# Healthy Minds
# Statistical Analysis Plan

**Evaluator (institution): National Institute of Economic and Social Research**
**Principal investigator(s): Heather Rolfe**

Education Endowment Foundation

Template last updated: August 2019

| | |
|---|---|
| **PROJECT TITLE** | Healthy Minds: A randomised trial of the impact of the Healthy Minds curriculum on academic outcomes[1,2] |
| **DEVELOPER (INSTITUTION)** | Bounce Forward (previously How to Thrive) |
| **EVALUATOR (INSTITUTION)** | National Institute of Economic and Social Research |
| **PRINCIPAL INVESTIGATOR(S)** | Heather Rolfe |
| **SAP AUTHOR(S)** | Lucy Stokes, Francesca Foliano, David Wilkinson |
| **TRIAL DESIGN** | Two-arm cluster randomised controlled trial with randomisation at school level |
| **TRIAL TYPE** | Efficacy |
| **PUPIL AGE RANGE AND KEY STAGE** | 11-16[3] (KS3 and KS4) |
| **NUMBER OF SCHOOLS** | 34 |
| **NUMBER OF PUPILS** | 10,029 |
| **PRIMARY OUTCOME MEASURE AND SOURCE** | GCSE Attainment 8 (NPD) |
| **SECONDARY OUTCOME MEASURE AND SOURCE** | Maths and English attainment, end Year 7 and end Year 8 (Hodder Education Access Reading and Maths Tests)<br><br>Exclusions and absenteeism, end Year 7 and end Year 8 (NPD) |

## SAP version history

| VERSION | DATE | REASON FOR REVISION |
|---|---|---|
| 1.2 [*latest*] | | |

[1] The title given in the trial protocol refers to the original project title
2 A separate report presents the findings of the impact of the programme on health outcomes (Lordan and McGuire, 2019), available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Healthy_Minds.pdf
[3] Note the programme runs up to Year 10 (so this would be age 15) – we specify 16 here as the analysis focuses on GCSE outcomes

| | | |
|---|---|---|
| 1.1 | | |
| 1.0 [*original*] | | *N/A* |

# Contents



## Introduction

The Healthy Minds[4] programme aims to improve pupils' wellbeing and health-related outcomes. The intervention comprises a Personal, Social and Health Education (PSHE) curriculum for Year 7 to Year 10 pupils. This comprises a set of 14 modules, bringing together individual elements which have been identified to be effective through previous research. The programme uses the principles of cognitive behavioural therapy and theory from positive psychology to help students improve their resilience, understand the link between thoughts feelings and behaviour, so they understand themselves better and have more empathy for others. The resilience skills form the basis and thread through the curriculum into the full range of personal, social, health topics. In schools receiving the intervention, the programme either replaced schools' existing PSHE lessons (where this time was allocated), and/or was built into the school week at other times. The lessons were taught by school staff (teachers or learning support assistants), who received in-depth training in each element.

Lordan and McGuire (2019) have analysed the impact of the programme on health outcomes. The main focus of this evaluation (and in the analyses set out in this plan) is whether the programme has an impact on academic attainment. The programme aims at

[4] Also known as the Developing Healthy Minds in Teenagers programme.

improving health-related outcomes, and if the evaluation shows no impact on academic attainment, this indicates that the programme improves health outcomes without adversely affecting academic achievement, despite diverting time away from traditional lessons. Furthermore, the skills which the programme aims to foster could potentially have a positive effect on academic results.

The primary research question to be addressed in this evaluation is:

- Does the Healthy Minds programme have an impact on academic attainment, as measured by pupil attainment at end Key Stage 4?

In addition, this evaluation will explore:

- Does the Healthy Minds programme have a differential impact on academic attainment (at end Key Stage 4) for disadvantaged pupils (as defined by eligibility for Free School Meals)?
- Does the Healthy Minds programme have a differential impact on academic attainment (at end Key Stage 4) according to pupils' prior academic achievement (as defined by attainment at end Key Stage 2)?
- Does the Healthy Minds programme have an impact on academic attainment (in Maths and English) at end Year 7 and end Year 8?
- Does the Healthy Minds programme have an impact on the likelihood of exclusion and absenteeism by end Year 7 and by end Year 8?

