Randomised Controlled Trial: Evaluation of the White Rose Maths Reception Jigsaw Statistical Analysis Plan – longitudinal follow-up

Evaluator: NFER
Principal investigator: Ben Styles

PROJECT TITLE	Randomised Controlled Trial: Evaluation of the White Rose Maths Reception Jigsaw
DEVELOPER (INSTITUTION)	White Rose Maths (Trinity MAT)
EVALUATOR (INSTITUTION)	National Foundation for Educational Research (NFER)
PRINCIPAL INVESTIGATORS	Ben Styles
PROTOCOL AUTHORS	Helen Poet, Ben Styles, Guido Miani, Joana Andrade, Anusha Ghosh
SAP AUTHORS	Gemma Schwendel
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at school level
TRIAL TYPE	Efficacy
PUPIL AGE RANGE AND KEY STAGE	Year 1 Pupils (age 5-6)
NUMBER OF SCHOOLS	138
NUMBER OF PUPILS	2689 (approximately 19 pupils per school)
PRIMARY OUTCOME MEASURE AND SOURCE	Maths attainment (New Progress in Understanding Mathematics – New PUMA)

SAP version history

VERSION	DATE	REASON FOR REVISION

Table of contents

Introduction

The Reception Jigsaw is a continuing professional development programme developed and run by White Rose Maths (WRM). The programme involves five cycles of a twilight CPD session followed by a half day coaching session with gap tasks completed in between. Reception Jigsaw is aimed at reception teachers but Year 1 teachers and school maths leads are expected to attend the five CPD sessions. Year 2 teachers and teaching assistants can also attend. For further detail on the programme see Poet *et al.*, 2021)

EEF commissioned WRM and NFER to run an evaluation of the Reception Jigsaw (Poet *et al.*, 2021). WRM delivered the Reception Jigsaw in schools allocated to the intervention group from November 2021 to May 2022. The report is due to be published in 2023. A longitudinal follow-up analysis was planned as part of the evaluation to answer the first of the secondary research questions:

> RQ2: What is the impact of the White Rose Maths Reception Jigsaw on pupils' maths attainment as measured by New PUMA tests at the end of Year 1, compared to 'business as usual'? (Poet *et al*., 2021).

As specified in the trial protocol, in order to answer this question the New PUMA will be re-administered to the same pupils who took part in the first stage of the trial at the end of Year 1 (in summer 2023). (The first stage of the trial assessed these pupils at the end of their reception year). The timing of the follow-up will have the added benefit of including the additional influence of Year 1 teachers who were also trained alongside reception teachers. The follow-up is measuring the impact of the Reception Jigsaw on the medium-term pupil attainment outcome as shown in the logic model (Poet *et al.*, 2021).

The decision as to whether to continue with the follow-up was taken by EEF following EEF guidance for longitudinal analysis (EEF, 2019). The findings from the longitudinal follow-up will be published as an addendum to the main report. This statistical analysis plan (SAP) details the exploratory analyses planned on the follow-up data collected in summer 2023.

Design overview

Upon signing the Memorandum of Understanding (MoU) during the recruitment stage of the trial, schools were made aware of the possibility of a longitudinal follow-up. The follow-up was confirmed to all schools that were randomised as part of the first stage of the trial (n=138) in an email sent in December 2022. In March 2023 all schools were provided with their list of around 20[1] pupils who had been sampled for the trial prior to randomisation and were asked to update any details such as Year 1 class name and whether any sampled pupils had left the school. Further detail of the randomisation and pupil sampling approach are available in the main SAP for the trial (Andrade, Styles & Morton, 2022).

As part of the pupil profoma, schools were asked whether they had participated in WRM's Primary Jigsaw during last academic year (2021-22) or this (2022-23). In addition we asked schools allocated to the control group whether they had taken part in WRM's Reception Jigsaw this academic year (2022-23). Schools allocated to the control group were required not to take part in the Reception Jigsaw in the main trial year (2021-22) but this requirement did not extend into 2022-23. It is important to gather this data as Year 1 teachers are expected to attend the twilight CPD sessions as part of Reception Jigsaw so control group schools accessing Reception Jigsaw in this current academic year (2022-23) may lead to contamination of the control group. Similarly, while schools signing up to take part in the trial had not participated in more than two sessions of Primary Jigsaw prior to the main trial, they may have signed up since. The data received from schools regarding participation in Jigsaw training was cross-checked with WRM. We also asked schools to provide some options for possible dates where NFER test administrators could attend and administer the New PUMA Year 1 assessments.

