<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>Teaching Assistant Campaign – Advocacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEVELOPER</td>
<td>Education Endowment Foundation</td>
</tr>
<tr>
<td>EVALUATOR</td>
<td>Institute for Fiscal Studies</td>
</tr>
<tr>
<td>STUDY</td>
<td></td>
</tr>
<tr>
<td>STATISTICIAN</td>
<td></td>
</tr>
<tr>
<td>STUDY CHIEF</td>
<td></td>
</tr>
<tr>
<td>INVESTIGATOR</td>
<td></td>
</tr>
<tr>
<td>SAP AUTHOR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Luke Sibieta</td>
</tr>
<tr>
<td>SAP VERSION</td>
<td>1</td>
</tr>
<tr>
<td>DATE</td>
<td>13/11/2017</td>
</tr>
<tr>
<td>EEF DATE OF</td>
<td>13/11/2017</td>
</tr>
<tr>
<td>APPROVAL</td>
<td></td>
</tr>
</tbody>
</table>
# Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Study design</td>
<td>4</td>
</tr>
<tr>
<td>Protocol and SAP changes</td>
<td>5</td>
</tr>
<tr>
<td>Sampling frame</td>
<td>5</td>
</tr>
<tr>
<td>Details of synthetic control approach</td>
<td>6</td>
</tr>
<tr>
<td>Outcome measures</td>
<td>8</td>
</tr>
<tr>
<td>Analysis</td>
<td>9</td>
</tr>
<tr>
<td>Report tables</td>
<td>12</td>
</tr>
</tbody>
</table>
Introduction

Teaching Assistants (TAs) make up approximately one quarter of the school workforce, yet the evidence summarised in the EEF Teaching and Learning Toolkit and elsewhere suggests that the current deployment of TAs in English schools is not improving pupil outcomes, and can even be detrimental to learning. TAs spend a disproportionately large amount of time with disadvantaged pupils, so the attainment of the EEF’s target group is particularly vulnerable to harmful practice. However, there is now a wide evidence base suggesting better ways TAs can be deployed. For example, when TAs are used to provide one-to-one or small group support using structured interventions, they can generate average gains of between three and four additional months (e.g. Switch-on Reading programme, Gorard et al., 2014, and the Catch Up Numeracy programme, NfER, 2014).

In summer 2015, EEF produced a guidance document setting out ways that schools maximise the impact of teaching assistants and launched a multi-stranded campaign to increase uptake of this advice and guidance, including:

1. National campaign materials, such as sending hard copies of the guidance report to all schools, emailing all schools with links to the guidance, various events and social media.

2. Advocacy support provided in a range of ways (such as training events, coaching, and consultancy) to primary schools in South and West Yorkshire. This was particularly aimed at underperforming schools, and focuses on enabling schools to act on the guidance.

3. Targeted interventions working with schools with high numbers of disadvantaged pupils within South and West Yorkshire. This involved evaluation grants, scoping grants and scale-up grants provided by the EEF; each of which will be subject to a separate evaluation.

The focus of this evaluation is the impact of the second strand of the campaign: the impact of advocacy and training in South and West Yorkshire on pupil attainment. Unfortunately, however, one cannot use experimental methods to address this question as the whole of South and West Yorkshire represents a single treatment group. South and West Yorkshire was a non-random choice and may be subject to different time trends as compared with other similar sized regions or England as a whole. One cannot use a matched comparison group based on contemporary information either as South and West Yorkshire may be subject to particular time trends. In principle, one requires a counterfactual group with similar characteristics and subject to similar time trends. In order to do this, we make use of synthetic control methods (precise details described below).

Given that the advocacy and training took place at the same time as the national campaign (strand 1 above), the synthetic control group will be one that is potentially more informed about better ways to use TAs than would otherwise have been the case. The impact of national campaign materials is the subject of separate EEF impact analysis. However, it is worth stating that the experimental variations to the national campaign (different forms of emails and incentives to engage with the information) had relatively low impact on levels of engagement with the TA campaign and are thus unlikely to be any source of bias in our impact analysis of advocacy and training on pupil attainment in South and West Yorkshire.

