changes in TA deployment and practice. The observations will again examine the activities of teachers and TAs in classrooms, and pupil interactions with adults. Observations in the intervention schools will be supplemented with audio recordings of TA-pupil interaction in order to determine any changes in the quality of the interactions following the MPTA training.

---

<sup>10</sup> The staff survey for the control group will serve as a way to keep in touch with them during the intervention year.

Additionally, three to five interviews per school will take place in the same five treatment schools as at baseline. These interviews will include the headteacher and/or SLT, teachers and TAs and will explore in more depth the usual practices surrounding TA deployment in the schools. The headteacher interviews in particular will address the moderating factors as part of the implementation fidelity component.

#### **FOLLOW-UP YEAR (2018-19)**

##### ***Staff surveys***

At the end of the follow-up year, we will conduct the last set of staff surveys in all treatment and control schools. This final survey will include questions on staff's perception of the intervention and its impact (treatment schools only) and will gather data on any TA-related activities since the trial began in the control schools.

##### ***School visits***

The final school visits will take place at the end of the follow-up year in the five treatment and five control schools visited at baseline. The observations will examine the quality of the interactions between TAs, teachers and pupils with an aim to identify, in the treatment group, whether changes in practice have taken place over time. As at baseline, audio recordings of TA-pupil interactions will be made in both intervention and control schools.

Additionally, three to five interviews per school will once again take place in these schools. These interviews will include the headteacher and/or SLT, teachers and TAs and will explore in depth how TAs are deployed in the schools. The headteacher interviews in particular will address the moderating factors as part of the implementation fidelity component.

#### **IMPLEMENTATION FIDELITY MEASURES (THROUGHOUT INTERVENTION)**

As part of the process evaluation, the team will assess measures of implementation fidelity which will be gathered throughout the intervention via instruments used as part of the intervention, as well as via the staff surveys, interviews and observations methods described in the previous section.

In terms of **fidelity** we would expect to see the following moderators (Table 7):

- **Headteacher/SLT buy-in and engagement (responsiveness):** attendance of school staff to MITA sessions, NLE school visits, MPTA staff training.
- **Adherence to the programme:** regular project development team meetings, readiness for MPTA training and completion of gap tasks.
- **Change capacity and environment** (school and staff readiness to make structural and operational changes): this will be observed on how SLT introduce change process, win and maintain staff support and manage resistance.
- **Scale of the challenge:** how far away the school is from exemplary practice, institutional blockers, and how effectively SLT integrate new practices and training into existing processes and practice.
- **Quality of teaching and SEND provision:** levels of teaching confidence and level of training.
- **Capacity** (staff turnover/availability).

Table 7 shows the source information by which each of these moderators will be measured.

TABLE 7: FIDELITY MEASURES

<b><i>HT/SLT buy-in and engagement (responsiveness)</i></b>	
Attendance at all MITA sessions	Register of attendance
Attendance at school visit meetings	School visit checklist
Attendance at MPTA staff training	MPTA training checklist
<b><i>Adherence to the programme</i></b>	
Regular development team meetings	School visit checklist
Readiness for MPTA training	MPTA readiness checklist
% Teachers and TAs completing MPTA training	MPTA trainer checklist
Completion of school visits	School visit checklist
Address all recommendations (I-IV)	School visit checklist
Completion of gap tasks	Audit component checklist Returns (e.g. action plan) School visit checklist
<b><i>Quality of teaching and SEND provision</i></b>	
Levels of teacher confidence, competence, level of training to meet needs of pupils with SEND	Usual practice survey TA Audit

## **2.10 REPORTING**

The findings from this study will be reported in the main study report containing details of the study design, implementation, process data and results for all included pupils. This will be reported in 2020 using the EEF reporting template. Academic journal articles will also be pursued as part of the project outputs.

