The Student Grouping Study

Further Appendices

April 2026

Jeremy Hodgen, Becky Taylor, Antonina Tereshchenko, Jake Anders, Laurie Jacques, Maria Cockerill, Rosa Kwok, David Sirl, Hester Burn, and Nicola Bretscher

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges and early years settings to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence**. Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds. Putting evidence to use.
- **Putting evidence to use**. Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were reendowed with an additional £137m from government, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Department for Education

What Works Network

Table of contents

Education Endowment Foundation

UCL Institute of Education

Appendix C.1: Memorandum of understanding and data sharing agreement.

The Student Grouping Study:
Investigating the impact of setting and mixed-attainment grouping

MEMORANDUM OF UNDERSTANDING

Aims of the study

UCL Institute of Education are investigating whether it is more effective to group pupils in sets or in mixed attainment groups. The research is funded by the Education Endowment Foundation.

The project

The project is investigating the effect of setting and mixed attainment (mixed ability) grouping in maths on pupils' attainment and self-confidence. The project will involve all the Year 7 classes starting secondary school in September 2022 from around 120 schools. The project will last until the end of Year 8 (Summer 2024). We are comparing schools that already group pupils for maths in mixed attainment groups to schools that already group pupils for maths in sets. **We are not asking you to change anything about what you already do in your school.**

This memorandum of understanding (MoU) explains what your school's participation in the study will entail. If you agree to take part and accept the terms and conditions outlined, please sign a copy of this form, and return by email to IOE.groupingstudents@ucl.ac.uk .

Structure of the study

This is an observational study with matched schools design. Schools are eligible for participation based on their current grouping practices for mathematics in Year 7 and Year 8. Each school that uses mixed attainment grouping in maths will be matched to two similar schools that use sets.

1. **Mixed attainment group** - Schools in this group will teach mathematics to mixed attainment groups in Year 7 (2022-2023) and Year 8 (2023-2024). Pupils grouped in 'mixed attainment groups' are grouped so that the range of prior attainment in each teaching group is similar to that of the year group as a whole. Schools may additionally have a 'nurture group' for the lowest attaining pupils.
2. **Setting group** - Schools in this group will teach mathematics to attainment sets in Year 7 (2022-2023) and Year 8 (2023-2024). Schools in this group will be matched to schools in the mixed attainment group. Pupils grouped in 'attainment sets' are grouped by attainment/'ability' in mathematics only for teaching in mathematics. Schools grouping in streams are not eligible to participate.

All schools in both groups will continue with their usual grouping and teaching practices for the duration of the study (September 2022 – July 2024).

The Study Team (IOE) will use school and pupil information provided by schools, and information from the National Pupil Database to assess the effectiveness of mixed attainment grouping and setting.

All schools will receive £1000 following the completion of all study requirements with staff/school and with the required pupils.

Ethical approval

The study has been granted full ethical approval by UCL Institute of Education Research Ethics Committee, reference REC 1139. If this study has harmed you in any way or if you wish to make a complaint about the conduct of the study you can contact Professor Phil Jones, Chair of the UCL Institute of Education Research Ethics Committee, using this email address: IOE.researchethics@ucl.ac.uk

Use of Data

The study has been approved by UCL Data Protection team, reference Z6364106/2018/11/03.

For the purpose of research and archiving, survey and test data collected from pupils will be linked with information about the pupils from the National Pupil Database (NPD) and shared with the Department for Education, the EEF, the EEF's archive manager and with the Office for National Statistics and potentially other research teams. Further matching to NPD and other administrative data may take place during subsequent research.

All pupil data will be treated with the strictest confidence and will be stored in accordance with the data protection legislation, including the General Data Protection Regulation (GDPR). Personal data will be processed as per condition 6(1)e of the GDPR under public interest purposes, because the research is considered to be a "task carried out in the public interest". The research will involve some 'special category personal data' (ethnicity) which will be collected from the NPD. For these data, we rely upon Article 9(2)(j) as the legal basis for data processing, i.e. where the processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. To ensure that parents/carers, and the pupils themselves, are comfortable for their, or their child's, data to be processed in this way, we are also providing an opportunity for parents/carers to discuss the research with their child and to withdraw their data from the research and any data processing (see requirements below for details).

All results will be anonymised so that no schools or individual pupils or teachers will be identified in any report arising from the research.

Further information about how UCL uses participant information can be found here: https://www.ucl.ac.uk/legal-services/privacy/ucl-general-research-participant-privacy-notice

If you are concerned about how personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk

Requirements for All Schools

- All schools are required to return pupil information for all Year 7 pupils starting in September 2022 (pupil first and surnames, date of birth, Unique Pupil Number (UPN), pupils' mathematics class and teacher) to the Study Team in September – October 2022.
- All schools will provide the study team with contact details for their data manager to facilitate the return of the requested data.
- All schools will sign a data-sharing agreement with UCL (appended to this Memorandum of Understanding) to establish the terms and conditions for sharing of personal data.
- All schools will deliver letters to parents giving them information about the study and an opportunity to discuss the research with their child and withdraw them from data processing in line with GDPR. Schools will inform the Study Team of any withdrawals by the end of September 2022.

- All schools agree to the Study Team obtaining the relevant pupils' KS2 mathematics and English scores and demographic data (gender, FSM status, ethnicity) from the National Pupil Database. The school's data manager or equivalent will provide the information requested above, i.e. pupils' names, date of birth and UPNs to enable this to be achieved (unless a parent/carer or pupil has requested that their data be withdrawn from processing, in line with GDPR).
- All schools will follow UCL Institute of Education guidance on the secure transfer of data.
- Facilitate a short, paper-based questionnaire for all Year 7 pupils in September 2022.
- Facilitate an online survey for all Year 8 pupils in June 2024.
- Facilitate all Year 8 pupils to take the GL Assessment Progress Test in Mathematics in Summer 2024. This will be at no cost to the school and will be supported by an administrator provided by the project.
- All mathematics teachers will be invited to complete questionnaires towards the end of the project (June 2024).
- The Head of Mathematics will be invited to complete a questionnaire during the project.
- Members of the Senior Leadership Team will be invited to complete a short questionnaire during the project.
- Some schools will be invited by the Study Team to become case study schools during the course of the project, although agreeing to do so is not a requirement of the study. Case study schools will allow the research team access to collect data (for example through observations and interviews). An additional payment of £500 is available to case study schools to facilitate data collection.
- To work closely with the Study Team.
- **If the school has to withdraw from the project for operational or other unavoidable reasons, it will notify the Study Team straight away and wherever possible still provide test data for the evaluation.**

Responsibilities of the Study Team:

- Act as the first point of contact for any questions about the study
- Provide information sheets and withdrawal forms for parents/carers
- Provide guidance to schools on how to collect and return data safely and securely
- Collect class and pupil level data including pupil names, date of birth, gender, Unique Pupil Number (UPN), eligibility for FSM, KS2 mathematics and English scores, pupil mathematics class and teacher
- Organise the distribution (to schools) and subsequent collection of a short questionnaire for all Year 7 pupils in Autumn 2022 and an online survey for Year 8 pupils in Summer 2024
- Make arrangements for all Year 8 pupils to take the GL Assessment Progress Test in Mathematics in Summer 2024
- Conduct surveys with mathematics teaching staff, Heads of Mathematics and at least one member of the Senior Leadership Team
- Contact schools asking them to be involved as a case study school during the Summer term 2023

- Request NPD data using pupil details
- Analyse the data from the project
- Disseminate the research findings

Education Endowment Foundation

UCL Institute of Education

Head Teacher agreement

☐ **I agree for my school to take part in the Student Grouping Study and I accept the eligibility terms and conditions listed above.**

☐ **I understand and agree that pupil data will be processed as per condition 6(1)e of the GDPR under public interest purposes and that 'special category personal data' (ethnicity) will be processed as per Article 9(2)(j) 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.'**

School Name			
Head Teacher Name			
Head Teacher Signature		Date	__ / __ / __
Head Teacher Email Address			
Head of Mathematics Name			
Head of Mathematics Email Address			
School Contact (if not Head Teacher)			
School Contact Email Address (if not Head Teacher)			
School Telephone Number			
Data Manager Name			
Data Manager Email Address			

Please answer the following questions about your school.

LA area and County	
School LA Establishment/DFE Number (a seven digit number)	
School admin email	
% students receiving Pupil Premium (Ever6 FSM)	
Ofsted grade	

Please confirm your school's grouping practices for mathematics in Year 7 and Year 8.

	Year 7	Year 8
All groups are completely mixed attainment		
Most groups are completely mixed, with a nurture group for very low-attaining pupils		
Most groups are completely mixed with a top group for the highest-attaining pupils		
Most groups are completely mixed with a top and a bottom group		
All groups are set by attainment/ability		

This MOU constitutes the school's agreement with UCL Institute of Education to participate in the Student Grouping Study.

Thank you for agreeing to take part in this research.

Please complete all information and return this form to: IOE.groupingstudents@ucl.ac.uk

Education Endowment Foundation

UCL Institute of Education

LONDON'S GLOBAL UNIVERSITY

Data Sharing Agreement

between

University College London

and

[SCHOOL NAME]

Date this Agreement comes into force: 1 September 2022

Parties to this Agreement

(a) **UNIVERSITY COLLEGE LONDON** a body corporate established by Royal Charter with company number RC000631 of Gower Street, London, WC1E 6BT (**UCL**); and

(b) [SCHOOL NAME] **[DESCRIPTION, COMPANY NUMBER AND REGISTERED ADDRESS]** (**[X]**).