[1] Although 20 pupils per school were originally sampled (where numbers allowed), there is an mean average of 19 per school due to some pupils moving from the school over the academic year 2021-22.

Table 1 shows the timings of planned activities going forward relating to the follow-up. Tests were paper-and-pen tests and test administrators will be blinded to randomisation group prior to visiting the school (although group allocation may be revealed by school staff during the visit). Training will be provided for the test administrators via a webinar scheduled for May 2023.

Timeline

Table 1: Timeline of activities related to longitudinal follow-up

Date	Activity	Who
December 2022	Email informing schools that the follow-up was taking place	
March 2023	Schools provide possible assessment dates and check pupil/teacher data	NFER, schools
April - May 2023	Assessment days booked in with test administrators and schools	NFER test administrators, schools
	Extension for NPD access application	NFER
May 2023	Webinar for test administrators	NFER test administrators, NFER
June - July 2023	Testing in schools, scripts returned to NFER	NFER test administrators, schools
July - August 2023	Marking, data input and data cleaning	NFER
August 2023	Data sent to ONS for linking to the NPD	NFER
August – October 2023	Analysis (main analysis not using NPD)	NFER
August – September 2023	Send schools pupil results	NFER
September 2023	Incentives paid to schools	NFER
September - October 2023	Access to NPD granted Analysis (subgroup analysis using FSM from NPD)	ONS, NFER
October 2023	Report submission	NFER

Baseline and outcome measures

Baseline measure

The baseline measure for the follow-up analysis was collected as part of the first stage of the evaluation and was derived from a bespoke 20-item checklist that was completed by reception teachers about their pupils prior to randomisation. The checklist, in Appendix A, was created by the evaluation team (with input from early years specialists at WRM) and was based on the Early Years Outcomes and the Early Learning Goals. All the items in the checklist have as possible responses: 'cannot do with support', 'can do with support', and 'can do independently' that were recoded as numeric values (1, 2, 3). The recoded numeric variables were used to derive the emerging numeracy (EN) scores.

Outcome measure

The outcome measure for this follow-up study is the Year 1 New Progress in Understanding Mathematics Assessment (New PUMA), a standardized test developed by Rising Stars (part of Hodder Education). The New PUMA is the new edition of the 2014 PUMA released in 2020. It is preferable in terms of cost and minimising burden on schools to use statutory assessments as outcome measure for longitudinal follow-up (EEF, 2019) however this is not possible here as the next formal assessment for this cohort is at end of Year 6 (Key Stage 2).

Analysis

The analyses will follow EEF's 2018 guidelines (EEF, 2018)[2] and assume intention to treat (ITT).

Primary analysis

These results will be analysed using a multilevel model similar to the one used in the primary analysis except for the dependent variable, which in this case will be maths attainment as measured by the version of New PUMA adopted for Year 1, rather than reception.

A multilevel random intercepts model with two levels (school and pupil) will be used to account for cluster randomisation. The primary analysis will investigate whether reception teachers having access to Recpetion Jigsaw training and support had an effect on their pupils' maths attainment by the end of Year 1. This will be determined by fitting a model with maths attainment at the end of Year 1, as measured by New Progress in Understanding Mathematics (PUMA) scores, as the dependent variable.

To control for prior ability, pupil-level EN scores assessed at baseline will be included in the model as a covariate. The model will also contain a dummy variable for stratum (school region versus school participation in the Mastering Number programme) to reflect the stratified randomisation (see main SAP: Andrade, Styles & Morton, 2022).

The two-level random intercepts model is given by:

$$\mathrm{PUMA}_{ij} = \beta_0 + u_{0j} + \beta_1 \mathrm{intervention}_j + \beta_2 \mathrm{baseline\ EN}_{ij} + \beta_3 \mathrm{stratum}_j + \epsilon_{ij}$$

Where PUMA_{ij} is the New PUMA score of pupil i in school j, u_{0j} is the random intercept in school j, $\mathrm{intervention}_j$ is the school-level intervention/control dummy variable, $\mathrm{baseline\ EN}_{ij}$ is the baseline EN score of pupil i in school j, and $\mathrm{stratum}_j$ is a dummy variable for the randomisation stratum of school j.

