Statistical Analysis Plan for ICCAMS Maths Evaluation

The University of Manchester

INTERVENTION	
DEVELOPER	University of Nottingham and Durham University
EVALUATOR	University of Manchester
TRIAL REGISTRATION NUMBER	ISRCTN12649501
TRIAL STATISTICIAN	Maria Pampaka and Graeme Hutcheson
TRIAL CHIEF INVESTIGATOR	Maria Pampaka
SAP AUTHOR	Maria Pampaka, Julian Williams, and Lawrence Wo
SAP VERSION	1.1
SAP VERSION DATE	14 May 2018
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Table of Contents

Introduction

The ICCAMS Maths intervention was developed in a recently completed 4.5 year ESRC-funded project[1], which focused on improving teaching and learning in Key Stage 3 mathematics. ICCAMS Maths is designed to teach two mathematical areas that are a key part of the Key Stage 3 curriculum, but which cause particular problems to students: algebra and multiplicative reasoning. The teaching development programme is comprised of 40 lesson plans, 20 associated assessment pre-tasks, and an extensive teacher professional development (PD) programme.

The current study aims to compare the effect of the ICCAMS Maths intervention (when delivered 'at distance' through the institutional network of Maths Hubs) to a 'business-as-usual' control group in a cluster randomised controlled trial, and with a particular focus on addressing the mathematical learning needs of low attaining students in deprived socio-economic contexts.

Study design

Description of population including eligibility criteria

Eligible schools were mainstream English state secondary schools (or middle schools) with more than two class intake for Year 7 (ideally not in special measures) and with, ideally, higher than average levels of FSM eligibility. Schools were only eligible to take part in the study if they agree to all of the study requirements outlined in the Participation Agreement between the Universities and Schools and the form was signed by the head teacher. The trial schools were recruited (by Durham, supported by NCETM and the Maths Hubs, and Nottingham) in five regionally-based groups to facilitate the hub-based PD. Recruitment also aimed to minimise the number of schools also taking part in the Schools, Students and Teachers Network SSAT trial, "Whole school Embedding Formative Assessment Project" or any other special program deemed related.

All students in Year 7 at the beginning of the 2016/17 school year are the target cohort (excluding those without parental consent). All teachers teaching Year 7 at the beginning of 2016/17 are also a group of interest in this evaluation.

Description of trial design and trial arms

The main trial is designed as a cluster randomised controlled trial with randomisation at school level, taking place in July 2016. The trial then runs for two academic years with the intervention schools running the intervention for all students in Year 7 initially and for the same students again when they are in Year 8 (September 2016 to July 2018).

Sample size

109 schools were recruited and then randomised (see next section). The resulted sample size based on the recruited schools size is estimated around 21,000 students.

Number and timing of measurement points

The final attainment and attitudinal outcomes will be collected in June 2018 (after the end of the intervention) along with students' and their teachers' perceptions of teaching practices and confidence with ICCAMS (only intervention teachers at the end). Students' mathematics dispositions and perceptions of teaching practices as well as teachers' perceptions are also

[1] The ICCAMS Maths intervention was developed in the ICCAMS 1 project (Increasing Competence and Confidence in Algebra and Multiplicative Structures), which was funded by the Economic and Social Research Council (ESRC), grant reference RES-179-25-009 (2008-2012). For more information, see: http://iccams-maths.org/

measured at baseline (September – November 2016). Baseline data (i.e. KS2 maths scores) will be collected from NPD.

Protocol changes

There have been no changes to the protocol at the time of writing the SAP.

Randomisation

Random allocation was at the school level and took place in July 2016 (with baseline measurement planned for September-October 2016), after the receipt of the school file with relevant school-level information (%FSM and %GCSEA*toC). The initial plan was to perform randomisation within each regional hub with expected maximum of 30 schools to randomise, and thus using block stratified randomisation (Torgerson & Torgerson, 2008). In order to ensure balance in regards to previous attainment and proportion of FSM, blocks were expected to be defined by the proportion of students in each school to achieve 5 A*-C in the 2015 GCSE examinations (above median and below median) and the proportion of students in each school to be eligible for Free School means (above and below median). This implied that there were up to 4 blocking variables (or strata) made up of the combinations of these two variables. Preliminary investigation of the given school information based on the medians of the two strata (FSM and GCSE %) within each area revealed some problems especially with confounding of the two variables in some areas. In order to account for this confounding, deal with the missing information (i.e. not available) for some schools, as well as ensure balance in the overall design and school split it was considered more useful to define the groups/blocks based on 3 categories per strata and with the steps detailed in the evaluation protocol.

