ASCENTS 121 Support for Science

Evaluator: The National Centre for Social Research
Principal investigator: Arnaud Vaganay

Education Endowment Foundation

PROJECT TITLE	ASCENTS 121 Support for Science
DEVELOPER (INSTITUTION)	University of Lincoln
EVALUATOR (INSTITUTION)	NatCen Social Research
PRINCIPAL INVESTIGATOR(S)	Arnaud Vaganay
PROTOCOL AUTHOR(S)	Rukmen Sehmi, Daniel Phillips
TRIAL DESIGN	Multi-site, two-arm cluster randomised trial with random allocation at the pupil level
TRIAL TYPE	Efficacy
PUPIL AGE RANGE AND KEY STAGE	14 to 16 year-olds, Key Stage 4
NUMBER OF SCHOOLS	46
NUMBER OF PUPILS	768
PRIMARY OUTCOME MEASURE AND SOURCE	Grade achieved in Full GCSE Double Award Science (NPD derived)
SECONDARY OUTCOME MEASURE AND SOURCE	1. Grade achieved for A- level or AS- level Biology, Chemistry or Physics (NPD derived) to denote progression to A- or AS-level Science 2. Level achieved in Full GCSE English (NPD derived) 3. Level achieved in Full GCSE Maths (NPD derived)

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	12/03/2020	N/A

Table of contents

Introduction

This analysis plan sets out the intended multi-site efficacy trial of ASCENTS 121 Support for Science (hereafter 'ASCENTS'). ASCENTS is a mentoring programme intended to promote achievement in science amongst disadvantaged Year 11 pupils; all pupils in the trial were eligible for Free School Meals (FSM). Mentees receive one-to-one science support from trained STEM undergraduates (Science, Technology, Engineering and Mathematics) throughout the final year of their GCSEs.

Mentors are required to attend two days of training prior to delivering mentoring sessions. They then deliver 23 weekly one-hour face-to-face ASCENTS sessions throughout Year 11 to the pupils they are paired with. Mentors are required to cover topics that are part of the GCSE science curriculum. ASCENTS also includes a University Revision Day whereby pupils visit the university where mentors are studying.

The evaluation will be conducted as a multisite two-arm randomised controlled efficacy trial with randomisation at pupil level. The primary outcome of interest is GCSE science attainment using NPD-derived GCSE Double Award Science attainment levels. Secondary outcomes are GCSE English and Maths attainment, also assessed using NPD-derived GCSE attainment levels. A further secondary outcome is to assess whether being offered ASCENTS has encouraged pupils to pursue A-level science two years after pupils sit their GCSEs.

Design overview

Trial design, including number of arms		Multi-site, two-arm cluster randomised trial with random allocation at the pupil level
Unit of randomisation		Individual
Stratification variables (if applicable)		School
Primary outcome	variable	Science GCSE attainment
	measure (instrument, scale)	Level achieved in Full GCSE Double Award Science (NPD derived)
Secondary outcome(s)	variable(s)	1) Maths GCSE attainment 2) English GCSE attainment 3) Progression to Science A- or AS- level[1]
	measure(s) (instrument, scale)	1) Level achieved in Full GCSE Maths 2) Level achieved in Full GCSE English 3) Grade achieved for A- level or AS- level Biology, Chemistry or Physics[2]
Baseline for primary outcome	Variable	1) Science (teacher assessed), Maths or English attainment at KS2[3]
	Measure (instrument, scale, source)	1) Level achieved in KS2 Science, Maths or English
Baseline for secondary outcome(s)	Variable	1) Maths KS2 attainment 2) English KS2 attainment 3) Science (teacher assessed), Maths or English attainment at KS2[3]
	Measure (instrument, scale, source)	1) Level achieved in KS2 Maths 2) Level achieved in KS2 English 3) Level achieved in KS2 Science, Maths or English

Sample size calculations overview

The trial is designed as a multi-site trial with randomisation at the individual level and blocked randomisation by school. Pupils within each school are randomly assigned to either a treatment group receiving the offer of ASCENTS or a control group receiving 'business-as-usual' teaching and support.