## **2.11 COSTS**

We will estimate average marginal costs per pupil per year for schools for the intervention. We will do this in three stages: 1<sup>st</sup> staff survey at baseline stage, in the 2<sup>nd</sup> staff survey during the intervention, and in the 3<sup>rd</sup> staff survey at follow-up. Average marginal costs per pupil will be estimated in line with the guidelines followed by other Education Endowment Foundation projects. Separate estimates of pre-requisite costs and of additional staff time will also be provided. These estimates together will provide a general indication of the costs involved with implementing the MITA research, though we recognise that costs will differ between schools and cluster areas.

Specific cost categories on which we will collect data include:

1. The direct costs of staff time. . This will identify separately the cost of new hires, supply staff, and any extensions made to the contract hours of teachers and TAs.
2. Pre-requisite costs (any cost linked to the training sessions, e.g., purchasing additional resources for the sessions, venue hire for MITA SLT sessions).
3. The cost of three components of the intervention (MITA SLT course, MPTA course and LLS-NLE input).
4. Additional (unpaid) staff time (completing surveys, meetings, etc.) reported by teachers and TAs.

The collection of cost data will be done through the survey of headteachers and SLT members.

We will collect additional data from the intervention team to validate the above estimates and to ascertain whether any planned changes if the intervention were to be made more widely available would have cost implications.

### 3. Ethics and registration

Following the EEF guidelines, this evaluation required approval for those institutions involved in the evaluation. The evaluation team obtained ethics approval from the University of Cambridge, Faculty of Education.

Parental opt-out from the additional testing for the evaluation component of the project will be required, with participants' parents or legal guardians being the decision-makers for individual pupils. The main reason for the opt-out approach is that the intervention will be delivered within the school day when schools act *in loco parentis*, and the intervention does not differ from standard practice in schools.

International Standard Randomised Controlled Trial Number (ISRCTN): 33648.

Fair processing of personal data: the project will fulfil the Condition 1 of processing personal data in Schedule 2 of the DPA as the data subjects will give their implicit consent in form of an opt-out letter to parents at the beginning of the trial. The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe. The study was reviewed by the RAND Europe ethics advisory board and approved by the University of Cambridge ethics review process.

RAND Europe adopts good industry practices regarding the protection of personal data as part of its obligations as a Data Controller under the Data Protection Act 1998 and takes appropriate technical and organisational measures conformant with ISO 27001 to protect personal data. Individuals targeted by the study have the right to oppose, have access to, rectify, or remove personal or sensitive personal data held by RAND Europe.

Any data sharing required will be governed by a data sharing agreement.

## 4. Personnel

There are several organisations involved in this evaluation. Table 8 below sets out the organisational roles and responsibilities.

**TABLE 8: ORGANISATIONAL ROLES AND RESPONSIBILITIES FOR MITA EVALUATION**

Organisation	Role	Responsibilities
RAND Europe and University of Cambridge	Evaluators	<ul style="list-style-type: none"> <li>• Overall responsibility for the outcome and process evaluations, trial design, analysis, reporting and quality assurance of the study.</li> <li>• Responsibility for oversight over the third party administering assessor blinded testing for all pupils in treatment and control schools</li> </ul>
UCL-IOE	Recruitment & intervention provider	<ul style="list-style-type: none"> <li>• Implementation of MITA through:               <ul style="list-style-type: none"> <li>○ Training of SLTs</li> <li>○ Training of teachers</li> <li>○ Training of TAs</li> <li>○ School visits by consultants</li> </ul> </li> <li>• Recruiting schools, teachers and TAs within the constraints placed on the trial.</li> <li>• Providing schools, teachers and TAs with background information to the trial.</li> <li>• Regular reporting on recruitment progress.</li> <li>• Collating a list of schools and pupils included in the trial that allow for later secure data linkage.</li> <li>• Creating and maintaining a key contact list for all schools during the study and sharing this with the evaluator team.</li> <li>• Facilitating access to schools for evaluator team fieldwork.</li> <li>• Reporting on implementation of MITA.</li> <li>• Distributing links to online surveys to teachers and headteachers.</li> <li>• Transcriptions and coding of audio recordings of TA-pupil interactions.</li> </ul>
Education Endowment Foundation	Funder	<ul style="list-style-type: none"> <li>• Funding RAND and intervention team.</li> <li>• Facilitating set-up phase.</li> <li>• Reviewing reports.</li> <li>• Signing off on evaluation.</li> </ul>