As a result of this process, there were 55 schools assigned to the Experimental group and 54 to the Control group in total (see Table 1). The distributions of FSM and GCSE are as shown in Table 2.

Table 1: Allocation of schools by Area

	Area					
Arm	**1**	**2**	**3**	**4**	**5**	**Total**
Control	9	10	9	13	13	54
Intervention	10	10	10	12	13	55
Total	**19**	**20**	**19**	**25**	**26**	**109**

Table 2: Average percentage of FSM and GCSE of allocated schools by arm and area

Average of FSM	**Area**					
	1	**2**	**3**	**4**	**5**	**Total**
Control	26.52	39.15	24.02	30.21	21.62	28.15
Intervention	31.51	38.47	21.39	29.65	23.77	28.82
Total	**29.15**	**38.81**	**22.71**	**29.95**	**22.7**	**28.48**
Average of GCSE	**Area**					
	1	**2**	**3**	**4**	**5**	**Total**
Control	51.57	65.11	55.11	47.5	63	56.4
Intervention	52.8	63.8	56.22	52.9	60.15	57.4
Total	**52.29**	**64.42**	**55.67**	**49.95**	**61.46**	**56.89**

Calculation of sample size

The sample size, i.e. the number of cluster (schools) needed for each of the two arms of this study, has been determined based on the following assumptions:

- a minimum detectable effect size of 0.15. This was deemed a worthwhile effect given the estimated cost of the intervention and the cascade delivery of PD within schools. The previous evaluation of ICCAMS (Hodgen et al., 2014) also suggests an effect size of this order is a reasonable target.
- 80% power and alpha of 0.05,
- ICC of 0.12, based on a combined consideration of suggestions/assumptions in relevant literature (Hedges & Hedberg, 2007; Spybrook & Raudenbush, 2009)
- A pre-test post-test design with 0.65 correlation[2]

With these assumptions it was estimated with PASS software (Donner & Klar, 1996; Donner & Klar, 2000) that a minimum of 50 schools will be required per trial arm (assuming number of students in Year7/8 in these schools ranging from 75 to 150 based on the eligibility criteria set during the discussions at that time). A target of 110 schools was thus set for recruitment, which was achieved with 109 schools to be randomised. The average size of recruited schools is more than initially estimated (about 200 per school) therefore power will be recalculated once all the information becomes available.

Outcome measures

Primary outcome

The primary attainment outcome for this evaluation is the raw scores on a slightly modified version of the Mathematics Assessment for Learning and Teaching (MALT) test for Year 8 (MALT 13). Administration of the tests will be directly invigilated by the independent evaluator team with invigilators and markers blind to condition and implemented under exam conditions in schools in June/July 2018.

Baseline measures of attainment will be obtained from the NPD. These will be the scaled scores in maths (KS2_MATSCORE)[3] collected when pupils were in Year 6.

Secondary outcomes

For the impact evaluation the following three secondary outcomes will also be used:

- An attainment sub-scale on MALT of “multiplicative reasoning” (secondary outcome)
- An attainment sub-scale on MALT of “algebra” (secondary outcome)
- Students disposition towards mathematics (secondary outcome)

Students’ disposition towards mathematics is measured with the 9 items presented in Appendix 1. This measure has been validated as part of our recent ESRC work with secondary students and drawing on previous versions for older students (Pampaka et al.,

---

[2] This was informed by a combination of references and guidelines and with a conservative decision in mind to ensure the required MDES https://ies.ed.gov/ncee/pubs/20114033/pdf/20114033.pdf , https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Pre-testing_paper.pdf

[3] As opposed to the raw scores which are not available as in the most recent NPD tables, https://www.gov.uk/guidance/scaled-scores-at-key-stage-2

2013; Pampaka & Wo, 2014). In this work (e.g. as in Pampaka & Wo, 2014) as well as the pilot study of ICCAMS which was distributed to both Year 7 and 8 students we administered 18 items about students feeling towards mathematics and found that such items fall into two related attitudinal measures: maths dispositions and maths self-identification. Previous work has also found the former to be more sensitive to teaching practices (Pampaka, Pepin, & Sikko, 2016; Pampaka & Williams, 2016; Pampaka et al., 2012) and thus we have chosen this measure as a secondary outcome for this evaluation.