Mentors were recruited from five partner universities, whereby each university had a pool of schools to which they supplied mentors to geographically proximate schools. The plan was to recruit 22 pupils from each of the 35 participating schools (770 pupils in total) and randomly assign pupils within schools to either receive mentoring or 'business as usual'. However, due

---

[1] NatCen will use an NPD outcome indicating grade achieved at A- or AS- level Science to create a binary indicator of whether pupils sat an exam or not; this will act as a proxy for progression to Science A- or AS- level. AS-levels are currently available on the curriculum but are 'decoupled' from A-levels so do not count towards overall A-level grades.

[2] We will use the grade achieved in the aforementioned subjects to indicate whether the student progressed to STEM-related A- or AS-level subjects

[3] Once we have the NPD data extract we will select the KS2 indicator that is the best predictor of Science GCSE (from KS2 Maths, Science, English), as the Science measure is categorical (KS2_SCITAOUTCOME: 1-6 national curriculum level achieved) and thus may be a poor predictor as it provides for little variation in assessed level (see Analysis section, page 6, for further explanation)

to variation in the number of pupils recruited by schools and mentoring capacity across the five partner universities, 845 pupils were recruited from 46 schools in total. To preserve the 1:1 allocation of mentors to students from the five participating universities, 77 pupils were randomly excluded from the evaluation resulting in a total of 768 pupils eligible for randomisation to treatment and control groups, of which 385 were randomly allocated to the intervention group and 383 to the control group (see Table 1). Randomisation was conducted by an independent analyst at NatCen Social Research using Stata v16 in September 2019[4].

Table 1: Minimum Detectable Effect Size Calculations

		PROTOCOL OVERALL	RANDOMISATION OVERALL
Minimum Detectable Effect Sizes (MDES)		0.16	0.16
Pre-test/ post-test correlations	level 1 (pupil)	0.59	0.59[5]
	level 2 (class)	n/a	n/a
	level 3 (school)	n/a	n/a
Intracluster correlations (ICCs)	level 2 (class)	n/a	n/a
	level 3 (school)	n/a	n/a
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		22	17[6]
Number of schools	intervention	n/a	n/a[7]
	control	n/a	n/a
	total	35	46
Number of pupils	intervention	385	385
	control	385	383
	total	770	768

N.B. No power calculation is provided for analysis of Free School Meal (FSM) pupils as the trial's recruitment criteria require that all pupils be eligible for FSM.

At the time of publishing the trial protocol, we used PowerUp![8] to estimate that this study is powered to detect an effect of 0.16 standard deviations based on the assumptions outlined in

---

[4] Stata Statistical Software: Release 16. College Station, TX: StataCorp LP.

[5] Torgerson, C. and Torgerson, D. (2013) `Randomised trials in education: An introductory handbook' London: EEF These authors reference a pre-test post-test correlation of 0.70. We use a more conservative correlation of 0.59 as our measure of baseline attainment is likely to be a poor predictor of the outcomes. See also footnote 3.

[6] Both the average cluster size for the protocol and randomisation are harmonised means.

[7] Randomisation is at the pupil-level so the number of schools is not applicable.

[8] Nianbo Dong and Rebecca Maynard, '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, no. 1 (1 January 2013): 24–67, doi:10.1080/19345747.2012.673143.

Table 1. The second column of Table 1 provides updated details on the calculation of our minimum detectable effects size at randomisation. Assuming the pre-test/post-test correlation of 0.59 at pupil level, we estimate the study to be powered to detect an effect of 0.16 standard deviations at the randomisation stage. Assuming that expected attrition at pupil level is 10% and 5% at school level, we estimate the study to be powered to detect an effect of 0.17 standard deviations at the analysis stage.

Analysis

The evaluation of ASCENTS aims to estimate the impact of the programme on Science, Maths and English attainment, and enrolment in further Science-related studies for Key Stage 4 pupils in England, using an intention-to-treat approach.