#### 4.1 TEAM

*Project Leader:* Dr. Alex Sutherland (RAND) | *Project Manager:* Rob Prideaux (RAND)

*Core fieldwork and analysis team:* Sonia Ilie, Sara Curran and Brier Rigby-Dames (Cambridge) | Miaoqing Yang (RAND) | Miriam Broeks (RAND)

*Project advisors:* Dr. Julie Bélanger (co-lead and advisor, RAND) | Janna van Belle (RAND, RCTs)

## 5. Risks

Risk	Assessment	Mitigation strategy
<b>Recruitment failure</b>	Likelihood: Moderate Impact: High	UCL-IOE to recruit schools and be the main contact for schools. UCL-IOE to allow sufficient window for recruitment, as well as strong network with cluster areas and other local partners Clear information about benefits of participation. UCL-IOE to recruit schools from within their personal networks of schools.
<b>Attrition</b>	Likelihood: Moderate Impact: Moderate to high	Clear information about expectations and requirements provided to participating schools. MoU to be signed with participating schools Intention to Treat (ITT) analysis to be used. Attrition to be monitored and reported according to CONSORT guidelines (Campbell et al., 2010). Strong communication strategy to keep schools engaged through Year 2.
<b>Missing data</b>	Likelihood: Moderate Impact: Moderate	Use of administrative data for all participating pupils.
<b>Pupil mobility</b>	Likelihood: Moderate Impact: Low	Pupils who are included in the study at the start of the school year and who move between study schools will be retained and analysed according to their original allocation to treatment / control. Pupils who migrate to non-study schools will be excluded from the analysis as these pupils will be tested with external tests. In the event that mobility to non-study schools exceeds 10% on average across all schools, then the evaluators will discuss with the EEF the possibility of additional funding to collect this information.
<b>Low implementation fidelity</b>	Likelihood: Low to moderate Impact: Moderate	Process evaluation to monitor and document fidelity of implementation (this is mainly a risk for headteachers, teachers, TAs). Consultant to maintain engagement and contact during intervention and implementation phase
<b>Cross-contamination</b>	Likelihood: Low Impact: High	Year group randomisation to be used. All teachers within the same year group will be randomised to the same treatment condition, limiting the possibility for contamination.
<b>Evaluation team members absence or turn-over</b>	Likelihood: Moderate Impact: Low	All RAND staff have a three month notice period to allow sufficient time for handover. The team can be supplemented by researchers with experience in evaluation from the larger RAND Europe pool.
<b>Low response rates for surveys</b>	Likelihood: Moderate Impact: Moderate	Surveys to be kept to a maximum of 10-15 minutes long. Respondents given the opportunity to complete survey online on multiple occasions if required. Sufficient data collection window given with real-time monitoring of response rates to allow for reminders to be targeted.
<b>Lack of coordination with larger teams (RAND,</b>	Likelihood: Moderate Impact: Moderate	Teams to attend initial meetings and agree on roles and responsibilities at the outset. Regular updates to be provided to the lead evaluators.

## 6. Data protection statement

RAND Europe adopts good industry practices regarding the protection of personal data as part of its obligations as a Data Controller under the Data Protection Act 1998 and takes appropriate technical and organisational measures conformant with ISO 27001 to protect personal data. Respondents to this survey have the right to oppose, have access to, rectify, or remove personal or sensitive personal data held by RAND Europe.