Although our main focus is the impact of advocacy and training, because additional interventions were also targeted at schools in South and West Yorkshire (strand 3 above), one cannot focus on the impact of the advocacy element without also considering the interactive effect of these additional targeted interventions. Indeed, our main analysis will combine the effects of advocacy with any effects of additional interventions targeted at South and West Yorkshire over the period. We will therefore undertake a small element of analysis where we compare the outcomes of schools in South and West Yorkshire that received
advocacy and training with those that received targeted interventions in addition. This is less robust, but should be indicative as to the degree to which the effects we observe are due to targeted interventions.

Synthetic control methods have not been used in previous EEF evaluations and the authors have not used them in previous work either. As a result, it is hard to pre-specify all analysis. Some of the analysis will necessarily be exploratory as we gain more experience in using these methods. We have sought to pre-specify as much of the proposed approach and analysis as is possible at this stage.

Study design

Advocacy and training support was offered to all 1,049 primary schools in South and West Yorkshire\(^1\) with the aim of working with approximately 525 schools. The support was provided by a set of around 5-10 intermediaries, who bid to EEF for the contracts to undertake the advice and training. Support consisted of a range of activities, such as training events, coaching, and consultancy. It was particularly aimed at underperforming schools, and focuses on enabling schools to act on the guidance.

As the treatment unit essentially represented all nine local authorities in South and West Yorkshire, we were unable to use an experimental or quasi-experimental designs using matching methods. We instead adopted a synthetic control approach.

The synthetic control method was first proposed by Abadie et al. (2003) in the context of estimating the effects of Basque Terrorism, further applied to assess the impact of a tobacco control law in California (Abadie et al., 2010) and of German reunification (Abadie et al., 2014). From a ‘donor pool’ of aggregate units (local authorities in our setting) not affected by the policy, a synthetic control group is constructed as the weighted average of those units that best resemble the treatment unit in terms of pre-treatment characteristics and outcomes for as long a time period as possible.

The context of the advocacy component is well suited to this kind of analysis as we are able to ensure comparability of pre-treatment outcomes stretching as far back as 2002. In particular, we will take the other 143 local authorities as our potential donor pool and find the set of time-constant non-negative weights across local authorities that best allows us to match pre-treatment trends in the primary outcome of interest (average Key Stage 2 fine points scores in Maths and English) in South and West Yorkshire. The argument to constraint weights to non-negative values that add up to one is to limit extrapolation between the observed cases and the synthetic control generated (Abadie et al. 2010).

To take a simple example, the weights could take the form of 0.5 for Manchester, 0.25 for Liverpool and 0.25 for York. These would then be used to re-create a synthetic comparison group for South and West Yorkshire back to 2002, which would have time trends in the pre-treatment outcomes that are as similar as possible. We would then apply these weights to the post-treatment outcomes across local authorities and compare this with the post-treatment outcomes in South and West Yorkshire. This will give an estimate of the effect of the advocacy campaign.

We will construct the synthetic control group(s) based on local-authority average pupil attainment and will do so before accessing post-test outcome data. This helps to build credibility that the control group was not constructed based on the final outcomes.

---

\(^1\) Sheffield, Rotherham, Doncaster, Barnsley, Leeds, Wakefield, Calderdale, Kirklees, Bradford
Protocol and SAP changes

There two main changes to the protocol along the course of the trial:

- The date of the primary outcome of interest was moved from one year after the start of the advocacy campaign (summer 2016) to two years after (summer 2017). It was felt one year was not long enough for the advocacy campaign to have any effect on pupil attainment. TA deployment decisions are generally made with a relatively long lag.
- Previous version of the protocol proposed creating two synthetic control groups in order to analyse the advocacy component against different baselines. The first was to be constructed using local authority averages based on pure information control schools from the guidance material experiment (i.e. calculating local authority averages amongst only schools in the control group for the national campaign), a business-as-usual baseline. The second was to be based on the group of schools receiving the micro-incentives, a baseline of (potentially) more actively engaged schools. We will still undertake this analysis. However, as the results of this separate experiment suggest that micro-incentives had small, negative impact on information engagement, it will not form a major part of the analysis.