We propose to conduct the primary analysis using a single-level model with fixed effects for schools[9]. A fixed effect model is used in preference to a multilevel model as it is more appropriate for the purposes of drawing 'conditional inference', where we do not attempt to generalise beyond the sample of schools within the trial. This type of inference is more appropriate for efficacy trials, such as ASCENTS.

However, as randomisation is stratified at the school level, a single-level OLS approach to estimation without any account for clustering within schools is inadequate as the correlation among observations within clusters is not taken into account, leading to downward biased standard errors and narrow confidence intervals. We therefore implement a single-level OLS model at the pupil level inclusive of a set of dummy variables representing schools (as covariates), to account for non-independence between observations across clusters (Primo, Jacobsmeier, Milyo 2007)[10].

Analyses will be conducted in Stata v16. Statistical significance will be assessed using two-sided tests at the 5% level and estimates of effect with 95% confidence intervals (CIs) and p-values will be provided.

Primary outcome analysis

The primary outcome analysis will explore the following hypothesis:

Research Question 1: Being offered ASCENTS improves academic Science attainment, as measured by the NPD-derived GCSE Double Award Science attainment levels.

The main analysis will estimate the intervention's impact on science attainment, as measured by NPD-derived science GCSE score. Following EEF guidance (2019), evidence of effectiveness and reported effect sizes will be obtained from baseline-adjusted analysis, in which the dependent variable is the NPD-derived science GCSE score.

The science GCSE score can be treated as a continuous variable as scoring ranges from 1 to 9; hence, the effects are estimated through a single-level OLS model including the most appropriate measure of baseline Key Stage 2 attainment (either Maths, Science or English). Theoretically, Key Stage 2 Science is the best predictor of GCSE Science attainment but, as previously mentioned (see Footnote 3), the measure is categorical and thus may have poor predictive power. We will therefore consider whether continuous measures assessing Key Stage 2 Maths and Key Stage 2 English attainment are possible alternatives. Bearing in

---

[9] In line with EEF Analysis Guidance, 2019
[10] Primo, David M., Matthew L. Jacobsmeier, and Jeffrey Milyo. 2007. "Estimating the Impact of State Policies and Institutions with Mixed-Level Data." State Politics & Policy Quarterly7(4): 446-59.

mind the order of preference – Key Stage 2 Science, Maths and then English - we will review these possibilities by a) taking account of the extent of missingness for each measure and b) examining the variance explained (using R squared values) when each measure is predicted on the outcome, GCSE Science attainment.

Alongside the most suitable measure of baseline Key Stage 2 attainment, a dummy variable indicator capturing treatment/control group membership and a set of dummy variables representing blocked randomisation by school to adjust for clustering within schools, will also be included in the model.

The basic form of the fixed effect model is,

(1) $Science\ attainment_i = \beta_1\ baseline_i + \beta_2 intervention + school + e_i$[11]

Where pupils (*i*) are clustered in schools. The intervention effect is estimated by β_2, β_1 represents the baseline measure of prior attainment at Key Stage 2, y represents the set of dummies denoting school strata at randomisation, and *e* is the error term.

In line with EEF analysis guidance (2019), other covariates will not be considered at this stage. Sensitivity tests will be conducted using an adjusted model with additional covariates which are described fully in the section entitled 'Sensitivity analyses'.

Secondary outcome analysis

The secondary outcome analysis will explore the following research questions:

Research Question 2: What is the impact of ASCENTS on the Maths attainment of disadvantaged Year 11 pupils in England?

Research Question 3: What is the impact of ASCENTS on the English attainment of disadvantaged Year 11 pupils in England?

To test RQ2, the intended measure is an NPD variable that records the highest level achieved in full GCSE Maths. Analysis for RQ2 will follow the same method as the primary analysis, on an intention-to-treat basis, implementing a single-level OLS model including a baseline measure of KS2 attainment in Maths, a dummy variable indicator capturing treatment/control group membership and a set of dummy variables representing fixed effects at school level.