Student dispositions will be collected at the start of 2016-17 academic year and at the end of the 2017-18 along with the primary attainment outcome. The first data collection took place via hard copies of surveys posted to schools in September 2017 by the evaluator team along with a free-post address to return. The second survey will be attached to the student test and will be administered along the MALT tests as explained earlier.

The primary and secondary attainment outcome measures will first be analysed using raw scores. However, owing to the small number of items in the secondary outcomes (e.g. MALT subscales) and the ordinal nature of items for the attitudinal measures for students these outcomes will also be calibrated using the Rasch modelling framework (Bond & Fox, 2001; Wolfe & Smith Jr., 2007a, 2007b) and analysed. Rasch modelling allows for objective measurement: the outcomes will thus be similar to standardised scores. Outcomes from both analyses will be compared and reported. This will help understand if the raw scores have violated assumptions of linear regression modelling.

Other measures

Additional measures of teachers' perception of teaching practice have been developed and validated by the evaluator team during the pilot stage. For their development, we drew on our previous work in which we have extensively validated a measure of teachers' perception of transmissionist (or more traditional) teaching practice (Pampaka & Williams, 2016; Pampaka et al., 2012). We hypothesise that involvement with ICCAMS PD will have an effect in this generic measure of teacher's practice. In addition we also acknowledge that there will be more direct effect on teachers' practice of formative assessment (FA) so we have extended our previous instruments with more items focusing on this aspect, drawing on details of how ICCAMS is delivered and other related work (Herman, Osmundson, & Silver, 2010; Wiliam, 2007). We have added such items to our previously validated transmissionist scale hypothesising that there should be a second dimension for practice more related to FA. Examples of new items are listed below:

- I provide feedback to students on their understanding of mathematical concepts
- I check students' understanding for maths during lessons to assess specific intended learning outcomes
- I assess students' maths conceptions and misconceptions in order to adapt my teaching
- I provide feedback on what students have understood in relation to what they should do next
- I encourage students to learn from each other

The results of the pilot study confirmed the validity of two measures of perception of teaching practice, consisting of the items as presented in Appendix 3: (a) perception of transmissionist teaching and (b) perception of teaching in accordance with Formative Assessment. These measures will be used as mediator/moderator variables in the models of effectiveness.

Teachers' confidence in teaching ICCAMS (for intervention teachers): This part of the teacher survey is only aimed for ICCAMS teachers and will be used in the post-survey. The 9 items were derived following previous work with teachers' confidence in teaching Core maths (report in press) – (see Appendix 4).

Students' perceptions of Transmissionist teaching
The student survey will also include a short version of our transmissionist scale with 15 items that were previously found to be giving an equivalent measure of students' perception of transmissionist teaching practice (Pampaka & Williams, 2016). The items are shown in Appendix 2 and the resulting measure will be used as a moderator/mediator variables and as a mitigating measure for the expected lower response in teachers' surveys.

The above measures will all be constructed with the use of the Rasch Model (Winsteps software) (as above, since the items they make up these measures are all in ordinal scale). The scores will be appended to the main dataset with the other outcomes and background variables for the analysis.

Fidelity measure
The relationship between fidelity and attainment outcomes will be analysed quantitatively by examining school fidelity to three key aspects of the ICCAMS intervention: attendance at PD sessions, the delivery of cascade training within schools and the number of lessons taught. These elements will be captured through the end of year (June 2018) teacher survey while the latter will also be captured from school-level report via email to school lead (and follow up during testing at schools by Manchester). The specific items and their scoring are shown in Appendix 5.

We will use Rasch modelling on the items scores to create a uni-dimensional continuous fidelity scale. Pilot analysis with the data collected by Durham in summer 2017 indicated that this is possible. This scale will then be used as a variable in the analyses as shown in the next section. We consider this to be a more meaningful approach than simply aggregating the scores for each item.

However, if it proves not possible to construct a valid uni-dimensional scale[4] we will report the analysis of the three fidelity items separately, placing most emphasis on the delivery of cascade training as the primary fidelity measure, with the other two aspects as secondary measures. Delivery of cascade training was selected by the Delivery and Investigator teams as the primary fidelity measure is this is dependent on the lead teachers attending the PD training and will also be linked to the number of lessons taught; fewer teachers would teach ICCAMS lessons in schools where no cascade training was delivered. It therefore encompasses all three elements to some extent.