Sampling frame

The main comparison will be between a treatment group consisting of all primary schools in South and West Yorkshire and a control group consisting of primary schools outside of South and West Yorkshire. The average outcome for South and West Yorkshire will represent the (pupil-weighted) average across the nine local authorities. The synthetic control outcome will represent the average across other local authorities in England, weighted by the synthetic control weights (calculation of which is detailed below). We will exclude the following groups of local authorities for reasons as specified:

- Isles of Scilly and City of London (both unusual and small)
- Inner London (very different trends and levels for attainment)

Using the Office for National Statistics (ONS) statistical definition of inner London\(^2\), this leaves us with 128 local authorities out of an initial donor pool of 143 before these exclusions.

Since the treatments were strongly focused on schools within South and West Yorkshire, we judge that spillover effects are likely to be small and include neighbouring local authorities in the donor pool. This is important as neighbouring local authorities are likely to be more similar to South and West Yorkshire than those further away.

A number of alternative samples will be constructed to analyse the likely impact of advocacy and guidance over and above additional treatments occurring as part of the wider TA campaign.

First, a large number of schools outside of South and West Yorkshire received micro-incentives to engage with the information in the TA campaign. To compare this with the effect of the advocacy and guidance, we will compare the average outcome of schools in South and West Yorkshire that did and did not received such incentives.

---

\(^2\) https://www.ons.gov.uk/methodology/geography/ukgeographies/administrativegeography/england
Second, some schools in South and West Yorkshire also received additional TA-led interventions. To benchmark the likely impact of these additional interventions on the average outcome for South and West Yorkshire, we will describe the outcomes of schools in South and West Yorkshire who did and did not receive these additional interventions (EEF to provide list of schools receiving additional TA-led interventions).

**Details of synthetic control approach**

In order to undertake the synthetic control approach, one needs to construct time-constant weights for each local authority in the donor pool that best approximate how outcomes for the treatment group would have evolved in the absence of the treatment.

To do so, we will use pupil-level data from the National Pupil Database (Key Stage 2 results data linked to pupil census data) from 2002-03 to 2016-17. The period from 2002-03 to 2014-15 will be classified as the pre-treatment phase and 2015-16 to 2016-17 as the intervention period, with 2016-17 representing our primary outcome date of interest.

Outcomes for the treatment group will represent the average (pupil-weighted) outcome across all nine local authorities in South and West Yorkshire.

The control group outcome will consist of a weighted-outcome from the donor pool of all other local authorities in England (excluding inner London, City of London and Isles of Scilly). The weights will be calculated at the local authority level to best approximate the evolution of the lagged outcomes and characteristics of South and West Yorkshire over the pre-intervention phase. In particular, our main measure to assess the quality of the match will be the mean squared prediction error (MSPE) between South and West Yorkshire and the synthetic controls in the lagged outcome. The resultant weights may differ across our two main primary outcomes and our two secondary threshold outcomes, but will be identical for the 2015-16 equivalents of our primary outcomes (i.e. weights will be the same for KS2 English fine point scores in 2015-16 and 2016-17). The fact that the weights are different across different outcomes is desirable as it is likely that different underlying factors drive different outcomes and, as a result, different sets of local authorities are likely to serve as suitable controls for different outcomes. One is seeking to find the best synthetic control for each specific outcome.

Although our main assessment of the match quality will be based on the MSPE in lagged outcomes, it is often helpful to seek a balance on wider characteristics that predict levels and trends in outcomes. We will select these from the list of control variables specified below. The lagged outcomes are the most important one to match, so we include all years of data here. We then also include a range of characteristics that are likely to affect attainment trends over time, including deprivation, ethnicity and language. We also consider recent institutional factors likely to influence the degree to which schools can act on TA guidance (recent use of TAs and proportion of Academies). The full set of predictor variables is listed below.