The model will be run in R (version 4.1.2) using the package 'nlme'[3].

Secondary analysis

In order to account for any schools who may have taken part in Jigsaw training programmes since randomisation (either control schools who have taken part in either the Reception Jigsaw or Primary Jigsaw after the first stage of the trial was complete, or intervention schools who have taken part in the Primary Jigsaw at any stage during the evaluation), the primary outcome analysis will be repeated, with two additional dummy covariates.

The two-level random intercepts model is given by:

$$\mathrm{PUMA}_{ij} = \beta_0 + u_{0j} + \beta_1 \mathrm{intervention}_j + \beta_2 \mathrm{baseline\ EN}_{ij} + \beta_3 \mathrm{stratum}_j + \beta_4 \mathrm{jigsaw\_r}_j + \beta_5 \mathrm{jigsaw\_p}_j + \epsilon_{ij}$$

Where PUMA_{ij} is the New PUMA score of pupil i in school j, u_{0j} is the random intercept in school j, $\mathrm{intervention}_j$ is the school-level intervention/control dummy variable, $\mathrm{baseline\ EN}_{ij}$ is the baseline EN

[2] Analysis will follow EEF's 2018 statistical analysis guidance rather than the updated 2022 version as the analysis will mirror that undertaken in the main trial which was completed under the previous 2018 guidance.
[3] Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2021). _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-153, <URL: https://CRAN.R-project.org/package=nlme>.

score of pupil i in school j, stratum_j is a dummy variable for the randomisation stratum of school j, jigsaw_r_j is a dummy variable for whether a control school j has taken part in Reception Jigsaw since the first stage of the evaluation has been completed (intervention schools will all be coded 0) and jigsaw_p_j is a dummy variable for whether school j has taken part in Primary Jigsaw at any stage of the evaluation.

The model will be run in R (version 4.1.2) using the package 'nlme'[4].

Subgroup analyses

As per the first stage of this evaluation, a subgroup analysis will be run to investigate possible differential effects of Reception Jigsaw on the attainment of children by the end of Year 1 eligible for free school meals. EVERFSM6 collected from the National Pupil Database (NPD) at the time of the first stage analysis (Autumn 2022) will be used as the identifier for the main subgroup analysis. While it is likely that there will be larger numbers of pupils in our sample flagged in the EVERFSM6 group in more recent updates due to updated school census data as compared to those pupils in this group in the NPD in early autumn 2022, retaining the subgroups from the first stage of the analysis is the simplest approach (EEF, 2019).

As discussed in the Randomisation section in the main SAP (Andrade, Styles & Morton, 2021), Mastering Numeracy, a programme with similar characteristics to the Reception Jigsaw was implemented in parallel to the trial in 47 of the 138 participating schools. The subgroup analysis on the schools that took part in the Mastering Numeracy programme will also be repeated in this follow-up study.

The analyses will be approached in two distinct ways: by running models with interaction terms (i.e. models that include both the subgroup indicator and the product of the subgroup indicator and randomisation group), and by running separate primary outcome models on:

1) Just the FSM eligible pupils
2) Schools that signed up for the Mastering Number and on schools that didn't sign up for the alternative programme.

Both approaches conform to the EEF 2018 guidelines.

The multilevel level random intercepts model with interaction terms for the FSM subgroup analysis will be given by:

$$\text{PUMA}_{ij} = \beta_0 + u_{0j} + \beta_1 \text{intervention}_j + \beta_2 \text{baseline EN}_{ij} + \beta_3 \text{FSM}_{ij} +$$

$$+\beta_4 \text{FSM}_{ij} * \text{intervention}_j + \beta_5 \text{stratum}_j + \epsilon_{ij}$$

With FSM_{ij} being a dummy variable for pupil i in school j's FSM eligibility status and the remaining variables as described in the Primary analysis section above.