Analysis

The standard procedure to be applied for statistical analysis in order to answer the research questions is as follows:

[4] To confirm the unidimensionality we will use the same psychometric approach we use for the other measures in this study (i.e. dispositions) and the Rasch model (and thus the guidelines suggested by Wolfe and Smith Jr (2007a, 2007b)). Decisions will be based on the item fit statistics (to be within acceptable ranges of 0.6 to 1.4) and dimensionality diagnostics, as performed and indicated in our other cited papers and as reported in the pilot of this study and can be shared on request.

- Data cleaning and screening ahead in preparation for analysis
    - Merging of online survey export files and hand input files.
    - Rasch Analysis for measure calibration and construction, for the secondary outcome measures and the other measures described above that will inform explanatory models and the Impact and Process evaluation (a benefit of this is that the analysis can deal with missing data at item level

- Basic descriptive analysis
    - Production of descriptive statistics (e.g. means, standard deviations for both intervention and control group).
- Intention-to-treat analyses: analysis of outcome data according to intention-to-treat principles, e.g. ignoring noncompliance, protocol deviations and other events that take place after randomisation (Gupta, 2011).
- Demonstration of equivalence at baseline (on the basis of baseline attainment and baseline mathematics attitudes)
- Sub-group and further explanatory modelling (as detailed in following sections)

Analysis will need to account for the fact that schools were randomised into groups, while the outcome measures are collected from the students. Therefore multilevel models will be employed to estimate a school-level and a student-level variance, in order to allow for schools to differ regarding their average outcome. Each model will include the outcome of interest as dependent variable (i.e. students' maths raw score) and the following covariates will be included as independent variables: an indicator of group membership (ICCAMS Maths Intervention vs Control), blocking stratifies (i.e. region and FSM) and student's KS2 (i.e. KS2_MATSCORE, as explained earlier) score.

All statistical models detailed next will be performed in Stata 14.

Primary intention-to-treat (ITT) analysis

ITT analysis will be employed in response to **Primary Question**:

> RQ1: Do students in schools implementing ICCAMS Maths over a two-year period demonstrate improvements in overall mathematical attainment compared to students attending control schools?

The unit of analysis is student level outcomes with students nested in control and intervention groups. We will thus apply a two level (random effect) multi-level model, where pupils are clustered at school level as shown with Equation 1.

$$Y_{ij}^{Post} = \alpha_1 + \alpha_2 . Cond_i + \alpha_3 . Y_{ij}^{Pre} + \alpha_4 . Region + \alpha_5 . FSMever + \varepsilon_{ij} \quad \text{(Equation 1)}$$

Where:

Y^{Post} = standardised MALT mathematics scores (raw scores for the primary outcome analysis)

Y^{Pre} = Scaled KS2 mathematics attainment scores

$\mathrm{Cond_i}$ = a dummy variable with the reference category indicating the control group

ε = error term for pupils clustered at school level

i = pupil i

j = school j

As the model above shows the only covariates in this model are the Condition (treatment versus control), KS2 results and other relevant covariates related to randomisation (i.e. regional indicators and FSMever). The coefficient a_2 associated with the condition dummy will be the main result of the trial.

Imbalance at baseline

We will check for any imbalance at baseline in relation to baseline KS2 data (from NPD) with difference being presented as an effect size. Comparisons in background characteristics (e.g. gender, FSMever) between the control and intervention schools, and between regions, will also be made and reported.

Similar comparisons will be performed with students' attitudinal baseline measures.

Missing data

Our approach for dealing with missing data will depend on the extent and the patterns of missingness (i.e. whether data are missing completely at random, at random or not at random). To determine these we will first report complete cases and establish the mechanism of missingness via logistic regression models where the probability of missingness will be modelled (on the basis of responses and complete student lists shared by the schools) with additional predictors (including school and class level, and other available information). If missing (student) cases in intervention and control groups (also considering school attrition) are greater than the 5% threshold which is considered safe for bias as per EEF's guidelines, analysis will proceed depending on the mechanism of missingness:

- If the outcome(s) are found to be missing conditional on covariates we run complete case analysis with covariates and compare with complete cases. If the results are similar then complete case analysis is unlikely to be biased. If results are not similar then missingness is likely not at random and we will proceed with sensitivity analysis.
- If covariates are missing conditional on other covariates or outcomes, then we will run multiple imputation and compare with complete case analysis. As with outcomes, if the results are similar then complete case analysis is unlikely to be biased. If results are not similar then missingness is likely not at random and we will proceed with sensitivity analysis

Sensitivity analysis by comparing a complete case analysis to multiple imputations for the MLM models. Multiple imputation procedures will be performed using the software REALCOM-IMPUTE assuming that the data is missing at random (Carpenter, Goldstein, & Kenward, 2011). Thus we will be able to include partially observed cases (i.e. cases that have not got a value for each of the variables in the model).