## 7. Timeline

Date	Activity
<b>Sept - October 2016</b>	Evaluation kick-off meetings
<b>Dec 2016 – June 2017</b>	Recruiting schools
<b>May – June 2017</b>	Baseline data collection and external testing of pupils in Year 1
<b>Late July 2017</b>	Randomisation
<b>Sept 2017</b>	Baseline interviews and classroom observations
<b>Sept 2017- June 2018</b>	Intervention phase (training sessions to school staff)
<b>Sept 2017- June 2018</b>	Mid-intervention data collection (surveys / interviews / classroom observation)
<b>Sept 2018 – June 2019</b>	Implementation and action plan in school phase
<b>May – June 2019</b>	End of implementation data collection
<b>June 2019</b>	Pupils outcome measures, including external testing of pupils in Year 3
<b>December 2020</b>	Reporting

## 8. Conflict of interest statement

None of the evaluation team has any conflicts of interest and all members of the study team have approved this protocol prior to publication.

Rob Webster and Paula Bosanquet have authored books relating to TA deployment that will be purchased for schools as part of the project. The authors receive royalties for each copy sold.

## References

- Alborz, A., Pearson, D., Farrell, P. and Howes, A., 2009. The impact of adult support staff on pupils and mainstream schools: A systematic review of evidence. London: Department for Children, Schools, and Families.
- Allison, N. 2012. Why maximum likelihood is better than multiple imputation. *Statistical Horizons*.
- Blatchford, P., Bassett, P., Brown, P. and Webster, R., 2009. The effect of support staff on pupil engagement and individual attention. *British Educational Research Journal*, 35(5), pp.661-686.
- Blatchford, P., Bassett, P., Brown, P., Martin, C., Russell, A. and Webster, R., 2007. Deployment and impact of support staff in schools: report on findings from the second national questionnaire survey of schools, support staff and teachers (Strand 1, Wave 2, 2006).
- Bosanquet, P., Radford, J. and Webster, R. 2015. *The Teaching Assistant's Guide to Effective Interaction: How to Maximise Your Practice*. London: Routledge.
- Bruhn, M. and McKenzie, D. 2009. In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, 1, 200-232.
- Clay, M.M., 2000. Concepts about print: What have children learned about the way we print language?. Heinemann.
- Dong, N. and Maynard, r. 2013. PowerUP!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6, 24-67.
- Greevy, R., Lu, B., Silber, J. H. and Rosenbaum, P. 2004. Optimal multivariate matching before randomization. *Biostatistics*, 5, 263-275.
- Higgins, J., Altman, D., Gøtzsche, P., Jüni, P., Moher, D., Oxman, A., Savovic, J., Schulz, K., Weeks, I. and Sterne, j. 2011. Cochrane bias Methods Group, Cochrane Statistical Methods Group (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928.
- Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. 2015. Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature.
- Morgan, K. L. and Rubin, D. B. 2012. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 1263-1282.
- Paris, S.G., 2005. Reinterpreting the development of reading skills. *Reading research quarterly*, 40(2), pp.184-202.
- Radford, J., Bosanquet, P., Webster, R. and Blatchford, P., 2015. Scaffolding learning for independence: Clarifying teacher and teaching assistant roles for children with special educational needs. *Learning and Instruction*, 36, pp.1-10.
- Rubin, D. B. 2012. Comment: The design and analysis of gold standard randomized experiments. *Journal of the American Statistical Association*.
- Sharples, J., Webster, R. and Blatchford, P., 2015. *Making Best Use of Teaching Assistants*. London: Education Endowment Foundation.
- Skinner, E. A., Kindermann, T. A., Furrer, C. J. (2008). A motivational perspective on engagement and dissatisfaction: Conceptualization and assessment of children's behavioural and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69 (3), 493-525.
- Snijders, T. A. B. and Bosker, R. J. (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling* (2nd edition). London: Sage Publishers.
- Taves, D. R. 1974. Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology & Therapeutics*, 15, 443-453.
- Webster, R., Russell, A. and Blatchford, P. 2015. *Maximising the impact of teaching assistants: Guidance for school leaders and teachers*. London: Routledge.
- Worrall, J. 2010. Evidence: philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice*, 16, 356-362.