The basic form of the fixed effect model is,

(2) $Maths\ attainment_i = \beta_1\ baseline^{12}_i + \beta_2 intervention + school + e_i$[13]

To test RQ3, the intended measure is an NPD variable that records the highest point score achieved in full GCSE English. Analysis for RQ3 will follow the same method as the primary analysis, on an intention-to-treat basis, implementing a single-level OLS model including a baseline measure of KS2 attainment in English, a dummy variable indicator capturing treatment/control group membership and a set of dummy variables representing fixed effects at school level.

---

[11] [15] [18] NB the overall intercept is removed from the OLS model as the dummy variables eliminate all between-school variation leaving only within-school variation to be explained by covariates. i.e. the school-level covariates are perfectly collinear with the dummies and thus drop out of the equation.
[12] Baseline measure is maths attainment at KS2

The basic form of the fixed effect model is,

(3) $\text{English attainment}_i = \beta_1 \text{ baseline}[14]_i + \beta_2 \text{intervention} + \text{school} + e_i[15]$

Subgroup analyses

Subgroup analyses by FSM will not be performed as all pupils in the trial will be eligible for free school meals.

Additional analyses

As some pupils may share a mentor[16], an additional analysis will test a single-level OLS model using the same method as the primary analysis, but with the exclusion of the sub-sample of pupils who shared a mentor.

Longitudinal follow-up analyses

One longitudinal follow-up analysis will be conducted to explore the following hypothesis:

Research Question 4: Being offered ASCENTS increases progression to A- or AS-level Science (Biology, Chemistry or Physics).

To address RQ4, the intended measure will be a binary variable indicating whether pupils sat an A- or AS-level science exam in any of Biology, Chemistry, or Physics subjects. This measure acts as a proxy for whether pupils go on to study science at A-level and this information will be collected in 2022, two years after collection of GCSE attainment data.

Analysis for RQ4 will follow the same method as previous analyses, on an intention-to-treat basis, implementing a single-level OLS model including the same KS2 attainment measure as per the primary analysis[17], a dummy variable indicator capturing treatment/control group membership and a set of dummy variables representing fixed effects at school level to adjust for clustering within schools.

The fixed effect model will take the following form:

(4) $\text{Progression to A-level Science}_i = \beta_1 \text{ baseline}_i + \beta_2 \text{ intervention} + \text{school} + e_i$

In line with EEF guidance, effect sizes for estimates pertaining to the longitudinal analysis, whether pupils progress to science A level, will be presented as risk ratios as they are simpler to interpret than other commonly used options. Relative risk is the ratio of the probability of an event occurring in an exposed or treatment group versus the probability of the event occurring in the non-exposed or control group (Ferguson, 2009)[18]. In the context of ASCENTS, the 'relative risk' is interpreted as the ratio of the probability of progressing to a Science A-level amongst pupils in the treatment group as compared to the probability of progressing to a Science A-level amongst those in the control group.

We will calculate the relative risk ratio, following a method outlined by Fleiss & Berlin (2009), as the measure of effect size.

---

[14] Baseline measure is English attainment at KS2

[16] The number of pupils who have shared a mentor will be made available when registers are collected at the end of the trial in June.

[17] We will take the same steps as per the primary analysis to review whether KS2 Science, Maths or English is the most suitable baseline measure, taking account of the degree of missingness and variance explained (using R squared values) when predicted on the outcome, progression to Science A level.

[18] Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. Professional Psychology: Research and Practice, 40, 532–538.

$$(5) \qquad\qquad RR = \frac{1 - exp^{-\alpha}}{1 + exp^{-\alpha-\beta}}$$

In (5) above α refers to the constant from the logistic model (1) and β refers to the treatment indicator.