**Predictor Variables**

- KS2 English/Reading Fine Points Score for 2002-03 to 2014-15
- KS2 Maths Fine Points Score for 2002-03 to 2014-15
- %FSM average over period
Performing synthetic control analysis effectively involves choosing two sets of weights: 1) the weights applied to each donor unit (here, local authorities); and 2), the weight applied to each predictor variable (list above). The first set of weights can be found as the solution to an optimisation problem, but the latter is effectively a subjective decision. We will follow a similar process to that laid out in Abadie (2003) for the choosing the second set of weights. In particular, each set of variable weights (call this V) is associated with a set of weights at the donor unit level (call this W). We will choose the V that achieves the W which is associated with the lowest MSPE on a subset of the predictor variables. This procedure is more efficient and performs better than simply choosing the weights that achieve the lowest MSPE on the subset of variables (Abadie, 2003).

The initial subset of predictor variables we will use for choosing V is as follows:

**Initial specification**

- %FSM average over period
- %White-British average over period, %Black average over period
- %Asian average over period
- %EAL average over period
- TA ratio average for 2010-11 to 2014-15
- % pupils in Academies 2010-11 to 2014-15

For robustness, we will then calculate the initial specification MSPE and compare this against a range of alternatives calculated as follows:

- **Vary specification** – increase number of control variables (e.g. more years of lagged outcomes) and reducing number of variables that are not lagged outcomes.
- **Vary size of donor pool** – excluding local authorities from donor pool where outcomes are x% greater or less than South and West Yorkshire on average over the pre-intervention phase or where change over time is outside a given tolerance thresholds. Both thresholds will be chosen with reference to data and the donor pool will not be reduced below a threshold of 18 local authorities (double the number in the treatment group). As an example, previous work (Abadie et al, 2010) has excluded units where the MSPE is

---

3 In principle, there is an infinite set of V weights. We will therefore employ a normalisation such that the Eucliedean norm of the weights is equal to one.
two, five and twenty times greater than the MSPE in the treatment unit. Checking sensitivity of the MSPE to the size of the donor pool is important as the number of donor units in previous cases has only ever varied between 12 and 38 (as compared with over 128 in our case).

In each case, we will report a minimum of 5 different specifications and sizes of the donor pool. These will be calculated in tandem, giving a total of 25 different alternatives.

The range of specifications and donor pools we analyse will be reported in a later version of this SAP, together with their MSPE. The final specification used will be that with the lowest MSPE.

Although the selection of the alternative specifications will be ad hoc, we will be considering a large range of specifications (at least 25). We will also undertake this analysis before we have access to final outcome data and we will therefore not be able to discern any effect of the chosen weights on our final impact analysis.

Furthermore, we will compare the MSPE in our final specification with the difference in outcomes across the distribution of local authorities, e.g. the standard deviation across local authorities and interquartile range.

**Outcome measures**

**Primary outcome**

Our primary outcomes will be as follows:

- KS2 English Fine Points Score 2016-17
- KS2 Maths Fine Points Score 2016-17

Both outcomes will be standardised amongst all pupils taking KS2 tests in 2016-17 and then averaged at the local-authority level. The primary outcome for the treatment group will be the (pupil-weighted) average across all 9 local authorities in South and West Yorkshire and the control outcome will be the average across the synthetic control group (i.e. local authorities weighted by the final synthetic control weights).

There have been reforms to Key Stage 2 assessments throughout the period of study, including in 2016-17 (year 2 of the outcomes). However, these reforms will affect synthetic control and treatment units in similar ways, particularly as we will have chosen a synthetic control group that is likely to react in a similar way to assessment reforms as the treatment group.

**Secondary outcomes**

We will also analyse the following secondary outcomes derived from KS2 test scores:

- Proportion achieving at least the expected level in KS2 English in 2016-17
- Proportion achieving at least the expected level in KS2 Maths in 2016-17
- KS2 English Fine Points Score 2015-16
- KS2 Maths Fine Points Score 2015-16
The first two secondary outcomes seek to test whether there is a greater impact lower down the KS2 distribution as it is likely to be struggling pupils who are the biggest focus of TAs. The second set of secondary outcomes test for early impact one year, rather than two years, after the start of the intervention.

Importantly, and as noted above, the two threshold outcomes could have different synthetic control weights than the primary outcomes.