## Design overview

| | | |
|---|---|---|
| **Trial design, including number of arms** | | Two-arm, cluster randomised |
| **Unit of randomisation** | | School |
| **Stratification variables** (if applicable) | | Proportion of pupils eligible for FSM; GCSE attainment; single or mixed sex school |
| **Primary outcome** | variable | GCSE attainment |
| | measure (instrument, scale, source) | Attainment 8 (range 0-90, NPD) |
| **Secondary outcome(s)** | variable(s) | Maths and English attainment<br><br>Exclusions<br><br>Absenteeism |
| | measure(s) (instrument, scale, source) | Maths and English attainment, end year 7 and end Year 8 (Hodder Education Access Reading and Maths Tests, age-standardised scores, range 65-130)<br><br>Exclusions, Year 7 and Year 8, total number of fixed exclusions for academic year (NPD) |

| | | |
|---|---|---|
| | | Absenteeism, Year 7 and Year 8, number of absence sessions (overall absence) during the academic year (NPD) |
| **Baseline for primary outcome** | **variable** | Attainment at end KS2 |
| | measure (instrument, scale, source) | KS2 average total points score (range 0-39, NPD) |
| **Baseline for secondary outcome** | **variable** | Maths: KS2 maths points score<br>English: KS1 reading score*<br><br>Exclusions, end KS2<br>Absenteeism, end KS2 |
| | measure (instrument, scale, source) | Maths: KS2 maths points score (range,0-51, NPD)<br>English: KS1 reading score (range 0-39, NPD)<br><br>Exclusions at end KS2, number of fixed exclusions for academic year (NPD)<br><br>Absenteeism at end KS2, number of absence sessions during the academic year (NPD) |

* Here we will use KS1 point scores, as a consistent KS2 reading point score is not available for all cohorts in our analysis (due to a change in the variables available within NPD for reading in 2012/13)

This is a cluster randomised trial, with school level randomisation. Randomisation was conducted using minimisation with schools stratified according to whether the percentage of pupils eligible for Free School Meals (FSM) is less than 13 per cent, between 13 and 25 per cent, or greater than 25 per cent; whether the percentage of pupils with 5 GCSEs at grades A*-C is below 59 per cent or not; and whether the school is single sex or mixed. School recruitment took place in two phases (discussed further below); the Phase 1 schools were randomised independently by the London School of Hygiene and Tropical Medicine, while Phase 2 schools were randomised by the LSE.

Recruiting schools in two phases was necessary as recruitment proved challenging and given the timing of the start of the study and the length of time required to engage and recruit schools. Thirteen Phase 1 schools were recruited to the project in the academic year 2012/13. Phase 1 schools assigned to the treatment group (N=7) implemented the intervention with their Year 7 pupils in the academic year 2013/14, while those schools assigned to the control group (N=6) did not implement the intervention in this year. Instead, the Phase 1 schools allocated to the control group implemented the intervention with their Year 7 pupils in the following academic year, 2014/15 (the waitlist control group). Thus, these Phase 1 control schools become part of the treatment group for the cohort of Year 7 pupils starting in 2014/15.

Twenty-one Phase 2 schools entered the project in academic year 2013/14. Phase 2 schools assigned to the treatment group (N=11) implemented the intervention with their Year 7 pupils in the academic year 2014/15, whilst Phase 2 schools assigned to the control group (N=10) do not implement the intervention.

In addition, for both phases, those pupils who were in Year 7 in the year preceding the implementation of the intervention also form part of the control group. For Phase 1 schools this is pupils who were in Year 7 in 2012/13, while for Phase 2 schools this is pupils who were in Year 7 in 2013/14. This is summarised in the table below.