Education Endowment Foundation

UCL Institute of Education

Purpose

(a) This Agreement establishes the terms and conditions under which the parties will share personal data in connection with the Student Grouping Study. The aim of this project is to investigate whether it is more effective to group students in sets or in mixed attainment groups. Personal data (names, dates of birth, UPN, mathematics class) will be processed as per condition 6(1)e of the GDPR under public interest purposes, because the research is considered to be a "task carried out in the public interest".

(b) The parties shall share the personal data described in 2(a) above only in accordance with the terms of this Agreement.

Term and termination

(a) This Agreement shall commence on the date set out at the beginning of it and shall continue until the end of this agreement unless terminated earlier in accordance with its terms.

(b) Either party may terminate this Agreement with immediate effect by giving written notice to the other party if that other party commits a material breach of any term of this Agreement which breach is irremediable or (if such breach is remediable) fails to remedy that breach within a period of 30 days after being notified in writing to do so;

(c) Clause 3 (Term and termination) and Clause 4 (Data protection) shall survive the termination or expiry of this Agreement, as shall any other Clause which, by its nature, is intended to survive termination or expiry.

(d) Termination or expiry of this Agreement shall not affect any rights, remedies, obligations or liabilities of the parties that have accrued up to the date of termination or expiry, including the right to claim damages in respect of any breach of the Agreement which existed at or before the date of termination or expiry.

Data protection

(a) In this Clause, the following terms have the following meanings:

(i) **Controller** means a person which, alone or jointly with others, determines the purposes and means of the Processing of Personal Data;

(ii) **Data Protection Laws** means all applicable statutes and regulations in any jurisdiction pertaining to the processing of Personal Data, including but not limited to the privacy and security of Personal Data;

(iii) **Data Subject** means the individual to whom the Personal Data relates;

(iv) **Personal Data** means any information relating to an identified or identifiable living individual;

(v) **Processing** means any operation or set of operations which is performed on Personal Data or on sets of Personal Data, whether or not by automated means, and Process, Processes and Processed shall be construed accordingly; and

(vi) **Personal Data Breach** means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, Personal Data transmitted, stored or otherwise processed.

(b) The Parties acknowledge and agree that in respect of the Personal Data disclosed by one Party to the other in connection with this Agreement:

i. the UCL is a Controller in respect of the Personal Data it Processes;
ii. the **[NAME OF OTHER DATA CONTROLLER]** is a Controller in respect of the Personal Data it Processes;
iii. the Parties are not joint Controllers; and
iv. neither Party Processes any Personal Data on behalf of the other Party as a Processor.

(c) In respect of the Personal Data a party Processes under or in connection with this Agreement, the party shall:

(i) comply at all times with its obligations under the Data Protection Laws;

(ii) notify the other party without undue delay after becoming aware of a Personal Data Breach; and

(iii) assist and co-operate fully with the other party to enable the other party to comply with their obligations under Data Protection Law, including but not limited to in respect of keeping Personal Data secure, dealing with Personal Data Breaches, complying with the rights of Data Subjects and carrying out data protection impact assessments.

(d) The parties shall work together to ensure that each of them is able to Process the Personal Data it Processes under or in connection with this Agreement for the purposes contemplated by this Agreement lawfully, fairly and in a transparent manner and in compliance with the Data Protection Laws. This shall include but not be limited to entering into such other written agreements as may be required from time to time to enable each party to comply with the Data Protection Laws.

Miscellaneous

(a) No variation of this Agreement shall be effective unless it is in writing and signed by the parties (or their authorised representatives).

(b) A failure or delay by a party to exercise any right or remedy provided under this Agreement or by law shall not constitute a waiver of that or any other right or remedy, nor shall it prevent or restrict any further exercise of that or any other right or remedy. No single or partial exercise of any right or remedy provided under this agreement or by law shall prevent or restrict the further exercise of that or any other right or remedy.

(c) If any provision or part-provision of this Agreement is or becomes invalid, illegal or unenforceable, it shall be deemed modified to the minimum extent necessary to make it valid, legal and enforceable. If such modification is not possible, the relevant provision or part-provision shall be deemed deleted. Any modification to or deletion of a provision or part-provision under this Clause shall not affect the validity and enforceability of the rest of this Agreement.

(d) This Agreement constitutes the entire agreement between the parties and supersedes and extinguishes all previous agreements, promises, assurances, warranties, representations and understandings between them, whether written or oral, relating to its subject matter.

(e) Each party agrees that it shall have no remedies in respect of any statement, representation, assurance or warranty (whether made innocently or negligently) that is not set out in this Agreement.

(f) Nothing in this Agreement is intended to, or shall be deemed to, establish any partnership or joint venture between any of the parties, constitute any party the agent of another party, or authorise any party to make or enter into any commitments for or on behalf of any other party.

(g) This Agreement does not give rise to any rights under the Contracts (Rights of Third Parties) Act 1999 to enforce any term of this Agreement.

(h) This Agreement may be executed in any number of counterparts, each of which when executed shall constitute a duplicate original, but all the counterparts shall together constitute the one Agreement.

(i) This Agreement and any dispute or claim (including non-contractual disputes or claims) arising out of or in connection with it or its subject matter or formation shall be governed by and construed in accordance with English law.

(j) Each party irrevocably agrees that the courts of England and Wales shall have exclusive jurisdiction to settle any dispute or claim (including non-contractual disputes or claims) arising out of or in connection with this Agreement or its subject matter or formation.

Appendix C.2: Parent information sheet and withdrawal form

The Student Grouping Study: Investigating the impact of setting and mixed-attainment grouping

Information for Parents/Carers

Research Ethics Committee Approval Number: REC 1139

Data Protection reference: Z6364106/2018/11/03

Principal Researchers: Prof Jeremy Hodgen & Dr Becky Taylor IOE.groupingstudents@ucl.ac.uk

What is this about?

UCL Institute of Education is carrying out a project funded by the Education Endowment Foundation (EEF) which aims to improve our understanding about the most effective ways to group students for teaching. This research has been reviewed and approved by the research ethics committee of UCL Institute of Education. The headteacher of your child's school has agreed that the school will take part in the research programme.

What will the project look like?

The project investigates the effect of different approaches to grouping students in maths. We are comparing schools that group students in sets with schools that use mixed attainment grouping. We have chosen your child's school because of the approach to grouping that it already uses. We are not asking schools to make any changes to what they already do.

We plan to work with around 120 schools, and the Year 7 students who start secondary school in September 2022. The project will continue until Summer 2024, when students finish Year 8.

What does this mean for me as a parent?

As part of measuring the effect of your school's approach to grouping in maths, your child will be asked to complete a short survey at the start of Year 7, and another survey towards the end of Year 8. The short survey will take about 15 minutes and the longer one will take about half an hour. We will also ask your child to complete a short maths test at the end of Year 8. The maths test will be marked by the test organisation and the marks shared with the research team. We are collecting this information for the purposes of the research project, to help us understand which type of grouping in mathematics is most helpful for children in Years 7 and 8.

We will ask the school for your child's name, date of birth, maths class and the name of their maths teacher. We will also ask the school for your child's UPN (Unique Pupil Number) to allow us to link the data with the National Pupil Database (held by the Department for Education, part of the UK Government). We will access your child's Key Stage 2 attainment data, gender, free school meal status and ethnicity from the NPD, but children's names and schools will not be available to us at that stage. Data will be analysed in an anonymous form so your child cannot be identified during analysis. For the purposes of research, your child's data will be shared with the Department for Education, the EEF's archive manager and, in an anonymised form, with the Office for National Statistics and potentially other research teams. Further matching to the National Pupil Database and other administrative data may take place during subsequent research.

No information that can identify individual children will be made available to anyone outside the project team, UCL's survey and testing contractors, the Department for Education and your child's school. Your child's data will be treated with the strictest confidence and will be kept behind secure firewalls. We will not use your child's name or the name of the school in any report arising from the research, and no information that could otherwise identify your child will be made public.

Because we are doing this research to improve understanding about what works in improving students' education, **if you are happy for information about your child to be used in this research project you do not need to do anything.** Thank you for your help with this research, your support is much appreciated.

This is an important project that we think will help improve teaching and learning. We expect that your child will enjoy their involvement in the project and they will be free to withdraw at any time. If you would prefer your child **NOT** to take part in any project surveys or testing, or their data not to be processed as above, please complete the enclosed form and return it to your child's school by Monday 12 December 2022.

If you have any questions you would like to ask, please contact Becky Taylor at the UCL Institute of Education by email at ioe.groupingstudents@ucl.ac.uk

---

Data Protection Privacy Notice

The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk. UCL's Data Protection Officer can also be contacted at data-protection@ucl.ac.uk.

Further information on how UCL uses participant information can be found here: https://www.ucl.ac.uk/legal-services/privacy/ucl-general-research-participant-privacy-notice The legal basis that would be used to process your child's personal data will be performance of a task in the public interest. The legal basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your child's personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this, and will try to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.