**Analysis**

**Primary intention-to-treat (ITT) analysis**

The primary ITT analysis will comprise the raw difference between the standardised primary aggregate outcomes for South and West Yorkshire and the synthetic control group. These figures will be calculated as described in the outcomes section above and will not account for any covariates (beyond those used to calculate the synthetic control weights). Results will be presented graphically, with numbers quoted in the text and graph. Analysis will be conducted using Stata 14.

This difference will capture the combined effect of advocacy and training in South and West Yorkshire, as well as the additional effect of targeted interventions. The baseline is likely to be more informed about effective uses of Teaching Assistants than would be the case in earlier years as a result of the National Campaign.

**Imbalance**

Balance will be assessed by comparing the level of the match assessment variables listed earlier in the SAP across South and West Yorkshire and the weighted Synthetic Control group. A key overall metric used to assess overall balance will be the mean squared prediction error (which we will compare with existing estimates in the literature). As additional comparators, we will also include the (pupil-weighed) average across local authorities with non-zero synthetic control weights and the national average. These will provide an indication of the effect of the synthetic control approach in achieving balance between the treatment and synthetic control groups.

**Robustness checks/sensitivity analysis of synthetic control approach**

We will compare our primary ITT analysis with a number of different approaches to both check the sensitivity of our results to different assumptions and show the effect of the synthetic control approach. In particular, we will compare the results against the following alternative ways of calculating the counterfactual outcome:

- 3 alternative synthetic control groups with the next lowest levels of MSPE
- Pupil-weighted average across local authorities with positive synthetic control weights
- National average, i.e. raw baseline
- Mahalanobis distance matching
- Kernel Propensity Score Matching
- Leave-one-out SC estimate (the idea is to iteratively exclude one LA which had received positive weight; this loses some goodness of fit but can help assess whether the results hinge on a specific LA being included).
As part of these analyses we will show the estimated difference in the primary outcomes compared with South and West Yorkshire, but also the achieved balance in match assessment variables and evolution of the lagged outcomes.

The two matching estimators will be undertaken at school-level with contemporaneous equivalents of the control variables (value of all matching assessment variables one year before the treatment phases). As already noted, these are likely to be less reliable than the synthetic control approach in principle, but will be useful to show the extra value of the Synthetic Control approach in practice.

**Comparisons with other elements of the TA campaign**

In addition to advocacy and guidance, the TA campaign also involved a national campaign (with some schools receiving micro-incentives to engage with information) and additional TA-led interventions. We will create a number of different baselines that allow us to benchmark the likely effect of advocacy and guidance with these other elements of the campaign.

First, we will create an alternative counterfactual amongst the synthetic control group based on schools that received and did not receive micro-incentives to engage with the national campaign. As the impact of the incentives on engagement was generally small and negative, we do not expect substantial differences in pupil attainment between these two groups. We will use identical weights to the primary ITT analysis as they pertain to periods before any type of treatment was implemented. This approach is valid under the assumption that the two groups of schools would have evolved in the same manner in the absence of the treatment differences. This is an appealing assumption given that we randomised across almost all primary schools in England.

Second, we will show the outcomes for schools in South and West Yorkshire offered advocacy and training split by whether or not they received additional interventions. This is not a robust approach, but it will be indicative as to the extent to which additional interventions could be driving some of the estimated difference in outcomes between South and West Yorkshire and the synthetic control group.

**Secondary outcome analyses**

The secondary ITT analysis will comprise the raw difference between the standardised secondary aggregate outcomes for South and West Yorkshire and the synthetic control group. These figures will be calculated as described in the outcomes section above and will not account for any covariates (beyond using them to construct the weights). Results will be presented graphically, with numbers quoted in the text and graph. Analysis will be conducted using Stata 14.

**Subgroup analyses**

Subject to data approval, we also seek to perform sub-group analysis for pupils eligible for free school meals and those with English as an Additional Language (both key targets groups of TAs). This will proceed in an identical fashion to the overall analysis except that the outcomes for each local authority will be defined for by those of the sub-group in question.
**EFFECT SIZE CALCULATION**

Continuous outcomes will be standardised prior to analysis and raw differences will thus be interpreted in effect size terms.