Secondary outcome analyses

RQ2: Do students in schools implementing ICCAMS Maths over a two year period demonstrate improvements in attainment in algebra (2a) and multiplicative reasoning (2b), compared to students attending control schools?

RQ4: Do students in schools implementing ICCAMS Maths over a two-year period change their dispositions to learn mathematics compared to students attending control schools?

The secondary attainment outcomes will be analysed in a similar manner to the primary outcome (see Equation 1). Two multilevel models will thus be derived:

- Secondary Outcome model 1 will have the dependent variable: an attainment sub-scale on MALT of "multiplicative reasoning".
- Secondary Outcome model 2 will have the dependent variable: an attainment sub-scale on MALT of "algebra".

Secondary Outcome model 3 (following equation 1) will have the dependent variable: students' mathematics dispositions scores and students mathematics disposition at baseline (collected via surveys in September/October 2016).

Subgroup analyses

In response to RQ3: Are effects on attainment different for students eligible for FSMever? If so, how?

Subgroup Analyses will be performed to answer this research question: The effect of the intervention on attainment (primary and secondary outcome measures) will also be analysed by repeating the primary analysis in the sub-group of students who are eligible for FSM. In other words, only pupils eligible for FSM will be selected to form a new subset. Models with equation 1 and in response to RQ2 above will be re-run without FSMever as a covariate.

Compliance

RQ5: Is there an interaction between fidelity and attainment change for the treatment schools?
A three level model (student – teacher/class – school) will be run further assuming we have at least one fidelity score per teacher. Given the information to be collected at school level, it is envisaged that a class level indicator (see Appendix 5) of compliance will be available and can be used in analysis. As per recent EEF guidelines, we will use group membership as an instrumental variable for the compliance indicator in a two stage least square approach: the first stage will model (predict) the compliance indicator (fidelity measure) using the treatment allocation as instrumental variable alongside all other covariates included in the second stage (as listed in equation 1).

Explanatory analyses

In order to explain the effects of the ICCAMS maths intervention we also intend to run additional models with further covariates at students and school level, which have been shown in previous literature to mediate or moderate such primary and secondary outcomes. The general form of these models are based on Equation 1, and extended to Equation 2:

$$Y_{ij}^{Post} = Y_{ij}^{Post} = \alpha_1 + \alpha_2.Cond_i + \alpha_3.Y_{ij}^{Pre} + \alpha_4.Region + \alpha_5.FSMever + bX_{ij} + \varepsilon_{ij}$$
(Equation 2)

As in equation 1 but with the addition of X as the vector of other related covariates

We thus intend to replicate the models as in equation 1 considering gender, age, and the measures of students' dispositions and perceptions of transmissionist teaching and its interaction with the intervention indicator. We also intend to check for the class/teacher level effect (also with covariates related to the teacher such as their perception of transmissionist teaching, and confidence and perception of teaching for FA, for the intervention teachers).

The models to be explored are summarised in the following table

Models	Outcome	Extra Covariates Condition + KS2 +FSMever+Region +	Structure	Justification
1 to 4	Attainment Algebra Multiplicative Math Disposition	Age +Gender	2 level (student in school)	previous research
5-7	Attainment Algebra Multiplicative	Age+ Gender+Math disposition 1	2 level	Mediating effect of disposition on attainment
8-11	Attainment Algebra Multiplicative Math Disposition	+ Age+ Gender+ TransTeaching 2 (we will also control for teaching at baseline by including TransTeaching1)	2 level	Mediating effect of students' teaching perceptions
12-19	Attainment Algebra Multiplicative Math Disposition	+ Age+ Gender+ TransTeaching2 + TeachersFA Age+ Gender+ TransTeaching 2+TeacherTrans+TeachersFA (we will control for teaching at baseline)	3 level	Mediating effect of teachers' teaching perceptions
19-21	Attainment Algebra Multiplicative Math Disposition	Age+ Gender+ TransTeaching 2+TeacherTrans+TeachersFA +Confidence TeachICCAMS	3 level – ICCAMS only	For confidence in teaching
22-25	Attainment Algebra Multiplicative Math Disposition	+Fidelity interactions (see p. 7 and Appendix 5)	3 level – ICCAMS only	For interaction of fidelity/compliance