The Student Grouping Study

(If you are happy for your child to participate in the research on grouping in maths, you DO NOT need to return this form.)

I **DO NOT** wish data about my child to be collected as part of this research.

Child's name: ....Date of birth: ....

Child's maths teacher: ....

School:..

Parent name (BLOCK CAPITALS) ....

Parent signature: ....

Date ....

(Please detach and return the completed form to your child's maths teacher by Monday 12 December 2022).

Appendix C.3: Factors influencing decisions around grouping.

Factor	Justification	Indicator(s)	Issues	Influence school decision on grouping practice?	Impact on outcomes?	Influenced by outcomes?	Use for matching?
Prior attainment	Aim to raise attainment across the board	Average KS2	Include historic data	Yes	Yes	No	Yes
	Low prior attainers – aim to close attainment gap / raise attainment of low prior attainers	Spread (SD) of prior attainment Or % of low prior attainers		Yes	Yes	No	Yes
	High prior attainers – maintain / raise attainment of high prior attainers	Spread (SD) of prior attainment Or % of high prior attainers		Yes	Yes	No	Yes
	Aim to close attainment gap between FSM and non-FSM at entry	FSM gap at KS2		Yes	Yes	No	Yes

Factor	Justification	Indicator(s)	Issues	Influence school decision on grouping practice?	Impact on outcomes?	Influenced by outcomes?	Use for matching?
Attainment outcomes (KS4)	Aim to raise attainment outcomes across the board	A8 P8 En/Ma GCSE	Outcome (e.g., GCSE) data would affect the decision, but, as above, is not exogenous to treatment. Over a period of time this might be influenced by outcomes but probably over a long period of time.	Yes	Yes	Yes	No
	Low prior attainers – aim to close attainment gap / raise attainment of low prior attainers	Low prior attainers: A8 P8 En/Ma GCSE Spread (SD) of GCSE outcomes	As above.	Yes	Yes	Yes	No
	High prior attainers; Maintain / raise attainment of high prior attainers	High prior attainers: A8 P8 En/Ma GCSE Spread (SD) of attainment	As above	Yes	Yes	Yes	No

Factor	Justification	Indicator(s)	Issues	Influence school decision on grouping practice?	Impact on outcomes?	Influenced by outcomes?	Use for matching?
Student characteristics	Aim to close attainment gaps	% FSM6 at intake	Over time, proportion of disadvantaged students could be affected by outcomes (e.g., as a result of changes to value added, or overall GCSE attainment, the school could become more or less attractive to disadvantaged or advantaged parents). To avoid this, take an average measure over time or a measure at some time point in the past (e.g. 5 years previously). Concern for FSM outcomes in high % FSM schools *may* make mixing more likely to avoid stigmatising low-attaining students receiving FSM or because of focus on raising attainment of FSM.	Yes	Yes	No	Yes
		% BAME	As for FSM. An indicator of greater diversity – can drive segregation (or forced integration)	Yes	Yes	No	Yes
		% EAL	As for FSM. Schools may choose to group learners by different stages of learning English.	Yes	Yes	No	Yes

Factor	Justification	Indicator(s)	Issues	Influence school decision on grouping practice?	Impact on outcomes?	Influenced by outcomes?	Use for matching?
		% SEND	As for FSM. Parents of some SEND students do have choice, so this may be influenced by the outcomes (though v small numbers – only those with EHCP). But we could avoid this by taking a measure over time or at some point in the past as for FSM? BUT would it be a major factor in a school's decision?	Yes	Yes	May have small effect	No
		% in high and low IDACI neighbourhoods		Yes	Yes	No	Yes
		Gender ratio	Influence on decision unclear. Possible influence for co-educational schools with highly unbalanced intake, but otherwise may not be an influence. May be difficult for stakeholders to understand if we don't match on this.	Unlikely	Yes	No	No
School capacity to implement change	Schools unlikely to *change* practices or take on 'riskier' MA	Age range (sixth form or not)	May be an indicator of capacity to change, but unlikely to affect decision of grouping practices at KS3	Yes	Yes	No	No

Factor	Justification	Indicator(s)	Issues	Influence school decision on grouping practice?	Impact on outcomes?	Influenced by outcomes?	Use for matching?
	grouping without capacity to do so. Lots of unobservables, but some observable factors are potential proxies	Size of school	Cannot set if cohort very small. Need at least 75 students/3 teaching groups in year group to make setting at 3 levels possible – implications for recruitment criteria.	Yes	Yes	No	Yes
		OFSTED grade	May be an indicator of capacity to change, but OFSTED grade may be influenced by outcomes / value added	Yes	Yes	??	Yes
		Academy Status	May be an indicator of capacity, and willingness, to change	Yes	??	No	Yes
		MAT membership	May be linked to capacity or impetus to improve.	Yes	Maybe	Yes	No
		Urban / rural	Level of competition between schools likely to influence decisions on grouping	Yes	No	No	Yes
Ethos / Values of SLT	MA (or setting) could be a values-driven decision	School policy documents Mission statement Soft Ofsted grades	Difficult to observe without cost Also likely endogenous variable	Yes	Yes	Maybe	No

Appendix C.4: Secondary outcome measure: student self-confidence scale

Self-Confidence in Mathematics

Items marked with an asterisk (*) are reverse-scored.

- Work in maths is easy for me
- I am not very good at maths*
- Maths is one of my best subjects
- I hate maths*
- I do well at maths
- I get good marks in maths
- I learn things quickly in maths lessons

General Self-Confidence in Learning

- I learn quickly
- Most things I do, I do well
- I am proud of my achievements at school
- I can do things as well as most people
- If I really try I can do almost anything I want to
- I am confident in my abilities

  I am generally high achieving in my studies

Both scales are scored on a 5-point Likert scale.

Appendix C.5: Opportunity to learn scale

Development and Validation of the Opportunity to Learn (OTL) Instrument

Our aim was to develop and validate an instrument to be completed by Y8 students to assess OTL operationalised in terms of learning time in class on the range of topics representing the mathematics expected to be covered in Y8 lessons.[1] In addition, we wanted an instrument that could be completed relatively quickly by students (thus minimising any additional burdens to schools, and avoiding dangers of test fatigue for students) and that could be delivered online (using a multiple-choice format similar to that used in the PISA OTL survey). Ideally, we wanted an instrument producing a unidimensional OTL scale consisting of between 20 and 25 items.

There were three stages to the validation process: validation interviews with Y8 students and mathematics education experts, followed by two 'pilot' rounds of validation.

The initial survey consisted of 34 items with a Likert scale. The survey was designed (and administered) using Research Electronic Data Capture (REDCap) tools hosted at UCL.[2]2

Cognitive interviews were conducted using 'think aloud' protocols with 8 students.[3] These enabled us to assess whether students understood the format of the items and the instruction to consider time spent on the topic (not to solve the item) as well as whether the items adequately represented the intended topics. Additionally, three mathematics education experts were interviewed to assess the mathematics coverage and whether the items would be easily understood by all Y8 students. As a result of the Stage 1 interviews, 10 items were deleted and 12 items added, resulting in a total of 36 items.

In the initial pilot survey (Autumn 2020), 187 Y8 students from one school completed the online survey. Surveys were completed on-site during mathematics lessons in the school's ICT room with the students' mathematics teacher present for the lesson. On average, students took 17 minutes to complete the survey (S.D. = 3 minutes). 17 students did not complete the whole survey. This first pilot indicated good internal consistency (Cronbach α = .86). Rasch[4] statistics were generally good, although factor analysis suggested the possibility of two factors. The Rasch modelling identified a number of potentially problematic items. As a result, adjustments were made to the wording of eight items, but no items were dropped at this stage.

In the second pilot survey (Winter 2021), 295 Y8 students from 4 schools completed the survey, again administered using REDCap. However, due to the pandemic and school closures, the surveys were completed by students at home. 44 students did not complete the whole survey. Additionally, 31 students completed the survey in less than 5 minutes (which, on the basis of the first pilot timings, was judged to be too fast to have considered OTL for all items) and 5 students took longer than 20 minutes (suggesting that

---

[1] Our expectation is that some classes, particularly lower sets, may cover some aspects of the KS2 curriculum, so the topics covered were slightly broader than the Y8/KS3 national curriculum for mathematics.

[2] REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources (Harris et al., 2009, 2019).

[3] The original intention had been to interview 25 students from 3 schools. However, school closures in Spring 2020, meant that it was only possible to interview students from 1 school.

[4] Rasch modelling is a form of Item Response Theory (IRT) and is commonly used to validate single trait (or uni-dimensional) instruments and tests.

they attempted to solve the items rather than consider OTL.) Hence, responses from 215 students were analysed.

Analysis consisted of factor analysis and Rasch modelling (using the partial credit model).[5] As a result of this analysis, several items were dropped (including problematic items identified in the first pilot) leaving a final survey of 22 items. Rasch statistics for these items are attached in Appendix 1 and are considered satisfactory (e.g., infit and outfit values of all items fall into the acceptable range between 0.4 – 1.6[6]). Internal consistency is good (Cronbach $\alpha = 0.85$). However, factor analysis indicated two factors 'explaining' 25% and 10% of the variance, respectively. Whilst not ideal, we judge this to be satisfactory for the purposes of the analyses that will be conducted using this measure.