The secondary threshold outcomes (i.e. achieving expected levels in Maths and English) will not be standardised as these outcomes are already readily understandable by schools in their raw form.

**INFERENCE PROCEDURE**

Although there is no uncertainty in terms of the values of the aggregate units (in this case, average attainment by local authority), there is uncertainty as to whether the synthetic control group best approximates the missing counterfactual for the treatment group. Inference cannot therefore be conducted in standard ways. Instead, Abadie et al. (2010) propose a placebo-test method, whereby each aggregate unit within the ‘donor pool’ is in turn assigned as the treatment group in order to create a distribution of treatment effects. We can then compare the estimated treatment effect of the actual intervention with the distribution of placebo treatment effects to ascertain whether the former is large and calculate a p-value.

As part of the approach proposed by Abadie et al. (2010), we will undertake the inference procedure in the following steps:

- Show the synthetic control estimates for all other donor LAs (with South and West Yorkshire back in donor pool)
- Exclude placebo LAs where pre-MSPE is “x” times larger than MSPE in South and West Yorkshire to deal with LAs that are unlikely to have suitable synthetic control. We will show this distribution with “x” set to different values, e.g. Abadie et al. (2010) use 2 and 5 times as large
- Compare both the levels and ratio of the post-MSPE to pre-MSPE in South and West Yorkshire with that in other LAs.

**ADDITIONAL ANALYSES**

No further analyses are planned beyond those already set out. However, given that synthetic control methods have only been used in a handful of cases to date, further ad hoc analysis may become necessary to check the robustness of the findings.
Report tables

Here, we provide a list of the proposed graphs and tables. Where a graph talks about outcomes, these will be presented for both maths and English in two separate panels.

Assessment of Match Quality

Table 1 – Comparison of time-constant control variables across raw groups (South and West Yorkshire, Rest of England, Selected Donor Pool), MSPE shown at bottom of table

Figure 1 – Evolution of primary outcomes across time for raw groups (South and West Yorkshire, Rest of England, Selected Donor Pool)

Table 2 – Comparison of time-constant control variables across treatment and synthetic control group, MSPE shown at bottom of table

Primary and Secondary ITT Analysis

Figure 2 – Evolution of primary outcomes over time across treatment and synthetic controls

Figure 3 – Difference in primary outcomes for South and West Yorkshire relative to synthetic controls

Figure 4 – Difference in secondary threshold outcomes for South and West Yorkshire relative to synthetic controls

Inference

Figure 5 – Evolution of difference in primary outcomes over time between placebo treatment LAs and placebo synthetic control group (likely to be over 100 very thin lines, with difference for South and West Yorkshire shown in bold)

Figure 6 – Repeat of Figure 5 narrowed down to placebo LAs where MSPE is less than “x” times greater than that in South and West Yorkshire before intervention

Figure 7 – Ratio of MSPE before and after intervention for South and West Yorkshire as compared with other placebo LAs where MSPE is less than “x” times greater than that in South and West Yorkshire before intervention. Graph to show position of South and West Yorkshire ratio in the overall distribution, giving an approximate p-value.

Comparison with Other Elements of the TA Campaign

Figure 8 – Evolution of difference in primary outcomes over time between schools in South and West Yorkshire with and without additional interventions.

Figure 9 – Evolution of difference in primary outcomes over time between schools in South and West Yorkshire and schools in the synthetic control group with and without incentives to engage with national campaign materials.

Appendix Tables

Table A1 – Comparison of time-constant control variables across South and West Yorkshire, preferred synthetic control, 3 alternative synthetic control groups and 2 matching specifications. MSPE shown at bottom of table.

Table A2 – Difference in primary outcomes between South and West Yorkshire and control group based on 2 matching specifications.
Figure A1 – Evolution of differences in primary outcomes between South and West Yorkshire, preferred synthetic control and 3 alternative synthetic control groups.

Figure A3 & A4 – repeat of Figures 6 and 7 for alternative “x”.

Figure A5 – Difference in primary outcomes over time between South and West Yorkshire, iteratively leaving out one local authority receiving positive weight.