## Appendix 1: Note on randomisation

### Stratification

Stratification is only possible on a limited number of variables. In practice, there may be many other variables that we desire to achieve balance on. One issue is that the more such variables there are, the more difficult achieving balance is unless one has many units to randomise.<sup>11</sup> Even in the simplest case of binary variables, the number of unique combinations quickly increases, thereby increasing the chance that some cells will be empty.<sup>12</sup> When baseline data is available, some other methods that allow the balance to be sought on multiple variables are recommended, including pair-wise matching (Greevy et al., 2004) and covariate adaptive randomisation (e.g. minimisation; see Taves 1974).

### Re-randomisation process step by step

- Collect covariate data and select covariates for the re-randomisation.
- Run 1,000 randomisations and assess the probability of imbalance for each stratifying measure.
- If the probability is deemed too great then proceed with randomisation.
- Specify a balance criterion determining whether a randomisation is acceptable.<sup>13</sup>
- Randomise the units to treatment groups.
- Check the balance criterion; if the criterion is met, go to the next step. Otherwise, return to the previous step to conduct re-randomisation.
- Conduct the experiment using the final randomisation.
- Analyse the results with a randomisation test.

If the random assignment to treatments produces some p-value above the desired level for any of the balance variables, a new randomisation is conducted. The process is repeated until all of the sampling variables are balanced. Covariate balance should be checked before the intervention has started, and if lack of balance is noted, a re-randomisation needs to be conducted until satisfied. Re-randomisation improves covariate balance between the treatment and control groups and if the covariate is correlated with the outcome, it provides us with more precise treatment effects.

---

<sup>11</sup> With  $k$  independent covariates, the chance of a significant difference between treatment and control groups for at least one covariate is  $1 - (1 - \alpha)^k$ , at significance level  $\alpha$ . For example, if we have 10 covariates and a 5% significance level, the probability is 40% (Morgan & Rubin, 2012). MORGAN, K. L. & RUBIN, D. B. 2012. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 1263-1282..

<sup>12</sup> Specifically, there are  $2^k$  unique combinations for  $k$  binary variables. E.g. with four variables there are 16 combinations ( $2^4$ ), but 256 with eight binary variables ( $2^8$ ).

<sup>13</sup> One way to set the criteria for acceptable balance is to use Mahalanobis Distance,  $M$ , to represent multivariate distance between group means. A randomisation is acceptable whenever  $M$  falls below a certain threshold,  $a$  (Morgan & Rubin, 2012). MORGAN, K. L. & RUBIN, D. B. 2012. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 1263-1282..

## Appendix 2: ACER pilot testing

### PILOT STUDY FOR THE MITA EVALUATION STUDY REPORT

Authors: ACER

Date: 28<sup>th</sup> April 2017

#### Purpose

To investigate which test, *ELMs Reading mid-year 1* test or *ELMs Reading Y2* test, would be more appropriate for use in the MITA evaluation study with Y1 pupils. *ELMs Reading Y2* can be rendered to print format, whereas *ELMs Reading mid-year 1* test is not suitable for use on paper.