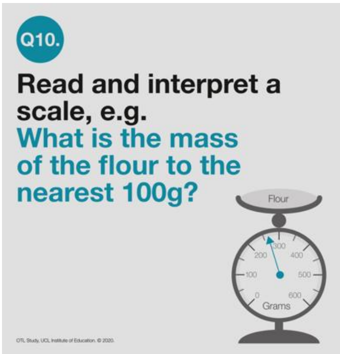
As a final validation, the instrument will be validated early in 2022 with a sample of at least 300 Y8 students from 3-4 schools under the conditions in which the instrument will be delivered (i.e., in students mathematics classes) and with the 22 items presented in random order. This has not been possible due to the impact of the pandemic on schools.

Example Items from the OTL Survey

Q9.

Recall and use division facts, e.g.

56 ÷ 8 =___



How often have you encountered these types of problems in your mathematics lessons?

- ◯ Frequently
- ◯ Sometimes
- ◯ Rarely
- ◯ Never

[5] The partial credit model is a rating scale model and, thus, appropriate to modelling a set of Likert scale items. PCM allows each item to have its own structure.

[6] Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum.

Appendix C.6: Teacher quality – student survey items

Your experiences of learning in maths

Thinking about your maths lessons, how much do you agree with the following statements? If you have more than one maths teacher, please think about the one who teaches you most of the time.

1. I sometimes feel inspired by my teacher's knowledge of maths
2. If I don't understand something in maths my teacher explains it another way
3. My teacher has several good ways to explain each topic that we cover in maths
4. When I need to correct something specific my maths teacher always has an activity to help
5. Before teaching a new topic or idea my maths teacher checks that all pupils understand and remember the previous work that it builds on
6. We enjoy listening to my maths teacher
7. My maths teacher regularly shows me how I can improve
8. We get helpful comments to let us know what we did wrong in our maths work
9. The comments that we get on our work in maths help us to understand how to improve
10. The feedback we get from our maths teacher tells us what to do to get better
11. We understand the feedback we get from our maths teacher
12. In maths we practise and repeat certain things to help us remember them very well
13. My maths teacher shows us how to plan our work

Response options:

- Agree strongly
- Agree
- No opinion
- Disagree
- Disagree strongly

Appendix C.7: Head of Mathematics Survey Questions

About grouping in maths at your school	
Do you have a nurture group in Year 7 for maths?	Yes/No
Briefly, why do you have a nurture group?	[free text]
How many maths classes are there in Year 7?	Options 2-16
How many different ability or attainment level sets are there? This may not be the same as the total number of sets. The following examples may help you choose your answer: If there is 1 top set, 2 middle sets and 1 bottom set, you would answer 3. If there are 2 top sets, 2 upper middle sets, 2 lower middle sets and 2 bottom sets, you would answer 4. If there are 8 sets all at different ability levels you would answer 8.	Options 2-12
Briefly, what is the main reason why you use your current approach to grouping for maths in Year 7?	[free text]
Whose decision is it to group in this way?	Head of Maths Senior Leadership Team Multi-Academy Trust Governors/Trustees By mutual agreement within the maths department Other (please specify)
If you would like to provide any additional information about grouping in maths at your school, please do so here	[free text]
Grouping information	
In some schools, pupils are not allocated to their maths sets straight away at the beginning of the autumn term. When were pupils in your school allocated to their maths groups?	Straight away In September, but after the beginning of term October After October half term Later in the school year
Which sources of information were/will be used to allocate students to Year 7 maths groups this year? Please tick all that apply, but exclude exceptional cases such as students arriving from overseas.	National Curriculum Key Stage 2 test results National Curriculum Key Stage 2 teacher assessments A commercially-available test (e.g. CATS/MidYIS) The school's own test of student attainment Teacher judgements of students' abilities Teacher observations of student behaviour A parent's judgement of their child's ability

	Information from students' feeder schools The results of random allocation of students Other (please specify)
When are commercially-available tests conducted?	Before the start of Year 7 (e.g. 'moving up' day) In the first two weeks of term Later in the first half-term After the autumn half-term holiday
When is the school's own test conducted?	Before the start of Year 7 (e.g. 'moving up' day) In the first two weeks of term Later in the first half-term After the autumn half-term holiday
Briefly, what is the main reason why you use these sources of information for allocating pupils to classes?	[free text]
[Setting schools only] How did you allocate Year 7 teachers to maths classes for this school year 2022-23? Please tick all that apply.	Subject expert teachers tend to be placed with the higher ability classes Subject expert teachers tend to be placed with the lower ability classes Teachers are rotated between ability classes and year groups from year to year Teachers are randomly allocated to different classes I have little choice over how teachers are allocated to Year 7 classes Teachers are allocated to classes based on their personal preferences Teachers are allocated to classes based on their particular strengths Experienced teachers tend to be placed with the higher ability classes Experienced teachers tend to be placed with the lower ability classes Experienced teachers tend to be placed with the middle classes Inexperienced teachers tend to be placed with the higher ability classes Inexperienced teachers tend to be placed with the lower ability classes Inexperienced teachers tend to be placed with the middle classes

	Teachers are allocated first to GCSE/A level classes and later to Year 7 Teachers are allocated so that they have a balanced set of classes (and ages) Other (please specify)
[Setting schools only] Which teacher strengths are taken into consideration when allocating teachers to groups?	Teaching pupils with SEND Teaching pupils with high attainment Managing challenging behaviour Building relationships with vulnerable pupils Subject knowledge Other strengths (please specify)
[Mixed attainment schools only] How did you allocate Year 7 teachers to maths classes for this school year 2022-23? Please tick all that apply.	Teachers are randomly allocated to different classes I have little choice over how teachers are allocated to Year 7 classes Teachers are allocated to classes based on their personal preferences Teachers are allocated to classes based on their particular strengths Teachers are allocated first to GCSE/A level classes and later to Year 7 Other

Appendix C.8: TRU Scoring rubric for case study field visits (an extract)

Date:	School code:		Teacher:	Observers:
Time:	Year Group:	Set/ MA:	No. of pupils:	
Notes for interview (transfer to interview schedule)				
Note: Text on white background refers to whole class and text on grey background refers to small group				

Domain Mathematical Content	Agreed evidence observed (Whole class)	Agreed evidence (Small group)
1 Classroom activities are unfocused or skills--oriented, lacking opportunities for engagement with key grade level content (as specified in the Common Core Standards)		
1 The mathematics discussed is not at grade level; **OR** discussions are aimed at “answer getting.” Explanations, if they appear, are largely procedural.		

2 Activities are at grade level but are primarily skills---oriented, with few opportunities for making connections (e.g., between procedures and concepts) or for mathematical coherence (see glossary). 2 Discussions are at grade level but are primarily skills---oriented, with few opportunities for making connections (e.g., between procedures and concepts) or for mathematical coherence (see glossary).		
3 Classroom activities support meaningful connections between procedures, concepts and contexts (where appropriate) and provide opportunities for building a coherent view of mathematics. 3 Explanation of and justification for central grade level mathematical ideas is coherent.		

Appendix C.9: Student paired interview after the observed lesson

TRU: Mathematics content

W/C 1. How did you find the lesson today?

Probe: Tell me what you liked/disliked about this lesson?

2. What was the lesson about?

Probe: Was the maths similar to any maths in previous maths lessons?

TRU: Cognitive Demand

General 3. What did you learn today that you didn't know about before today?

W/C 4. Which parts in the today's lesson did you find most challenging?

Prompt: think about a task you worked on in the lesson

Probe: Did you manage to work it out? How?
Which did you find the easiest?
Learnt the most from?

5. What were you expected to do to be successful in your maths lesson today?

Prompts: give correct answers, solve problems using the steps the teacher showed me, explain my thinking, listen and make sense of other students' reasoning

Probe: Is that different to other maths lessons?
What would other students in your class say? Which ones? How do you know?

S/G [If no paired/group work, ask if they ever do it. If yes, ask how they like it, then ask 7 and 8 only and skip remaining S/G questions]

6. What was the purpose of your small group work today?

Prompts: solve a problem using the steps our teacher showed us, check with my group members to see if my answers were correct, investigate a mathematical idea, share the different ways students in my group were solving a problem

Probe: Is that what small group work in your maths lessons is usually for?

7. Do all groups usually work on the same/different task/activity, like today?

Prompt: we usually all have the same task, the teacher gives some groups different tasks, we choose the task we want to work on

8. Do you usually work with the same students for group work?

Prompt: the teacher tells us who to work with, we always sit in the same groups, and why do you think that is

9. How useful did you find other students' ideas for your learning during small group/paired work today? And other students, how useful do you think your ideas were for their learning?

10. Did your group struggle with the task today? Y: What did you do help yourselves?

Prompt: ask the teacher, ask someone in a different group, listen to each other's ideas, try out different ways and compare them, wait for the teacher to come

Probe: Is that what normally happens?

TRU: Access

W/C 11. How comfortable were you sharing your ideas in the lesson today?
Probe: Is that usually how you feel in your maths lessons (e.g., comfortable or not sharing)?

Probe: For 'Yes' – Can you give an example of something that you shared in the lesson today?
Prompts: How did the teacher or other students react to what you shared? How was what you shared used by the teacher or other students?