**Summary of treatment and control groups**

| **Intervention group** | **N school cohorts** |
|---|---|
| | |
| Phase 1 schools allocated to intervention<br><br>Cohort 2013 (Pupils in Year 7 in 2013/14; in Year 11 in 2017/18) | 7 |
| Phase 1 schools (previously waitlist control group)*<br><br>Cohort 2014 (Pupils in Year 7 in 2014/15; in Year 11 in 2018/19) | 5 |
| Phase 2 schools allocated to intervention<br><br>Cohort 2014 (Pupils in Year 7 in 2014/15; in Year 11 in 2018/19) | 11 |
| | |
| **Control group** | |
| Phase 1 schools allocated to control<br><br>Cohort 2013 (Pupils in Year 7 in 2013/14; in Year 11 in 2017/18) | 6 (5) |
| Phase 1 schools (all)<br><br>Cohort 2012 (Pupils in Year 7 in 2012/13; in Year 11 in 2016/17) | 13 (12) |
| Phase 2 schools allocated to control<br><br>Cohort 2014 (Pupils in Year 7 in 2014/15; in Year 11 in 2018/19) | 10 (9) |
| Phase 2 schools (all)<br><br>Cohort 2013 (Pupils in Year 7 in 2013/14; in Year 11 in 2017/18) | 21 (20) |

*One of the schools in the waitlist control group ultimately never became part of the treatment group.

In total therefore, 34 schools were initially recruited to be part of the study, comprising 73 school cohorts, as shown in the table above. In practice, 2 of these schools dropped out (after randomisation); the resulting numbers of school cohorts are shown in parentheses in the table. Thus 32 schools, forming a total of 69 school cohorts, are included in the analysis.

## Sample size calculations overview

| | | Protocol | | Randomisation | |
|---|---|---|---|---|---|
| | | OVERALL | FSM | OVERALL | FSM |
| **Minimum Detectable Effect Size (MDES)** | | 0.25 | - | 0.27 | 0.33 |
| **Pre-test/ post-test correlations** | level 1 (pupil) | - | - | 0.5 | 0.5 |
| | level 2 (class) | - | - | - | - |
| | level 3 (school) | - | - | - | - |
| **Intracluster correlations (ICCs)** | level 2 (class) | - | - | - | - |
| | level 3 (school) | 0.13 | - | 0.13 | 0.13 |
| **Alpha** | | 0.05 | 0.05 | 0.05 | 0.05 |
| **Power** | | 0.8 | 0.8 | 0.8 | 0.8 |
| **One-sided or two-sided?** | | 2 | - | 2 | 2 |
| **Average cluster size** | | 80 | - | 145 | 22 |
| **Number of schools** | intervention | 16 | - | 18 | 18 |
| | control | 16 | - | 14 | 14 |
| | **total** | 32 | - | 32 | 32 |
| **Number of pupils** | intervention | 1280 | - | 3706 | 556 |
| | control | 1280 | - | 6323 | 948 |
| | **total** | 2560 | - | 10,029 | 1,504 |

At the point of preparing the protocol, the power calculations assumed 160 pupils per year group per school, based on analysis of year group size in the preceding year in the selected schools. The initial plan for the trial was to assess outcomes using standardised tests of Maths and English skills, such that 80 pupils per school would take the Reading test and 80 pupils would take the Maths test. The protocol power calculations were therefore based on these numbers. These also assumed an ICC of 0.13, significance level of 0.05 and power of 0.8. Based on these figures, the required number of schools to detect an effect size of 0.3 standard deviations is 23, whilst to detect an effect size of 0.25 standard deviations requires 32 schools (the scenario presented in the table above). Meta analysis of similar programmes (Durlak et al., 2011) indicates an average effect size of 0.27 standard deviations. Power calculations were not reported for the FSM subgroup in the trial protocol (and the trial was not designed to be powered to detect an effect on the FSM subgroup).

Ultimately 34 schools were initially recruited to the trial across Phase 1 and Phase 2. In addition a special school was also recruited (and randomised) but it was decided that this school would be excluded from the analysis of academic attainment (this is not specified in the protocol). As noted above, with dropout of 2 of the 34 schools (after randomisation), 32 schools are included in the final analysis. Following concerns around measurement burden and attrition resulting from testing burdens in the first two years of the trial, it was decided

that attainment data would no longer be collected directly from schools, and that instead the primary outcome would be GCSE attainment taken from the National Pupil Database (NPD). As data on GCSE outcomes should be available for (almost) all pupils per school, a greater number of pupils will be included in the primary analysis than originally anticipated.