Effect size calculation

As most of the models we will be using are multilevel, effect sizes will be calculated following the guidelines set by EEF for multilevel models[5]. As suggested, for cluster randomised trials, the total variability is decomposed into random variation between pupils (σ^2_{error}) and heterogeneity between schools (σ^2_S). Effect sizes will thus be calculated as follows (assuming equal cluster size and using total variance):

---

5 https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf

$$ES = \frac{(\bar{Y}_{intervention} - \bar{Y}_{control})_{adjusted}}{\sqrt{\sigma_s^2 + \sigma_{error}^2}}$$ where the $(\bar{Y}_{intervention} - \bar{Y}_{control})_{adjusted}$ denotes the mean differences between intervention and control groups adjusting for baseline scores and other stratification variables (and the covariates defined earlier, e.g. see equation 1)

A standardised effect size will be calculated and reported using Hedge's *g* for consistency (even though small sample bias correction will not be needed in this study) as shown below. Cohen's *d* is first calculated as the difference between the two groups means divided by the pooled standard deviation, with the formula presented next (with an example with teachers scores).

$$SD_{pooled} = \frac{\sqrt{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}}{n_1 + n_2 - 2}$$

Where

n_1 = number of teachers in the treatment group

n_2 = number of teachers in the control group

s_1 = standard deviation of teachers scores in the treatment group

s_2 = standard deviation of teachers scores in the control group

the following correction factor will then be applied to Cohen's *d* to result to Hedge's *g*:

$$G = D\left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

Effect sizes will be accompanied by 95% confidence intervals as per EEF specifications (https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Analysis_for_EEF_evaluations_REVISED2.docx.pdf) . In particular, statistical significance will be assessed using two-sided tests at the 5% level unless otherwise stated. Estimates of effect with 95% confidence intervals (CIs) and p-values will be provided as appropriate.

Report tables

The EEF trial report template[6] contains several tables whose structure is pre-specified. Evaluators should paste these into the SAP and populate them with their chosen variables. Templates for any tables and charts additional to those in the report template should also be specified in the SAP.

Example Table 1: Summary of impact on primary outcome

[6] https://educationendowmentfoundation.org.uk/evaluation/resources-centre/writing-a-research-report/

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	EEF cost rating
ICCAMS vs. control				
ICCAMS FSM vs. control				

Example Table 3: Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Correlation between pre-test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	110/11000 (55/5500 each arm)	0.65	0.12	GCSE and FSM	80%	0.05	0.15
Randomisation	109						
Analysis (i.e. available pre- and post-test)							

Example Table 4: Baseline comparison

Variable	Intervention group		Control group	
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
School type Academy Converter Academy sponsor led …				
Ofsted rating Outstanding Good				

Location Urban Rural				
	…	…	…	…
School-level (continuous)	**n (missing)**	**[Mean or median]**	**n (missing)**	**[Mean or median]**
Number of Y17 pupils (at 2016) **Attainment** **% attaining level 4 in KS2 Maths**				
FSM eligibility				
	…	…	…	…
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM				
Male				
	…	…	…	…
Pupil-level (continuous)	**n (missing)**	**[Mean or median]**	**n (missing)**	**[Mean or median]**
Pre-test score (KS2) **Maths disposition at baseline**				

Example Table 5: Primary analysis

	Raw means				Effect size		
	ICCAMS Maths group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
MALT 13							

References

Bohlig, M., Fisher, W. P. J., Masters, G. N., & Bond, T. (1998). Content Validity and Misfitting Items. Rasch Measurement Transactions, 12(1), 607.

Bond, T. G., & Fox, C. M. (2001). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. NJ: Lawrence Erlbaum Associates Inc.

Carpenter, J. R., Goldstein, H., & Kenward, M. G. (2011). REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. Journal of Statistical Software, 45(4), 1-14.

Donner, A., & Klar, N. (1996). Statistical Considerations in the Design and Analysis of Community Intervention Trials. The Journal of Clinical Epidemiology, 49(4).

Donner, A., & Klar, N. (2000). Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold.

Gupta, S. K. (2011). Intention-to-treat concept: A review. Perspectives in Clinical Research, 2(3), 109-112. doi:10.4103/2229-3485.83221

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass Correlation Values for Planning Group-Randomized Trials in Education. Educational Evaluation and Policy Analysis, 29(1), 60-87. doi:10.3102/0162373707299706

Herman, J., Osmundson, E., & Silver, D. (2010). Capturing quality in formative assessment practice: Measurement challenges (CRESST Report 770). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Linacre, J. M. (1998). Detecting Multidimensionality: Which Residual Dara-type Works Best? Journal of Outcome Measurement, 2(3), 266-283.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational Measurement (Third ed., pp. 13-103). USA: American Council of Education and the Oryx Press.