Probe: For 'No' – Can you give an example of something that you shared (or wanted to share?) in the lesson today but felt uncomfortable doing so? What was your reason for feeling uncomfortable?
Prompt: How did you think your teacher/ or other students might react if you shared your thinking?
12. Did listening to other students in today's whole class discussion help make your thinking better?
Prompt (if yes): Can you give an example [thinking about an activity or a problem in today's lesson]?
Prompt (if no): Did anything another student said in the lesson confuse you? Can you give an example [thinking about an activity or a problem in today's lesson]?

S/G 13. Did you participate in your small group today?
Probe: Is that what it's usually like in your group work?

14. What happens in small group/paired work if you disagree with another student in your group?
Prompt: Did this happen today? How did you work through it? Share your answer and your reason? Compare your reason with other students. Check the calculation again using the teacher's way? Say nothing and assume you were wrong? Say nothing and assume you were right? It would depend on who the other student was.

TRU: Agency, ownership and identity

W/C 15. Which students talked the most in the lesson today?
Prompts: students who knew the right answer; who had ideas to share; had questions to ask
Probe: Is that what it's usually like in your maths lessons?
Which students usually share their ideas more / less than others? Ask more questions than others? Students that are good or bad at maths?

And you, when do you talk in maths lessons?

16. What do you think was that the purpose of today's whole class discussion?
Prompts: share how we solved problems using the steps our teacher showed us, learn the way the teacher showed us to solve the problem, learn different ways that work to solve a problem from other students, share a mathematical idea we came up with on our own, check to see if our answers are correct
Probe: What would happen if *you* had a different answer from the teacher or another student?

S/G 17. How comfortable were you sharing your thinking in your small group/with your partner today?

Probe: Is that usually how you feel in your group work?

Probe: Y – How did other students react to what you shared today?

Prompt: Did the students agree with what you shared? Did your group use your idea?
Y – [<u>if the teacher was with the group</u>] How did the teacher react to what you shared?

N – What was your reason for not feeling comfortable to share your thinking in today's group work?
Prompt: How did you think your group might react if you shared your thinking?

18. Who talked the most in your group today?

Prompt: Is it because of the task you worked on or is it not activity-related? Some students talk more than others, students that are good or bad at maths?
Probe: Why do you think those students had more to say?

TRU: Formative assessment

W/C 19. How well do you think you understood the maths in today's lesson?

Prompt: very well, some parts were difficult, hard.

Probe: How did how well you understood the maths make you feel?

20. Do you think your teacher knows if you understood the lesson? How do you know that?
Prompt: he/she checked my work in the lesson, he/she asked me harder questions, he/she asked me to share my work with others.

S/G 21. And when you work in groups/pairs, does the teacher come to your group?
Prompt: why or why not (e.g. the teacher thinks we can do the work without him/her), who does the teacher go to, how the teacher chooses a group he/she is going to work with, do you mind if the teacher doesn't come to you?

W/C 22. What do you think the students in other sets/classes might have been learning in today's maths lesson?
Prompt: same maths, harder/easier maths

Appendix C.10: Teacher post lesson observation interview (mixed attainment)

* Lesson Specific Clarification Questions

1. How does mixed attainment grouping work in your school?

Prompts: which year groups, how many maths classes in each year group

Probe: How long have you been teaching students using mixed attainment grouping practice? (including ITE)

How do you feel about this approach to grouping students?

TRU: Mathematics

2. What were your mathematical goals for this lesson?

Prompts: New learning? Key mathematical ideas? Concept? Procedure? Prior learning? Assessment?

Probe: Were the goals the same for all students in the class?

Prompts: Did all students achieve the mathematical goal for your lesson today (e.g. the focal students in particularly?) If so, how do you know? If not, why do you think that?

What mathematical connections do you think students made to prior learning?

What mathematical goals for future learning, will today's learning lead on to?

TRU: Cognitive Demand

3. How challenging did you imagine the students might find the mathematics content in the lesson?

Prompt: Misconceptions, common errors

Probe: What kinds of mathematical thinking did you want your students to experience in today's lesson?

Prompt: remember facts, vocabulary, use or apply strategies, make connections, develop partial understandings, draw on prior knowledge, interpret/use

Probe: How well did you feel the students managed the level of challenge in the lesson today?

Prompts: Would you say that is the same for all students, with LPA and HPA, e.g. focal students?

How did students manage any challenges they had? (e.g. asking the teacher a question? Asking/questioning another student?)

How do you see your role in managing these challenges?

TRU: **Access**

4. In what ways did you ensure that all students in the class today accessed the mathematics?

Prompt: ways did you 'differentiate'

To what extent are the low attaining students considered in your lesson?

The lesson required the students to... talk, write, lean in, listen hard, manipulate symbols, make diagrams, interpret text, use manipulatives, connect different ideas, etc

Teacher: used prompting questions, scaffolding questions, giving clues, redirecting student attention to a critical feature?

Tasks: asking open ended/single answer, offering low threshold-high ceiling/narrow focus, providing opportunity for multiple solution strategies/ routine practice of worked example

Probe: Were there any students who participated more or less than others in the lesson? Why do you think that was?

Prompt: What can you tell me about the students who contributed most/less in the lesson today? Specifically, LPA/ HPA students (e.g. our focal students maybe)? Typical patterns of participation?

TRU: Agency, ownership and identity

5. In what ways did you anticipate how different groups of students might contribute during the lesson?

Prompt: Were there parts of your lesson where you had considered how to involve LPA/ HPA?

Probe: Which students contributed to the whole class lesson today?

Prompt: Lower PA/ Higher PA

Probe: Were there **any particular things** that one or more students contributed today that influenced what happed next in the lesson?

Prompt: Were these Lower PA/Higher PA students? How typical are these patterns of interaction for these students?

(Or draw attention to an example observed in the lesson)

TRU: Formative assessment

6. What did you do differently from what you had planned to do today? Why?

Prompt: student responses? Timings? Questions you asked? Student difficulties?

Probe: What did you learn in today's lesson about your students' understanding of the mathematics in this lesson?

Prompt: how LPAs different from HPAs

Prompt: What did students say or do that revealed what they understood about the mathematical ideas in today's lesson?

7. Considering the lessons today, can you imagine in what ways it would be different if you were using setting?

Prompts: in terms of: a) your students' experiences of the mathematics (Specifically for LPA / HPA); b) your teaching practices (e.g. with LPA/ HPA)

8. How many of the Y7 classes do you teach? (out of how many classes?)

Probe: Do you teach other year groups?

Probe: Have you every taught this mathematical content to a different class grouping? How was it the same/ different?

Prompts:

Student experiences: less/ more opportunity … to discuss ideas with other students/ teacher, to work independently/ in a group, to use multiple representations, to solve authentic problems

Teacher practice: different content, seating arrangements, timing, tasks/exercises, resources, representations, cognitive demand, pace

9. Finally, I'd like to ask about your background as a teacher:

- How many academic years have you taught mathematics since you qualified?
- What's your highest mathematics qualification?
- What was your route to achieve QTS?

Appendix C.11: Focus group schedule

Sorting Task 1 Mathematical Content

Sorting statements

Introducing the card sort

- We would like to know more about learning maths in your Y8 lessons and how your Y8 teacher helps you. We have some statements that we'd like you to look at together as a group and discuss where you think you would place the cards under each of the three headings: Most of the time, some of the time, rarely/never
- We'd like you to say out loud to one another your reasons for your choices. If you can't agree, that is fine (in fact it is good). Just try to explain why *you* think the card goes where you would place it and give some examples to back up your suggestion and we'll work out where the card could go.
- Do you have any questions before we begin?

- But before we do the sorting task, we'd like to ask you - what makes maths enjoyable or less enjoyable for you. [Allow students to respond individually].

[*Prompt*: when it's easy/ hard, getting it right, making you think, eventually working out the answer, working with friends]

- OK, we'll take a few cards at a time now and I will ask you a few other questions to help you think about the statements on the cards.

Notes for interviewer:

1. Make it clear for the recording where they are positioned e.g. "So you would put 1.1 with rarely/never?".
2. Describe the theme for the cards as you lay them out. E.g. for 1. Mathematical facts and procedures – "These cards are about learning mathematical facts and procedures".
3. If you sense a strong opinion about the statement, invite the students to suggest whether they would prefer it to be in a different position and to state which, aloud (for the recording).

1. Mathematical Facts and procedures

Lay **1.1** and **1.2** out. Invite a student to read each one aloud. "So where would you put these? Can you discuss with one another?"

1. Learning maths is about remembering facts (e.g. times tables, formulas, definitions) and methods (e.g. long multiplication, finding areas of a 2D shape.)

2. Learning maths is about copying and using the teacher's method from the board.

Probes

i. What facts/methods do you have to remember? How does your teacher help you to remember these facts and methods?

[*prompt:* times tables? formulas? Definitions?]

ii. Are there any facts or methods that are hard for you to remember? Why do you think that might be?

[*prompt*: practice at the beginning of a lesson, short tests, homework]

iii. What happens when you go wrong with a method or can't remember a fact?

Lay out **1.3** and **1.4**. Read the statements to the students. "So where would you put these? Can you discuss with one another again?"

3. Learning maths is about using facts to find other facts or comparing different methods to solve a problem.
4. Learning maths is about finding your own methods to solve problems.