And the two multilevel level random intercepts models with interaction terms for the participation/non-participation in Mastering Number subgroup analysis will be given by:

$$\text{PUMA}_{ij} = \beta_0 + u_{0j} + \beta_1 \text{intervention}_j + \beta_2 \text{baseline EN}_{ij} + \beta_3 \text{MN}_j +$$

$$+\beta_4 \text{MN}_j * \text{intervention}_j + \beta_5 \text{stratum}_j + \epsilon_{ij}$$

Where MN_j is the indicator of whether school j has signed up for the Mastering Number programme or not, and the remaining variables as described above in the Primary analysis subsection.

[4] Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2021). _nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-153, <URL: https://CRAN.R-project.org/package=nlme>.

Power analyses will also be performed to determine if subgroup analyses are underpowered.

Missing data

Although a low level of pupil attrition is to be expected in the context of the Reception Jigsaw trial we will nevertheless tackle the problem of missing data if the proportion of participants with missing data exceeds five per cent.

After evaluating to what extent data are missing and counting the number of complete cases, we will proceed to identify patterns of missingness in terms of the primary outcome variable. By design only pupils who were assessed at baseline were included in the trial, and so we are not expecting to find missing cases in the data corresponding to any of the covariates of the primary analysis model (baseline emerging numeracy and randomisation group and block variables whose values are already known). As such, we will not investigate missingness in terms of any variables other than primary outcome.

We will start by investigating patterns of missing data by means of a two-level (pupil and school) logistic model where the outcome is missingness, with baseline EN, school region (Essex or Yorkshire), whether the school signed up to Mastering Number, and randomisation group indicators as covariates. Additional variables that may be associated with missingness, but which are not included in the primary analysis, will also be included as covariates.

After this stage the analyses will follow the road-map from EEF 2018 analysis guidance[5].

If necessary, sensitivity analysis built on a multilevel multiple imputation will be implemented. The missing primary outcome values will be imputed using predictive mean matching, with five plausible values derived for each case. The primary analysis model will then be re-run on the five sets of imputed plausible values and the estimates for each model will be pooled into a single set of estimates and standard errors that will be compared to the results of the original analysis.

The missing data analysis will be run in R (version 4.1.2) using the packages 'mice'[6] and 'smcfcs'[7] (pooling of the results of the plausible values models).

Compliance

The compliance analysis that was carried out in the main study will be repeated for this follow-up study, with the two compliance measures used remaining the same:

1. A pupil was taught by a reception teacher who completed the intervention if the teacher participated in nine or more sessions out of the ten available (five Twilight in-depth training sessions and five half-day coaching sessions); this binary measure will form an optimal compliance indicator in the CACE analysis.
2. The second CACE analysis will use a pseudo-continuous dosage measure of compliance (0-10 sessions completed by the teacher).

There is a small group of children in the trial's sample that were taught by more than one teacher at reception. We will not define compliance measures for these children or include their data on the CACE analyses.

To evaluate if there is an association between teachers having completed the intervention and maths attainment of their pupils we will adopt the instrumental variables methodology (IV) prescribed by the EEF 2018 guidelines: we will run an instrumental variable regression by two-stage least squares model

[5] We are working under the expectation that there will be will no missing values among the models' covariate under MAR (missing at random), and that it will be possible to obtain valid estimates by including covariates predictive of non-response in the substantive models. The models' interpretation is conditional on these covariates being included.
[6] Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL https://www.jstatsoft.org/v45/i03/.
[7] Jonathan Bartlett, Ruth Keogh and Edouard F. Bonneville (2021). smcfcs: Multiple Imputation of Covariates by Substantive Model Compatible Fully Conditional Specification. R package version 1.6.0. https://CRAN.R-project.org/package=smcfcs

with group allocation as the instrumental variable. The same approach will also be used to investigate the presence of an association between dosage (number of training sessions attended by a teacher) and pupil maths attainment. Both models will be fit using the function ivreg from the R package 'ivreg'[8] and the estimation of causal effects will be done resorting to the functions contained on the 'ivpack'[9] package.

The analyses will, as before, be run in R (version 4.1.2).