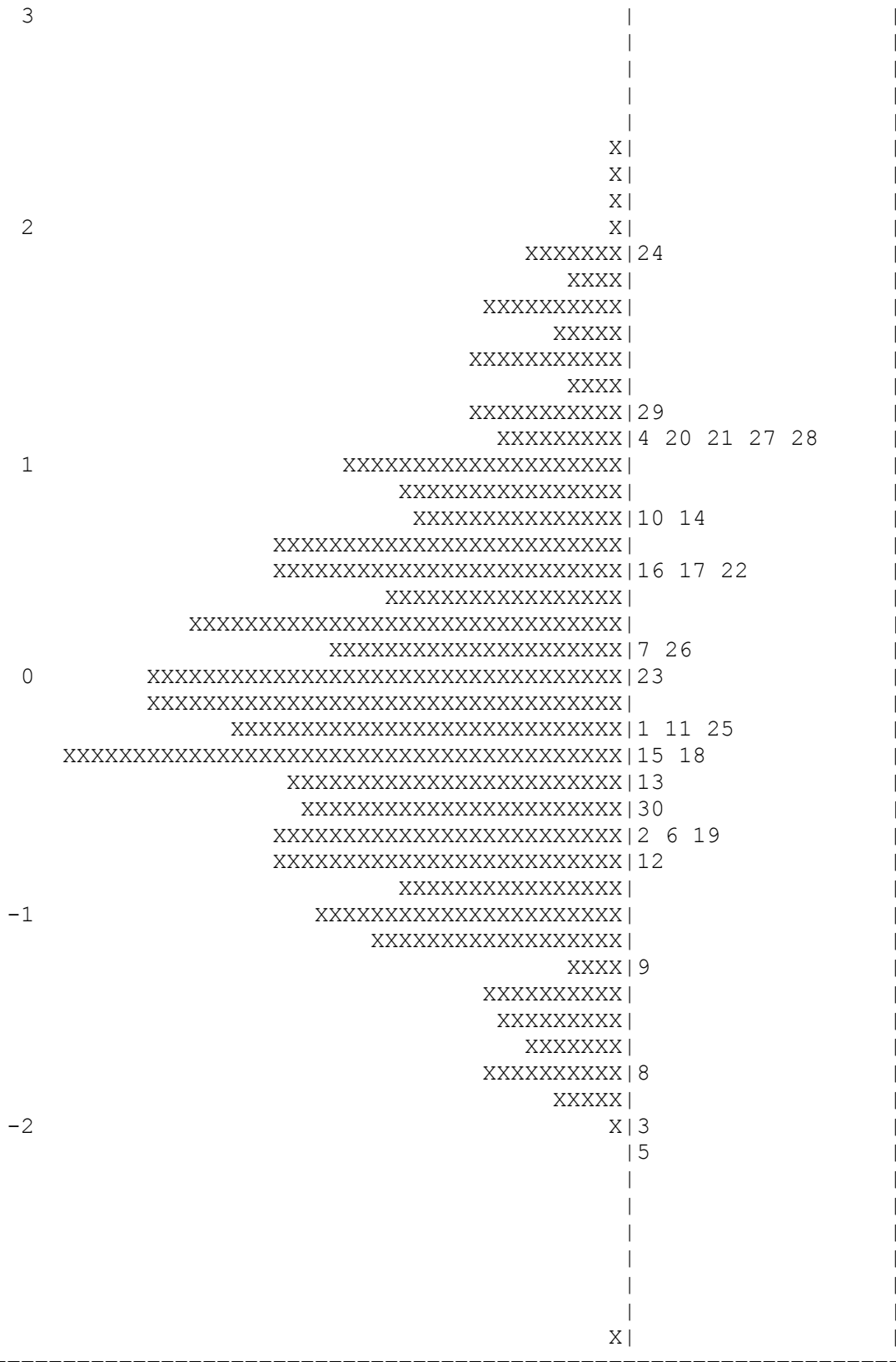
#### Method

Between 30<sup>th</sup> March and 5<sup>th</sup> April 2017 up to three classes of mixed-ability Year 1 children were assessed in each of the participating three primary schools. The two test forms were randomly allocated to the participating cohort, so that approximately half the children in each school took the mid-year 1 and half the Y2 test. ACER staff administered and invigilated the testing.

ACER psychometricians analysed the results for item and test difficulty. The results are available as item maps in Figures 3 and 4.