Pampaka, M., Pepin, B., & Sikko, S. A. (2016). Supporting or alienating students during their transition to Higher Education: mathematically relevant trajectories in the contexts of England and Norway. International Journal of Educational Research.

Pampaka, M., & Williams, J. (2016). Mathematics teachers' and students' perceptions of transmissionist teaching and its association with students' dispositions. Teaching Mathematics and its Applications. doi:10.1093/teamat/hrw007

Pampaka, M., Williams, J. S., Hutcheson, G., Black, L., Davis , P., Hernandez-Martinez, P., & Wake, G. (2013). Measuring Alternative Learning Outcomes: Dispositions to Study in Higher Education. Journal of Applied Measurement, 14(2), 197-218.

Pampaka, M., Williams, J. S., Hutcheson, G., Wake, G., Black, L., Davis, P., & Hernandez - Martinez, P. (2012). The association between mathematics pedagogy and learners' dispositions for university study. British Educational Research Journal, 38(3), 473-496.

Pampaka, M., & Wo, L. (2014). Revisiting Mathematical Attitudes of students in Secondary Education. In Liljedahl, P., Oesterle, S., Nicol, C., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36, Vol. 4, pp. 385-392. . Paper presented at the The Joint Meeting of PME 38 and PME-NA 36, Canada, Vancouver.

Spybrook, J., & Raudenbush, S. W. (2009). An Examination of the Precision and Technical Accuracy of the First Wave of Group-Randomized Trials Funded by the Institute of Education Sciences. Educational Evaluation and Policy Analysis, 31(3), 298-318. doi:10.3102/0162373709339524

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 67-114). London Lawrence Erlbaum Associates, Publishers.

Torgerson, D. J., & Torgerson, C. J. (2008). Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction: Palgrave Macmillan.

Wiliam, D. (2007). Five "Key Strategies" for Effective Formative Assessment. National Council of Teachers of Mathematics (Research Brief).

Wolfe, E. W., & Smith Jr., E. V. (2007a). Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part I - Instrument Development Tools. Journal of Applied Measurement, 8(1), 97-123.

Wolfe, E. W., & Smith Jr., E. V. (2007b). Instrument Development Tools and Activities for Measure Validation Using Rasch Models: Part II - Validation Activities. Journal of Applied Measurement, 8(2), 204-234.

Wright, B. D., & Masters, G. N. (1982). Rating Scale Analysis. Chicago: MESA Press.

Appendices

Appendix 1: The Items for Students' Disposition towards Mathematics

[Please circle the appropriate number in each line]		**Scale**
1.	Mathematics is important to me	Strongly Disagree (1), Disagree (2), Unsure (3), Agree (4), Strongly Agree (5)
2.	Learning maths is enjoyable for me	
3.	I am interested in learning new things in maths	
4.	I never want to take another mathematics course	
5.	I prefer my future studies to include a lot of maths	
6.	I look forward to studying more mathematics after school	
7.	I would like to be a mathematician	
8.	Maths is one of the most interesting school subjects	
9.	Maths is important for my future (after school)	

Appendix 2: The Items for Students' 'Perceptions of transmissionist teaching practice'

In this section, we want to find out how you are taught maths in general. Please tell us how often does each of the following happen in your normal weekly maths lessons? *[Please circle the appropriate number for each line]*	
1.	We (students) use only the methods the teacher taught us.
2.	We choose which questions to tackle.
3.	We compare different methods for doing questions.
4.	The teacher draws links between different topics.
5.	We work collaboratively in small groups.
6.	We discuss our own ideas.
7.	We work collaboratively in pairs.
8.	We invent our own methods.
9.	The teacher tells us which questions to tackle.
10.	The teacher asks questions to check what we understood.
11.	The teacher teaches each topic separately.
12.	What we learn is related to everyday real life situations.
13.	We learn from each other (NEW – for FA)
14.	We explain our work to the whole class.