Intra-cluster correlations (ICCs)

ICCs will be estimated from the variance of the random intercept and residual variance of multi-level models by means of the formula:

$$ICC = \frac{\sigma^2_{intercepts}}{\sigma^2_{intercepts} + \sigma^2_{residuals}}$$

Pre-test ICCs will be computed considering random intercepts two-level (school and pupil) models with no covariates, and post-test ICCs will be derived from the primary ITT and secondary ITT models described above**.**

Effect size calculation

As specified in the EEF 2018 guidelines, the results of the analyses of continuous outcomes by means of multi-level regression models will be reported as Hedges' g. The effect size will be calculated according to the formula

$$\boldsymbol{g = \frac{\bar{o}i - \bar{o}c}{s^*}}$$

The numerator for the effect size calculations can be calculated as the coefficients of the intervention group from the regression models, and the denominator as the unconditional total variance from the corresponding models without covariates. The effect sizes thus computed are equivalent to Hedges' g.

Confidence intervals for each effect size will be computed by multiplying the standard errors of the intervention group coefficient by the 2.5th percentile of a Student's t-distribution with the number of degrees of freedom associated with the sample size. The confidence intervals for the coefficient will be converted to effect size confidence intervals using the same formula as the effect sizes themselves.

[8] John Fox, Christian Kleiber and Achim Zeileis (2021). ivreg: Instrumental-Variables Regression by '2SLS', '2SM', or '2SMM', with Diagnostics. R package version 0.6-1. https://CRAN.R-project.org/package=ivreg
[9] Yang Jiang and Dylan Small (2014). ivpack: Instrumental Variable Estimation.. R package version 1.2. https://CRAN.R-project.org/package=ivpack

References

Andrade, J., Styles, B., and Morton., C., 2022. *Randomised Controlled Trial: Evaluation of the White Rose Maths Reception Jigsaw: Statistical Analysis Plan*. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reception-jigsaw [accessed 16 January 2023].

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), pp. 1-48. https://doi.org/10.18637/jss.v067.i01.

EEF, 2018. *Statistical analysis guidance for EEF evaluations, March 2018*. London: EEF. Available at: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf [accessed 20 April 2023].

EEF, 2019. *Longitudinal analysis of EEF trials*. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design [accessed 20 March 2023].

Kish, L., 1965. *Survey Sampling*. New York: Wiley.

Poet, H., Styles, B., Miani, G., Andrade, J., and Ghosh, A., 2021. *Randomised Controlled Trial Evaluation of the White Rose Maths Reception Jigsaw: Evaluation Protocol*. London: EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reception-jigsaw [accessed 16 January 2023].

Robinson-Smith, L., Fairhurst, C., Stone, G., Bell, K., Elliott, L., Gascoine, L., Hallett, S., Hewitt, C., Hugill, J., Torgerson, C., Torgerson, D., Menzies, V. and Ainsworth, H., 2018). *Maths Champions: Additional Appendices*. London: EEF [online]. Available at: https://educationendowmentfoundation.org.uk/public/files/Maths_champions-appends.pdf [Accessed 27 July 2021].

Appendix A

Observational Check list for the Primary and Secondary I analyses baseline measure

Please indicate for [insert pupil's name] whether, based on what you have observed since they started Reception, they can do each of the following tasks with support or independently.	cannot do with support	can do with support	can do independently
1. Recognises numerals 1 to 5			
2. Counts up to 3 or 4 objects by saying one number for each item in order			
3. Subitises (recognises quantities without counting) up to 5			
4. Counts objects up to 10			
5. Count objects beyond 10			
6. Counts out a smaller number of objects (up to 6) from a larger group - e.g. "give me four of the…"			
7. Selects the correct numeral to represent 1 to 5 objects			
8. Selects the correct numeral to represent 6 to 10 objects			
9. Estimates how many objects they can see and checks by counting them (up to 10)			
10. Uses the language of 'more than' and 'less than/fewer' to compare two sets of objects			
11. Finds the total number of items in two groups by counting all of them (up to 10)			
12. Finds one more or one less from a group of up to 5 objects			
13. Finds one more or one less from a group of up to 10 objects			
14. Records mathematical ideas using marks or objects they can interpret and explain			
15. Selects a particular named shape			
16. Can describe the relative position of objects or shapes such as 'behind' or 'next to'			
17. Uses familiar objects and common shapes to create and build models			
18. Continues, copies and creates repeating patterns			
19. Uses everyday language related to time			

20. Orders and sequences familiar events			