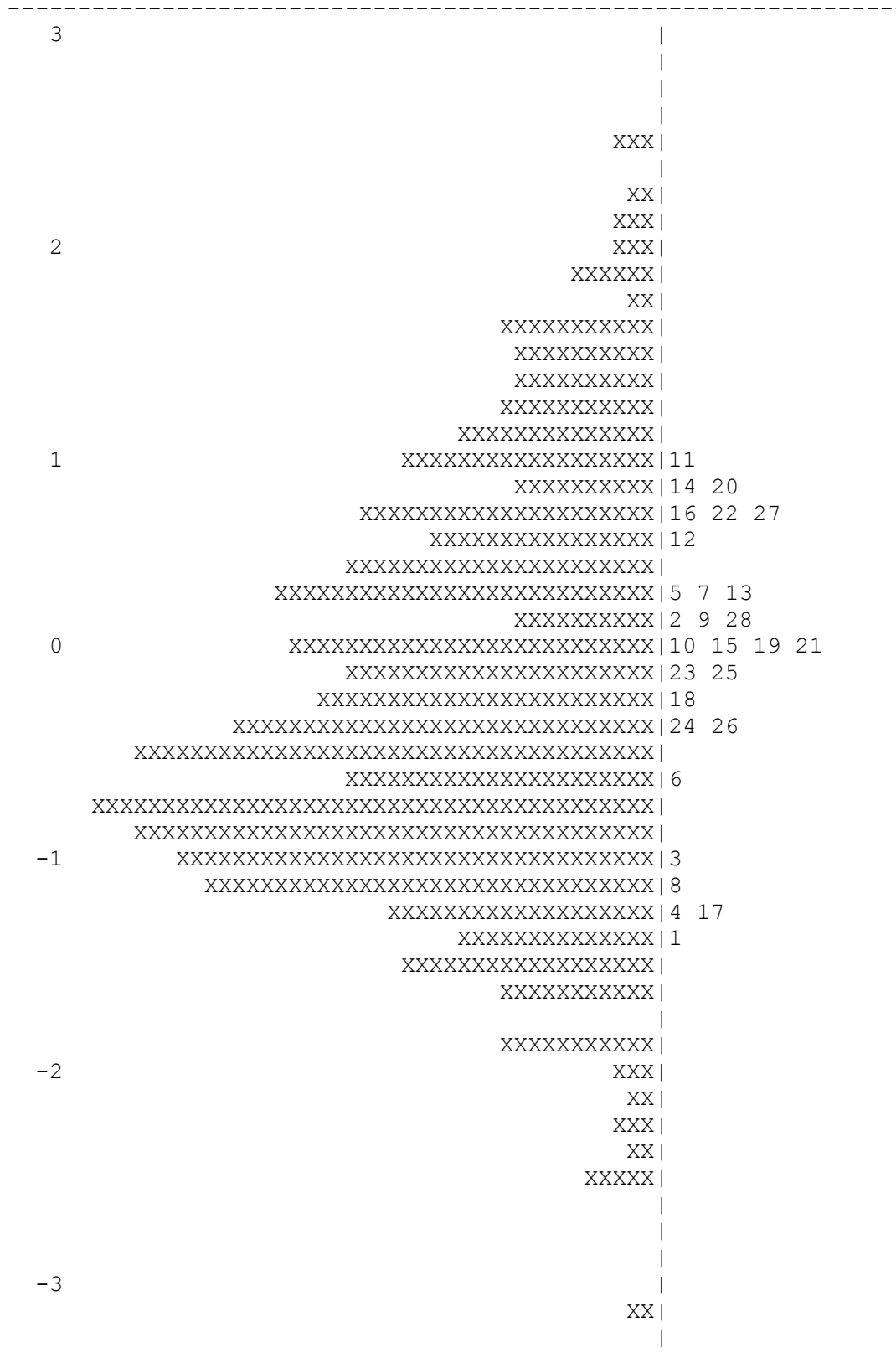
	<b>School type</b>	<b>No of participating pupils</b>	<b>% SEN (nat. ave 2.6%)</b>	<b>EAL (nat. av. 20%)</b>	<b>FSM (nat. 25.4%)</b>
School 1 South London	Mixed community school	90 (three classes)	0.8%	44.7%	13.1%
School 2 Reading	Mixed community school	60 (two classes)	0.5%	9.2%	20%
School 3 Reading	Mixed community school	60 (two classes)	2%	11.4%	4%

**FIGURE 3: MAP OF ELMS READING MID-YEAR 1 TEST SCORES**



Each 'X' represents 0.2 cases

**FIGURE 4: MAP OF ELMs READING YEAR 2 TEST SCORES**



Each 'X' represents 0.1 cases