The power calculations presented in the "at randomisation" column of the above table are therefore based on assuming GCSE outcomes data are available for all pupils in a year group. Assumptions around the inclusion of a pre-test are not stated in the protocol. For the power calculations at the point of randomisation presented above, we assume a pre-test-post-test correlation of 0.5. On this basis, the MDES (keeping all other assumptions the same), stands at 0.27. Based on an assumption of 15 per cent of pupils being eligible for FSM in the schools participating in the trial (this is consistent with figures presented in the interim report for the evaluation), the MDES for the FSM subgroup stands at 0.33. Throughout, the MDES calculations are based on the number of schools, rather than school cohorts, to allow for conservative estimation.

## Analysis[5]

The analysis will be carried out using multilevel regression models to reflect the clustered nature of randomisation, following the approach set out in the protocol and as adopted previously in the interim reports presenting analysis of attainment at end Year 7 and end Year 8. These analyses were for EEF internal use only but will be included as appendices in the report (see Interim Analyses).

The main analyses will be conducted on an intention-to-treat basis, with all treatment and control schools included, regardless of drop-out from the programme. However, we will also conduct a compliance analysis to explore the robustness of the main results to excluding those schools that have withdrawn from the programme, as discussed further below.

### *Primary outcome analysis*

The primary outcome for the trial is Attainment 8, which captures attainment at the end of Key Stage 4. Data on Attainment 8 will be obtained through linkage to data from the National Pupil Database (NPD).

As schools were recruited in two phases, and with multiple cohorts within schools, it is important to be aware of changes in the form of GCSE examinations across years. The analysis involves pupils who sat their GCSEs in summer 2017, 2018 and 2019 (the table shown earlier in the design overview section indicates the academic year in which pupils were in Year 11 for each phase and cohort). In summer 2017, the first reformed GCSEs were introduced for English language, English literature and maths, and graded on a new scale running from 9 to 1; in 2018, reformed GCSEs were introduced for a further 20 subjects, and by 2019, almost all new GCSEs were in place with points awarded on a 9 to 1 scale. An attainment 8 score is available from NPD for all three years, but it is important to be aware of these underlying changes in GCSE examinations. While the main analysis will combine pupils from all phases and cohorts, we will conduct an additional analysis to test the sensitivity of the results to including only pupils sitting GCSEs within the same academic year.

[5] Please see the Statistical Analysis Guidance.

Following the standard approach set out in the EEF guidance, the primary analysis model will control for prior attainment, treatment allocation and the variables used in randomisation (proportion FSM pupils, single or mixed sex school, prior school-level GCSE attainment – based on the definitions used for randomisation as described in the Design overview section on page 4). It will also include an indicator for the pupil cohort.

Prior attainment will be measured based on pupil's attainment at end Key Stage 2, using the average total points score, also obtained through linkage to the NPD.

The equation to be estimated is:

$y = X\beta + Z\mu + \epsilon$

where:

$y$ = vector of outcome scores

$X$ = covariate matrix (treatment group indicator, KS2 scores, variables used in randomisation, recruitment phase and cohort indicator)

$Z$ = design matrix identifying which school or cluster an individual attended

$\mu$ = vector of school random effects

$\beta$ = fixed effect parameters

$\varepsilon_{ij}$ = residual error term for j-th member of cluster (school) i

The analysis will be conducted in Stata (version 15).

### *Secondary outcome analyses*

Secondary outcomes will be analysed using essentially the same approach as for the primary outcome. As noted above, the secondary outcomes to be reported will be performance on the Hodder Education Access Reading Test and the Hodder Education Access Maths Test, as assessed at the end of Year 7 and at the end of Year 8. Here the measure of prior attainment will use subject-specific measures of KS2 attainment, rather than the KS2 average points score (although for reading this will use a measure of KS1 attainment in order to have a consistent measure across all cohorts). Additional secondary outcomes to be reported are exclusions and absenteeism at end Year 7 and end Year 8; for these models, rather than prior attainment, the model will include exclusions and absenteeism at end KS2.

### *Interim analyses*

Analyses based on these measures of Reading and Maths were documented in interim reports to the EEF. The models reported included a wider set of control variables[6] in addition to those specified for the primary analysis above. These analyses will also be presented within an appendix to the final evaluation report.