Probes

iv. Can you give an example of how you might use a fact to solve another fact? [e.g. if you know 8 x 2 = 16, what other problems could this fact help you to solve?]. Does your teacher ever ask you think in that way?
v. Does your teacher ever use students' own methods to help you to learn?

2. Mathematical concepts

Lay out **2.1**. Read aloud to the students. "So where would you put this?"

1. Learning maths is about the teacher *telling us about* mathematical ideas e.g. telling us what ratio is.

Probes

vi. Can you give an example of some maths that you feel you understand well. So, what makes you think you understand it well?

[*prompt*: can you do it without thinking? explain why or how it works? It makes sense to you? know when to use it? do it quickly?]

vii. What does your teacher do to help you understand the maths you are working on in your lessons?

[*prompt:* reminds you of what you've learned before, gives examples of when you might use the maths, explains things clearly, prompts you to say what you already know? uses students' mistakes]

3. Mathematical language

Lay out **3.1** and **3.2**. Read aloud to the students. “So where would you put these?”

1. Learning maths is about using correct mathematical words.
2. Learning maths is about explaining your ideas in your own words.

Probes

viii. What mathematical words have you been using recently to help you learn maths?
ix. Can you give some examples of when you’ve been asked to explain your ideas to the class?

[prompt: What happens after you’ve explained your ideas?]

x. Do you think there are some students who get chosen more often than others to explain their ideas? Why do you think that is?

Lay out **4.1** and **4.2**. (Mathematical representations). Read aloud to the students. “So where would you put these?”

4. Mathematical representation

1. Learning maths is about reading and writing equations and using mathematical symbols.
2. Learning maths is about representing our ideas in different ways. E.g. drawing a bar model, using cubes.

Probes

xi. Can you give some examples of how you use different representations to help you learn the maths in your lesson?

[*prompt*: drawing something? Moving something around? Using technology – e.g., phone, ipad, computer app]

xii. Does your teacher help to link the maths you are learning to real life? Can you give some examples?

5. Mathematics as a discipline

Lay out **6.1** and **6.2**. Read aloud to the students. “So where would you put these?”

1. Maths is about getting the right answer.
2. Mathematics is about looking for and using patterns and rules.

Probes

xiii. Do you ever get asked to explain why an answer is correct? Can you give an example?
xiv. What do you think it means to be good at maths?

[*prompt*: you? think of someone in your class]

Sorting Task 2 Cognitive demand

Sorting statements

Preparing for the card sort

- This time we would like to know more about your experiences of struggle in your Y8 maths lessons. Like before, we have some statements that we'd like you to look at together.

- But before we do the sorting task we'd like to ask you what you understand by the term "struggle". [Allow students to respond].

[*Prompt*: being stuck, making you think hard, feeling challenged, you are learning/ not learning]

Probes

I. Do you think it is a good or a bad thing to struggle in your maths lessons? Why do you think that? [*prompt*: been given right/wrong level of work]

II. What kinds of things do you struggle with in maths?

[*prompt*: remembering things? Making sense of a problem? Understanding the teacher/ the students? Keeping up with others?]

III. Which students struggle in your maths class?

[*prompt*: students who are better at maths/ not as good at maths] How do you know when they are struggling?

Beginning the sorting task

- Here are some cards about struggle, we'd like you to discuss where you think you would place the cards under each of these headings.
- We'll take a few cards at a time and sometimes I will ask you a few other questions to help you think about the statements on the card.

Notes for interviewer:

1. Make it clear for the recording where they are positioned e.g. "So you would put 1.1 with rarely/never?".
2. Describe the theme for the cards as you lay them out. E.g. for 1. Mathematical facts and procedures – "These cards are about learning mathematical facts and procedures".
3. If you sense a strong opinion about the statement, invite the students to suggest whether they would prefer it to be in a different position and to state which aloud (for the recording).

Most of the time, some of the time, rarely/never

6. Teachers managing struggle

Lay **6.1** and **6.2** out. Read each one aloud. “So where would you put these? Can you discuss with one another?”

1. When we struggle our teacher asks us to explain what we are thinking.
2. The teacher shares things we are struggling with, with the whole class.

Probes

I. Do you think your teacher wants all students to struggle in your maths lessons? Why?

[*prompt*: no, a little bit, all students need to]

II. How does your teacher respond when someone struggles in your lesson?

[*prompt*; the teacher … reassures/ encourages you, gives you a hint, repeats instructions, uses your idea with the class, asks other students to help you, gives different work]

III. How does your teacher know when students are struggling or when they find the work easy?

7. Students managing struggle

Lay **7.1** and **7.2** out. Read each one aloud. “So where would you put these? Can you discuss with one another?”

1. When we are struggling, we put up our hand for our teacher to come to us.
2. We share our answers even if we aren’t sure we are correct.

Probes

IV. Can you give me an example of when you have been stuck on something in maths. How did you try and get ‘unstuck’? Was that alone or were you working with others?

[*prompt*: hands up? Anything before hands up? Look back at book, check board, discuss with others around me; use resources/ draw something?]

V. Does explaining your ideas or listening to others’ ideas help you to makes sense of the maths in your class? Why/how?

Appendix C.12: Head of Mathematics interview

* Notes from HOM survey/ Observations from lesson – specific queries

*

1. I'd first like to ask you some background information

   - How many academic years have you been a HOM?

   - In this school?

   - Others?

   - Did they also use similar grouping practices?

   - How many academic years have you been a teacher of mathematics?

   - Do you/ have you had any other roles in the school?

   - What is your highest mathematics qualification? (*GCSE, A level, mathematical degree?*)

   - What was your route to achieving QTS? (*PGCE, GPT, TF other?*)

   - Which year groups are you teaching maths this year? (*sets? Nurture group?*)

About the type of grouping practice

2. Tell me about how you group students in Y7 and Y8 for mathematics in your school.

   Prompts: How many sets/ classes? Nurture groups? Who decides who goes into which set/class?

   Probe: Are there any other provisions for students in Y7/8 for different groups of students (e.g. Lower PA / higher PA?)

   Prompts: homework club, interventions, enrichment activities?

3. What are the reasons for using this approach to grouping your students in Y7 and 8?

   Prompt: Similar grouping practices in older year? Why? Why not? Certain student characteristics of one/all cohorts? Who decided to group students this way?

   Probe: How well does setting /mixed attainment teaching work?

   Prompt: What are the benefits? Challenges?

4. How do you feel about leading a department that uses this approach to grouping students?

Probe: Do you think this is the best way to group students? Why?

Prompt: Have you ever considered grouping students in other ways? Have you worked in schools that have grouped students in a different way?

Reflecting on the values and beliefs underpinning the grouping approach

TRU Equitable access

5. How do you support ***all learners*** in Y7, so that they have an opportunity to make progress in mathematics in your school?

Prompt: what about for those students with lower prior attainment? Higher prior attainment? Teacher allocation, student-to-class allocation?

Probe: How do you know whether your (LPA/ HPA) students' progress benefit from your grouping approach?

Prompt: Have there been any particular challenges to creating equitable access to mathematical learning? For LPAs or HPAs?

Probe: (Setting only) What would your advice be about doing setting as equitably as possible?

TRU: Agency, ownership and identity

6. From your perspective as a HOM overseeing the department, what do teachers do to enable all students to *participate* in learning mathematics in Y7/8?

Prompt: work in groups, discuss, explain, listen to teacher, work in silence, volunteer answers

Probe: (Setting) How would student agency compare in different sets?

Prompt: lower/higher sets; agency is where students have a voice in their classroom

Probe: So, how do you think your grouping approach promotes student agency in their mathematics learning?

7. To what extent would Y7 students in your school see themselves as maths people?

Prompt: (MA) Would Lower/higher attainers see themselves in the same way or (setting) in higher or lower sets

Probe: What do think for? your students might mean to be good at maths?

TRU: Mathematical content

8. Tell us about the origins of your scheme of work?

Prompt: SoW written by…school? MAT? NCETM? White Rose? Purchased from… so would you say you use a "mastery" approach?

Probe: (MA) So how do teachers use this SoW to set learning objectives that account for different prior attainment in the class?

Prompt: Lower prior attainers / higher prior attainers?

Probe: (Setting) So how do teachers use this SOW to set learning objectives for different sets?

Prompt: experience the same/different mathematical content? If no, how do you decide what content is included/omitted?

TRU: Cognitive Demand

9. How is challenge included in your scheme of work?

Prompt: different types of tasks – open ended, reasoning, different contexts

Probe: What would you expect *challenge* to look like in mathematics lessons in Y7/8?

Probe: Would challenge look different in different sets? / for students with different prior attainment?

TRU: Formative assessment

10. How do you expect teachers in Y7/8 to use formative assessment in the classroom?

Prompt: Clear learning objectives for lessons? Indicators of success/ difficulty? Assessment policy? Marking policy? Feeding back to students?

Probe: How might teachers in Y7/8 in your school monitor the learning of students with differing prior attainment in their lessons?

Probe: How well do you feel teachers' use of formative assessment impacts on LPA students and HPA students in Y7/8? / students in different sets?

Reflecting on the quality of teaching for different groups of students

11. What do you think are important qualities for teaching mathematics?

Prompting: teachers' subject knowledge, pedagogical knowledge (how to teach), pedagogical content knowledge (content knowledge for teaching); knowledge about learners; school T&L policy?