15.	The teacher questions our methods.

Appendix 3 – Items for Teachers' Perceptions of Teaching Practice Measures

ID	Item	FA Practice	Transmissionist
1	I introduce a new topic by first determining what the students already know about it	✓	✓ R
2	I use activities in contexts that the students can engage with	✓	✓ R
3	I use activities which allow connections to be made between mathematical ideas	✓	✓ R
4	I allow students to work at their own pace		✓ R
5	I teach the whole class at once	✓ R	✓
5	Students start with easy questions and work up to harder questions	✓	✓
7	When a student asks a question, I give clues instead of the correct answer	✓	✓ R
8	I ask students to explain their reasoning when giving an answer		✓ R
9	I encourage students to discuss the mistakes they make		✓ R
10	Students use only the methods I taught them		✓
11	Students choose which questions to tackle		✓ R
12	Students compare different methods for doing questions		✓ R
13	Students work collaboratively in small groups.	✓	✓ R
14	Students discuss their ideas.		✓ R
15	Students work collaboratively in pairs.	✓	✓ R
16	Students invent their own methods.	✓	✓ R
17	I tell students which questions to tackle.	✓ R	✓
18	I teach each topic separately		✓
19	I provide feedback to students on their understanding of mathematical concepts	✓	
20	I check students' understanding for maths during lessons to assess specific intended learning outcomes	✓	
21	I assess students' maths conceptions and misconceptions in order to adapt my teaching	✓	
22	I provide feedback on what students have understood in relation to what they should do next	✓	
23	I encourage students to learn from each other	✓	

R-indicates that reverse coding is needed for the construction of one-dimensional measures.

For FA practice: Higher scores mean more formative assessment practice

For Transmissionist: Higher scores mean more transmissionist teaching

Appendix 4: Items for Teachers' Confidence in Teaching ICCAMS

The next question is only for ICCAMS teachers (teachers who used the ICCAMS material); if you have not used them please skip this part

Considering the ICCAMS material and lessons you taught, please tell us how much you agree with the following statements.

(Please circle the appropriate number in each line)	SD	D	N	A	SA
I feel confident teaching the ICCAMS lessons.	1	2	3	4	5
Teaching ICCAMS lessons is no more demanding for me than the other lessons I am teaching.	1	2	3	4	5
The materials for ICCAMS have helped me feel confident.	1	2	3	4	5
Other support for ICCAMS has helped me feel confident.	1	2	3	4	5
The training I received was useful	1	2	3	4	5
Teaching ICCAMS lessons matches my teaching skills and experience well.	1	2	3	4	5
I would feel confident to teach these lessons again next year if I am asked to do.	1	2	3	4	5
I would prefer to teach the ICCAMS lessons instead of other maths courses/units, if I had a choice.	1	2	3	4	5

Appendix 5: Items for Fidelity School Measure

These three key aspects, and the initial proposed fidelity levels, are as follows:
<u>Attendance at PD Sessions</u>

High: 16-18 sessions attended in total between the 2 lead teachers

Medium: 10-15 sessions attended

Low: 1-9 sessions attended

This data is gathered throughout the 2 years by PD registers. Ignore any changes in staffing; it

doesn't matter if the attendee isn't the official/original lead teacher, as long as somebody attends.

Context: There are 9 PD sessions with an expected 2 lead teachers to attend each.

<u>Cascade training</u>

The year 2 lead teacher survey will ask a new question: 'Between both lead teachers, have you taught all of the cascade sessions over the 2 years?'

High: 'Yes'

Medium: ‘No, missed up to 2’

Low: ‘No, missed more than 2’

Data gathered at end of year 2. The year 1 survey data was difficult to use; this question is designed to capture this information in a more simple way.

Lessons taught

The year 2 teacher survey will ask both lead and cascade teachers how many ICCAMS lessons they have delivered to each of the classes they teach (see question below). High, Medium and Low fidelity at school level will be defined based on the total numbers taught (classrooms x lessons) and with the following cut offs (as averages):

High: 18-20 lessons per class

Medium: 15-17 lessons per class

Low: <15 lessons per class

Context: Teachers are expected to teach 20 lessons over the 2 years; but some of these can be options to repeat earlier lessons. By the time this survey is sent to schools, it will be expected that they all should be finished ICCAMS teaching.

Question in Teacher Survey:

Please list the Year 8 class you teach this year and tell is whether you have delivered ICCAMS and if so how many lessons.

Class Name (please list)	**ICCAMS taught?**	**How many ICCAMS lessons have you delivered? (out of the 20 specified)**
	Yes/No	
	Yes/No	
	Yes/No	
	Yes/No	
	Yes/No	
	Yes/No	
	Yes/No	
	Yes/No	