[6] They include additional controls compared to what would be standard for the primary analysis, namely: pupil age in months, gender, a dummy for eligibility for FSM. Additional school level variables controlled for are the proportion of pupils in the school with Special Educational Needs (SEN) and the proportion of pupils in the school with English as an Additional Language (EAL), as well as a dummy variable for those schools that took the tests later than expected

As the intervention comprises a four-year programme, it should be noted that assessment of outcomes at end Year 7 and end Year 8 may have been too soon to detect effects. It is also important to note that difficulties in testing resulted in significant missingness on the reading and maths outcome measures, and so these analyses should also be interpreted with caution for this reason.

### *Subgroup analyses*

As identified in the trial protocol, the analysis will also be conducted separately for the subgroup of pupils eligible for FSM (using the EVERFSM_6_P variable available through linkage to the NPD). We will run analyses interacting treatment allocation with FSM status, as well as running separate models for the FSM and non-FSM subgroups. The effect size for FSM pupils will be presented on the basis of the subgroup model.

We will also explore differences according to level of prior attainment, interacting treatment allocation with prior attainment. This will explore impacts for pupils with lower levels of prior attainment, based on those pupils with scores in the bottom third of KS2 scores, and for pupils with higher levels of prior attainment, based on those pupils with scores in the top third of KS2 scores.

In presenting the subgroup analyses it should be noted that the trial was designed to detect effects in the full sample rather than for subgroup analyses, and thus findings for these subgroups will be exploratory.

### *Additional analyses*

We will run an additional model for the primary outcome that includes a broader set of control variables. The exact specification will depend on the variables that it is possible to obtain from the NPD, but as far as possible we will replicate the models using the full specification that was used for the interim analyses - that is: additionally controlling for pupil age in months, gender, FSM eligibility, as well as the proportion of pupils in the school with Special Educational Needs (SEN) and the proportion of pupils in the school with English as an Additional Language (EAL). In particular, we will include variables to capture those aspects on which the sample appears unbalanced. Thus two analyses will be conducted; one which uses the broad set of available covariates, and one using the same specification as for the primary analysis but also including any additional covariates on which the sample appears unbalanced.

Some pupils will have moved school during the course of the project. Subject to the success of the matching process, it should be possible, through the NPD, to identify pupils who remain in the same school at the point of taking their GCSEs. As the primary analysis is on an intention-to-treat basis, all pupils will be included, but an additional analysis will test the robustness of the results to including only pupils who remain in the same school throughout.

Finally, due to the multiple cohorts (year groups) involved in the study, we will also conduct a sensitivity analysis for the primary outcome model to check whether the results are affected if analysis is confined to pupils within the same year group (acknowledging that such an analysis would inevitably be based on a smaller sample size). That is, we will run the same model used in the primary analysis, but separately for the sample of pupils that form the 2013 cohort, and then separately for those pupils that form part of the 2014 cohort.

### *Longitudinal follow-up analyses*[7]

A longitudinal follow-up analysis is beyond the scope of the current planned analysis of academic attainment and report but in future a longitudinal analysis could potentially be conducted to investigate whether any impact on KS5 outcomes is evident.

### *Imbalance at baseline*

We will report on the characteristics of the sample both at randomisation and at the point of analysis. This will include school characteristics (including Ofsted rating, school type (academy status), urban/rural locations, number of pupils, percentage FSM pupils, percentage EAL pupils, percentage SEN pupils) and pupil characteristics (age, gender, FSM eligibility, prior attainment (KS2 and KS1 attainment), absenteeism and exclusions. Reporting will follow the standard EEF template, with means and standard deviations reported for continuous variables and counts and percentages in each category given for categorical variables. In line with EEF guidance, differences in "pre-test" variables will be reported as effect sizes. In this trial no specific pre-test was administered so this will be captured through prior attainment as measured at the end of KS2.

### *Missing data*

We will report the number of complete cases (those without missing data). We will consider separately missingness in outcome data and missingness in covariates, and will report the distribution of missing observations by treatment arm.