Probe: Are there different qualities needed for teaching different students in different grouping arrangements? mixed attainment classes/top set/bottom set?

Prompt: Teacher allocation – revisit if relevant – allocating teachers with different skills/qualities to particular class/sets

Probe: (MA only) I guess you'll have a range of teachers with different expertise in Y7 and Y8. How are they supported to teach mixed attainment mathematics in your school?

12. (MA) What advice would you give to a HOM considering adopting mixed attainment grouping in a school that currently sets?

Appendix C.13: Teacher endline survey: full tables for IPE items

Teacher endline survey: About students learning maths. Categories collapsed: SD/D – strongly disagree and disagree; N – neither agree nor disagree; A/SA – agree and strongly agree.

	Mixed attainment						Setting					
	SD/D		N		A/SA		SD/D		N		A/SA	
	N	%	N	%	N	%	N	%	N	%	N	%
All students should have the opportunity to learn the whole maths curriculum, regardless of prior attainment.	33	16.9	5	2.6	157	80.5	119	29.0	20	4.9	271	66.1
All students benefit from the same types of pedagogy when learning maths, regardless of prior attainment.	72	36.9	21	10.8	102	52.3	187	45.6	46	11.2	177	43.2
Students benefit from learning in maths classrooms with students of a similar prior attainment level	41	21.0	51	26.2	103	52.8	13	3.0	21	5.1	376	91.7
Students benefit from learning in maths classrooms with a wide range of prior attainment in maths	43	22.1	36	18.5	116	59.5	247	60.2	75	18.3	88	21.5
All students can do well in maths	19	9.7	14	7.2	162	83.1	69	16.8	39	9.5	302	73.7
All students benefit from being challenged mathematically, regardless of their prior attainment	9	4.6	5	2.6	181	92.8	22	5.4	17	4.1	371	90.5
All students benefit from support when learning mathematics, regardless of their prior attainment	9	4.6	10	5.1	176	90.3	9	2.2	12	2.9	38.9	94.9

Teacher endline survey: Thinking about your teaching of mathematics to Year 7 and Year 8, how often do you do the following? Categories collapsed: N/O – never and occasionally; F/A – frequently and always.

	Mixed				Setting			
	N/O		F/A		N/O		F/A	
	N	%	N	%	N	%	N	%
I present a summary of recently learned content	72	37.3	121	62.7	126	32.1	267	67.9
I set goals at the beginning of instruction	66	34.2	127	65.8	105	26.7	288	73.3
I explain what I expect the students to learn	44	23.0	147	77.0	35	8.9	358	91.1
I explain how new and old topics are related	39	20.2	154	79.8	48	12.2	344	87.8
I present tasks for which there is no obvious solution	104	53.9	89	46.1	296	74.7	100	25.3
I give tasks that require students to think critically	48	24.9	145	75.1	140	35.5	254	64.5
I have students work in small groups to come up with a joint solution to a problem or task	130	67.4	63	32.6	326	82.3	70	17.7
I ask students to decide on their own procedures for solving complex tasks	89	46.6	102	53.4	231	58.3	165	41.7
I tell students to follow classroom rules	9	4.7	183	95.3	25	6.3	371	93.7
I tell students to listen to what I say	13	6.8	179	93.2	18	4.6	377	95.4
I calm students who are disruptive	10	5.2	182	94.8	27	6.9	367	93.1
When the lesson begins, I tell students to quieten down quickly	12	6.3	179	93.7	33	8.3	363	91.7
I refer to a problem from everyday life or work to demonstrate why new knowledge is useful	56	29.3	135	70.7	133	33.7	262	66.3
I let students practise similar tasks until I know that every student has understood the subject matter	48	25.0	144	75.0	75	19.0	320	81.0
I have students come up to the board to show their ideas	114	59.1	79	40.9	280	70.7	116	29.3
I give students manipulatives to use in lessons to help them learn	137	71.4	55	28.6	326	82.3	70	17.7
I use representations in lessons to help students learn	34	17.7	158	82.3	121	30.9	271	69.1
I observe students when working on particular tasks and provide immediate feedback	17	8.8	176	91.2	20	5.1	376	94.9
I seat students who need the most help where I can easily get to them	10	5.2	183	94.8	17	4.3	38	95.7
I give students problem solving tasks as an extension	42	21.9	150	78.1	107	27.0	289	73.0
I give students problem solving tasks as a starting point for learning	118	61.1	75	38.9	310	78.3	86	21.7

Experiences of grouping

Which of the following types of grouping for maths have you experienced as a teacher? Please select all that apply.

	Mixed		Setting	
	N	%	N	%
Mixed attainment	194	99.0	217	52.3
Setting	159	81.1	397	95.7
Streaming	78	39.8	107	25.8
Nurture group	106	54.1	106	25.5
Total in group	196	100	415	100

Beliefs about students learning maths

All students should have the opportunity to learn the whole maths curriculum, regardless of prior attainment.

	Mixed		Setting	
	N	%	N	%
Strongly disagree	5	2.6	27	6.6
Disagree	28	14.4	92	22.4
Neither agree nor disagree	5	2.6	20	4.9
Agree	73	37.4	137	33.4
Strongly agree	84	43.1	134	32.7

All students benefit from the same types of pedagogy when learning maths, regardless of prior attainment.

	Mixed		Setting	
	N	%	N	%
Strongly disagree	23	11.8	53	12.9
Disagree	49	25.1	134	32.7
Neither agree nor disagree	21	10.8	46	11.2
Agree	54	27.7	105	25.6
Strongly agree	48	24.6	72	17.6

Students benefit from learning in maths classrooms with students of a similar prior attainment level

	Mixed		Setting	
	N	%	N	%
Strongly disagree	12	6.2	4	1.0
Disagree	29	14.9	9	2.2
Neither agree nor disagree	51	26.2	21	5.1
Agree	67	34.4	156	38.0
Strongly agree	36	18.5	220	23.7

Students benefit from learning in maths classrooms with a wide range of prior attainment in maths

	Mixed		Setting	
	N	%	N	%
Strongly disagree	10	5.1	96	23.4
Disagree	33	16.9	151	36.8
Neither agree nor disagree	36	18.5	75	18.3
Agree	86	44.1	66	16.1
Strongly agree	30	15.4	22	5.4

All students can do well in maths

	Mixed		Setting	
	N	%	N	%
Strongly disagree	8	4.1	13	3.2
Disagree	11	5.6	56	13.7
Neither agree nor disagree	14	7.2	39	9.5
Agree	79	40.5	155	37.8
Strongly agree	83	42.6	147	35.9

All students benefit from being challenged mathematically, regardless of their prior attainment

	Mixed		Setting	
	N	%	N	%
Strongly disagree	5	2.6	5	1.2
Disagree	4	2.1	17	4.1
Neither agree nor disagree	5	2.6	17	4.1
Agree	41	21.0	108	26.3
Strongly agree	140	71.8	263	64.1

All students benefit from support when learning mathematics, regardless of their prior attainment

	Mixed		Setting	
	N	%	N	%
Strongly disagree	5	2.6	4	1.0
Disagree	4	2.1	5	1.2
Neither agree nor disagree	10	5.1	12	2.9
Agree	28	14.4	92	22.4
Strongly agree	148	75.9	297	72.4

Appendix C.14: Student endline survey: full tables for IPE items

Student Grouping Study items:

- I feel comfortable talking about my mistakes in class
- I understand ideas in maths better when I can use a diagram, picture, number line or graph
- I understand ideas in maths better when I can use objects to help me, e.g. cubes
- My teacher uses diagrams to show how to correct a mistake
- My teacher links what we are learning to real-life situations
- My teacher asks students to come up to the board to show their ideas
- My teacher often helps me individually with my work

How much do you agree with the following statements about maths lessons? Categories collapsed: SA/A – strongly agree and agree; D/SD – disagree and strongly disagree.

	Mixed attainment				Setting			
	A/SA		SD/D		A/SA		SD/D	
	N	%	N	%	N	%	N	%
I feel comfortable talking about my mistakes in class	2218	50.4	2184	49.6	5033	51.6	4712	48.4
I understand ideas in maths better when I can use a diagram, picture, number line or graph	3002	68.2	1399	31.8	6890	70.8	2844	29.2
I understand ideas in maths better when I can use objects to help me, e.g. cubes	2567	58.4	1827	41.6	5834	59.9	3898	40.1
My teacher uses diagrams to show how to correct a mistake	2797	63.6	1604	36.4	6135	63.0	3603	37.0
My teacher links what we are learning to real-life situations	2659	60.6	1729	39.4	6162	63.4	3553	36.6
My teacher asks students to come up to the board to show their ideas	2496	56.6	1912	43.4	4957	51.0	4771	49.0
My teacher often helps me individually with my work	2664	60.7	1725	39.3	6162	63.3	3570	36.7

How much do you agree with the following statements about maths lessons? Comparing responses for pupils in mixed attainment and in setting schools, with high, middle and low prior attainment. Categories collapsed: SA/A – strongly agree and agree; D/SD – disagree and strongly disagree.