Imbalance at baseline

We will explore potential imbalance at baseline in pupil characteristics. Individual characteristics, such as gender or differences in baseline attainment, could impact outcomes of interest such as GCSE attainment in Science. We will examine variation in pupil characteristics for the 'as analysed' and 'as randomised' samples to explore potential imbalance resulting from randomisation or attrition.

At pupil level, the comparison will include the following factors:

- Gender
- Baseline Science attainment
- Baseline Maths attainment
- Baseline English attainment

Potential imbalance for categorical data will be reported as cross-tabulations, including a count and percentage for treatment and control group allocation, and tested with a Chi-Square test. Continuous variables will be summarised with descriptive statistics (n, mean, standard deviation, range, median and effect sizes) by group allocation and a paired sample t-test, and differences in test scores reported as Hedge's g effect sizes. An effect size of more than 0.05 indicates possible imbalance. Where imbalance is indicated, an additional model will be estimated as a sensitivity analysis including the unbalanced variables.

Missing data

Missing outcome data for science attainment (i.e. the outcome of interest for the primary analysis) and pupil-level covariates may occur due to factors such as pupils moving to other schools or absences at the time of baseline and follow-up data collections. If missing data exceeds five per cent[19] and provided observed covariates can predict missingness, the sensitivity of the estimated effect will be assessed using multiple imputation under the assumption that outcome data are missing at random.

First, we will use logistic regression analysis to examine predictors of 'missing on science outcome and covariates' modeled as a binary outcome measure; we will establish which pupil-level covariates outlined in the 'imbalance at baseline' section are predictive of missingness. We will also explore whether school-level factors are predictive of missingness, including Ofsted rating, setting type (i.e. independent, academy, etc.), rural or urban setting and proportion of pupils in settings eligible for FSM, and include in the imputation model if predictive.

If missingness can be predicted by observed pupil and/or school-level factors, we will use multiple-imputation to take account of any associated bias and present results alongside headline impact estimates for comparison. Multiple imputation by chained equations (MICE) will be implemented (using the **mi** suite of commands in Stata 16 SE) if there are both binary and categorical covariates in the imputation model. As a rule of thumb, the number of imputations implemented should be at least equal to the percentage of missing observations

---

[19] In line with EEF Analysis Guidance, 2019

(White, Royston and Wood, 2011)[20]. However, statistical power for small effect sizes may be diminished if an insufficient number of imputations is implemented, so we will ensure that the number of imputations will be considerably higher than the percentage missing on observations to reduce the chance of simulation error[21]. This may be particularly important if the proportion of missing information is high in our sample. Further, the first 100 iterations of the imputation will not be used (i.e. 'burn in') to ensure that convergence is stable. The **mi estimate** commands will then be implemented to run the imputed models and provide estimates.

The benefit of using a chained equations approach is that imputation occurs on a variable-by-variable basis, hence different variables can be modeled according to their own distribution (e.g. binary variables modeled using logistic regression, continuous variables modeled using OLS regression). As we are likely to include different types of variables, a chained equations approach may be preferable as it can be more robust than a mis-specified joint modelling approach[22]. Furthermore, a chained equations approach allows for passive imputation which may be helpful in improving specification of the model.

Compliance

Participation in ASCENTS is measured at the pupil-level, so a measure of compliance is created using pupil attendance at mentoring sessions (awarded one point for each of 23 sessions) and at the University Revision day (awarded three points). Scores for the number of mentoring sessions attended and the University Revision day will then be summed to produce a compliance scale with a possible range of $0 \leq Comply_A \leq 26$, which will then be re-scaled to a range of $0 \leq Comply_A \leq 1$. Trial pupils in the treatment group with higher scores within this range are deemed 'more' compliant, and those with lower scores are deemed 'less' compliant.