In the event of greater than 5% missing data at either cluster or individual level, or a significant difference in missingness between treatment and control arms, we will conduct further investigation into the mechanisms of missingness This will be explored separately for missingness in outcomes and for missingness in covariates, as discussed below Any multiple imputation undertaken in response to missing data will focus on the primary analysis model, in line with EEF guidance. The extent to which multiple imputation is appropriate depends on the missingness mechanism.

Missing outcome data

For our primary outcome, as our measure is obtained from administrative data, we do not expect to see significant levels of missingness.

Nevertheless, if more than 5% of outcome data is missing we will investigate the extent to which baseline characteristics (at school and pupil level) are correlated with missingness, using logistic regression, where the dependent variable is a binary indicator for missingness. If this shows significant associations with any of the characteristics, we would conduct an additional analysis including those covariates in the model to assess the robustness of the main results.

Missing covariates

Again, where more than 5% of data is missing for any of the covariates in the primary analysis model we will undertake further investigation into the nature of missingness. This

[7] Please see the longitudinal analysis guidance.

will follow the same approach outlined above of a logistic regression to explore the extent to which observed baseline characteristics are correlated with missingness. Where this shows significant associations with the included characteristics, and where missingness does not depend on unobserved characteristics, the missing data are considered to be missing at random (MAR), and in this case, proceeding with multiple imputation (MI) is appropriate. This MI would use those variables shown to be significant in the logistic regression alongside an indicator of treatment arm, using ten imputations. We will then compare the estimated treatment effect based on the imputed data with that from the primary analysis.

If however, missing data depend on unobserved variables, the missing observations are missing not at random (MNAR), and this cannot be resolved through MI. In this case we would conduct further sensitivity analysis exploring the robustness of the results to differing definitions of the sample, through including and excluding variables for which observations are missing.

### *Compliance*

Not all schools allocated to the treatment group completed delivery of the Healthy Minds programme. The Bounce Forward team have a record of whether schools delivered the programme to the satisfaction of the Bounce Forward team (thus compliance is defined at school level). This can be used to form a binary indicator, taking a value of 1 if the school is considered to have delivered the programme to the satisfaction of the Bounce Forward team, and equal to zero otherwise (as in the compliance analysis conducted for the separate evaluation of health outcomes by Lordan and McGuire (2019)).

We will use this information on compliance in conducting a Complier Average Causal Effect (CACE) analysis to estimate intervention effects on treated pupils. We will estimate the CACE using two stage least squares (2SLS) regression by estimating a (first stage) model of compliance, using the binary measure of compliance described above. The predicted values from the first stage are then used in the estimation of a model of our outcome measure. This analysis will be conducted using the ivregress command in Stata, which allows the necessary adjustments to be made to standard errors in response to the clustered nature of the data.

### *Intra-cluster correlations (ICCs)*

We will estimate the ICCs for the pre-test and post-test at school-level using empty hierarchical linear models including school-level random effects as follows:

$$\boldsymbol{Y_{ij} = \beta_0 + \eta_j + \varepsilon_{ij}}$$

where $\boldsymbol{Y_{it}}$ is the pre- or post-test of individual $\boldsymbol{i}$ in school $\boldsymbol{j}$, $\boldsymbol{\beta_0}$ is a constant term, $\boldsymbol{\eta_j}$ is a school-level random effect and $\boldsymbol{\varepsilon_{ij}}$ is an individual-level idiosyncratic error term. The ICC estimate is recovered as follows:

$$\boldsymbol{ICC = \frac{var(\eta_j)}{var(\eta_j) + var(\varepsilon_{ij})}}$$

### *Effect size calculation*

Effect sizes will be calculated using Hedges' g, following the standard approach for EEF trials as set out in the EEF analysis guidance. This will therefore be calculated as:

$$ES = \frac{(Y_T - Y_{C)adjusted}}{s*}$$

where s* is the pooled unconditional variance of the treatment and control groups. All relevant parameters will be provided in the report so that readers are able to compute alternative definitions of effect sizes.

Ninety-five per cent confidence intervals will be reported around the resulting effect.

## References


Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82(1), 405-432.

Education Endowment Foundation (2018) Statistical analysis guidance for EEF evaluations. Available at: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Lordan, G. and McGuire, A. (2019). Healthy Minds: Health outcomes, Evaluation report and executive summary, Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Healthy_Minds.pdf