Some individuals may not conform to their assigned treatment arm, whether they be in the treatment or control group[23]. We will therefore use an Instrumental Variable (IV) approach where compliance is applied using allocation status (Angrist and Imbens, 1995[24]). The IV regression uses a two-stage least squares where the first equation estimates:

$$Comply_i = \propto + \beta_1 Treat_i + \varepsilon_i$$

The predicted values from the first stage equation, $\widehat{Comply_i}$, will then be used in the estimation of the second stage equation, as follows:

$$Y_i = \propto + \beta_1 Treat_i + \beta_2 \widehat{Comply_i} + \beta_3 Baseline_i + \omega_i$$

Stata 16 will be used to conduct the IV regression analyses using the command **ivregress**. Endogeneity tests will be used to assess whether the treatment allocation is suitable for the

[20] White, I. R., Royston, P. & Wood, A.M. (2011). Multiple Imputation Using Chained Equations: Issues and Guidance for Practice. Statistics in Medicine, 30, 377– 399.

[21] Graham, J.W., Olchowski, A.E., & Gilreath, T.D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science, 8, 206-213.

[22] Huque, M. H., Carlin, J.B., Simpson, J. A. & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. BMC Med Res Methodol, 18, 168.

[23] Participation data was collected by way of a register to document the attendance of trial pupils and also asked to collate a list of all pupils who, at any given point, received mentoring but were not in the treated group.

[24] Angrist, J. and Imbens, G. (1995) `Two-stage least squares estimation of average causal effects in models with variable treatment intensity.' American Statistical Association, 90(430), pp431-442

purposes of applying instrumental variable techniques (Wooldridge, 1995[25]), and F-statistics and p-values will be reported, in line with EEF guidance.

Intra-cluster correlations (ICCs)

Although the evaluation uses a one-level model, we will also run a two-level model for the purposes of informing future research. The unconstrained intra-cluster correlation will be calculated separately to the analysis model by running a multilevel model, including only 'treatment' as a covariate and a random effect for schools. The ICC: **ρ** will be estimated with the post-estimation command **estat icc** in Stata 16 SE, using the following formula:

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

Where σ_B^2 is the between-school variance, σ_W^2 is the within-school variance. Values of p range from 0 to 1, where values closer to 0 implies that the within-cluster variance is much greater than the between cluster variance.

Effect size calculation

In line with EEF guidance, estimates for the primary outcome, science attainment, as well as two secondary outcomes, Maths and English attainment will be reported as standardised effect sizes using Hedge's g with 95% confidence intervals.

The Hedge's *g* effect size will be estimated following Hedge's (2007)[26] formulae for the effect size d_t for designs with unequal sample sizes. The effect size, g_t is estimated as follows:

$$g_t = J \times \left(\frac{\overline{Y}_{\blacksquare\blacksquare}^T - \overline{Y}_{\blacksquare\blacksquare}^C}{S_T} \right)$$

Where $\overline{Y}_{\blacksquare\blacksquare}^T$ and $\overline{Y}_{\blacksquare\blacksquare}^C$ are the grand means of the treatment and control groups.

The remaining terms are calculated as follows:

The correction factor J is defined as:

$$J = 1 - \left(\frac{3}{4(N^T + N^C - 2) - 1} \right)$$

The pooled standard deviation, S_T is defined as:

$$S_T = \sqrt{\frac{\sum_{i=1}^{m^T} \sum_{j=1}^{n_i^T} (Y_{ij}^T - \overline{Y}_{\blacksquare\blacksquare}^T)^2 + \sum_{i=1}^{m^C} \sum_{j=1}^{n_i^C} (Y_{ij}^C - \overline{Y}_{\blacksquare\blacksquare}^C)^2}{N - 2}}$$

The variance term is calculated as follows:

---

[25] Wooldridge, J. M. 1995. `Score diagnostics for linear models estimated by two stage least squares'. In `Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao', ed. G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66-87. Oxford: Blackwell

[26] Hedges, L. V. (2007) 'Effect Sizes in Cluster-Randomized Designs' Journal of Educational and Behavioral Statistics 32(4): 341–370

$$V\{g_t\} = \left(\frac{N_T + N_C}{N_T N_C}\right) + \frac{\delta_g^2}{2(N-2)}$$