		High				Middle				Low			
		A/SA		SD/D		A/SA		SD/D		A/SA		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%
	Mixed	836	55.2	678	44.8	770	51.2	735	48.8	547	43.9	699	56.1

I feel comfortable talking about my mistakes in class	Setting	1776	54.3	1493	45.7	1726	51.0	1660	49.0	1320	48.5	1403	51.5
I understand ideas in maths better when I can use a diagram, picture, number line or graph	Mixed	984	65.1	527	34.9	1066	70.9	437	29.1	861	68.9	388	31.1
	Setting	2300	70.5	962	29.5	2438	72.0	947	28.0	1891	69.5	828	30.5
I understand ideas in maths better when I can use objects to help me, e.g. cubes	Mixed	762	50.4	750	49.6	919	61.0	588	39.0	801	64.6	438	35.4
	Setting	1754	53.8	1504	46.2	2085	61.6	1300	38.4	1759	64.6	965	35.4
My teacher uses diagrams to show how to correct a mistake	Mixed	972	64.4	537	35.6	966	64.0	543	36.0	778	62.4	468	37.6
	Setting	2019	62.0	1240	38.0	2156	63.6	1234	36.4	1726	63.4	995	36.6
My teacher links what we are learning to real-life situations	Mixed	949	63.0	557	37.0	908	60.3	598	39.7	726	58.5	515	41.5
	Setting	2217	68.0	1041	32.0	2091	61.9	1285	38.1	1613	59.5	1100	40.5
My teacher asks students to come up to the board to show their ideas	Mixed	292	52.7%	262	47.3%	490	54.7%	405	45.3%	612	52.5%	554	47.5%
	Setting	2273	53.9%	1947	46.1%	2037	51.0%	1961	49.0%	1471	52.5%	1330	47.5%
My teacher often helps me individually with my work	Mixed	896	59.4	613	40.6	898	60.0	599	40.0	784	62.9	462	37.1
	Setting	2029	62.2	1235	37.8	2095	61.9	1290	38.1	1784	65.7	932	34.3

SGS items by FSM

Table A.14.1 I feel comfortable talking about my mistakes in class

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	713	54.1	606	45.9	616	50.5	605	49.5	359	42.1	493	57.9
	Setting	1597	54.9	1312	45.1	1418	51.1	1358	48.9	966	49.5	984	50.5
FSM	Mixed	123	63.1	72	36.9	154	54.2	130	45.8	188	47.7	206	52.3
	Setting	179	49.7	181	50.3	308	50.5	302	49.5	354	45.8	419	54.2

Table A.14.2 I understand ideas in maths better when I can use a diagram, picture, number line or graph

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	857	65.1	459	34.9	868	71.3	350	28.7	596	69.7	259	30.3
	Setting	2052	70.7	851	29.3	2017	72.7	759	27.3	1370	70.2	582	29.8
FSM	Mixed	127	65.1	68	34.9	198	69.5	87	30.5	265	67.3	129	32.7
	Setting	248	69.1	111	30.9	421	69.1	188	30.9	521	67.9	246	32.1

Table A.14.3 I understand ideas in maths better when I can use objects to help me, e.g. cubes

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	650	49.4	666	50.6	755	61.7	468	38.3	559	65.9	289	34.1
	Setting	1562	53.9	1338	46.1	1713	61.8	1059	38.2	1274	65.3	678	34.7
FSM	Mixed	112	57.1	84	42.9	164	57.7	120	42.3	242	61.9	149	38.1
	Setting	192	53.6	166	46.4	372	60.7	241	39.3	485	62.8	287	37.2

Table A.14.4 My teacher uses diagrams to show how to correct a mistake

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	838	63.7	477	36.3	783	63.9	442	36.1	524	61.6	327	38.4
	Setting	1778	61.3	1123	38.7	1750	63.0	1028	37.0	1242	63.7	708	36.3
FSM	Mixed	134	69.1	60	30.9	183	64.4	101	35.6	254	64.3	141	35.7
	Setting	241	67.3	117	32.7	406	66.3	206	33.7	484	62.8	287	37.2

Table A.14.5 My teacher links what we are learning to real-life situations

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	831	63.4	480	36.6	735	60.2	486	39.8	511	60.1	339	39.9
	Setting	1978	68.3	920	31.7	1709	61.8	1058	38.2	1161	59.8	782	40.2
FSM	Mixed	118	60.5	77	39.5	173	60.7	112	39.3	215	55.0	176	45.0
	Setting	239	66.4	121	33.6	382	62.7	227	37.3	452	58.7	318	41.3

Table A.14.6 My teacher asks students to come up to the board to show their ideas

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	776	58.8%	543	41.2%	658	53.8%	566	46.2%	470	54.8%	387	45.2%
	Setting	1497	51.6%	1404	48.4%	1379	49.7%	1395	50.3%	1001	51.5%	943	48.5%
FSM	Mixed	107	54.9%	88	45.1%	181	63.7%	103	36.3%	231	58.6%	163	41.4%
	Setting	185	51.5%	174	48.5%	309	50.6%	302	49.4%	381	49.4%	391	50.6%

Table A.14.7 My teacher often helps me individually with my work

FSM	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	128	66.0	66	34.0	176	62.0	108	38.0	254	64.8	138	35.2
	Setting	210	58.5	149	41.5	381	62.5	229	37.5	513	66.5	259	33.5
FSM	Mixed	768	58.4	547	41.6	722	59.5	491	40.5	530	62.1	324	37.9
	Setting	1819	62.6	1086	37.4	1714	61.8	1061	38.2	1271	65.4	673	34.6

SGS items by sex

Table A.14.8 I feel comfortable talking about my mistakes in class

Sex	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	552	66.3	280	33.7	462	61.9	284	38.1	318	53.9	272	46.1
	Setting	1127	63.9	636	36.1	983	62.7	586	37.3	682	59.3	468	40.7
Girl	Mixed	284	41.6	398	58.4	308	40.6	451	59.4	229	34.9	427	65.1
	Setting	649	43.1	857	56.9	743	40.9	1074	59.1	638	40.6	935	59.4

Table A.14.9 I understand ideas in maths better when I can use a diagram, picture, number line or graph

Sex	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%

Boy	Mixed	537	64.7	293	35.3	533	71.4	214	28.6	414	69.8	179	30.2
	Setting	1238	70.5	519	29.5	1145	73.1	422	26.9	812	70.7	337	29.3
Girl	Mixed	447	65.6	234	34.4	533	70.5	223	29.5	447	68.1	209	31.9
	Setting	1062	70.6	443	29.4	1293	71.1	525	28.9	1079	68.7	491	31.3

Table A.14.10 I understand ideas in maths better when I can use objects to help me, e.g. cubes

Sex	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	444	53.4	387	46.6	465	62.2	282	37.8	385	65.3	205	34.7
	Setting	990	56.3	768	43.7	986	63.0	578	37.0	768	66.4	388	33.6
Girl	Mixed	318	46.7	363	53.3	454	59.7	306	40.3	416	64.1	233	35.9
	Setting	764	50.9	736	49.1	1099	60.4	722	39.6	991	63.2	577	36.8

Table A.14.11 My teacher uses diagrams to show how to correct a mistake

Sex	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	582	70.0	249	30.0	516	68.9	233	31.1	404	68.2	188	31.8
	Setting	1175	66.7	586	33.3	1089	69.4	480	30.6	777	67.3	377	32.7
Girl	Mixed	390	57.5	288	42.5	450	59.2	310	40.8	374	57.2	280	42.8
	Setting	844	56.3	654	43.7	1067	58.6	754	41.4	949	60.6	618	39.4

Table A.14.12 My teacher links what we are learning to real-life situations

Sex	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	534	64.3	297	35.7	472	63.4	273	36.6	344	58.5	244	41.5
	Setting	1214	69.2	541	30.8	1035	66.3	527	33.7	707	61.5	442	38.5
Girl	Mixed	415	61.5	260	38.5	436	57.3	325	42.7	382	58.5	271	41.5
	Setting	1003	66.7	500	33.3	1056	58.2	758	41.8	906	57.9	658	42.1

Table A.14.13 My teacher asks students to come up to the board to show their ideas

Sex	Group	High	Middle	Low

		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	515	61.8%	319	38.2%	422	56.6%	324	43.4%	354	59.6%	240	40.4%
	Setting	922	54.6%	766	45.4%	872	55.7%	694	44.3%	645	56.1%	505	43.9%
Girl	Mixed	368	54.1%	312	45.9%	417	54.7%	345	45.3%	347	52.8%	310	47.2%
	Setting	690	45.9%	812	54.1%	816	44.9%	1003	55.1%	737	47.1%	829	52.9%

Table A.14.14 My teacher often helps me individually with my work

Sex	Group	High				Middle				Low			
		A/F		S/N		A/F		S/N		A/F		S/N	
		N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	530	63.7	302	36.3	460	62.0	282	38.0	388	65.7	203	34.3
	Setting	1163	66.0	599	34.0	1064	67.9	502	32.1	779	67.7	371	32.3
Girl	Mixed	366	54.1	311	45.9	438	58.0	317	42.0	396	60.5	259	39.5
	Setting	866	57.7	636	42.3	1031	56.7	788	43.3	1006	64.2	561	35.8

PISA items

- Mathematics is an important subject for me because I need it for what I want to study later on.
- I will learn many things in mathematics that will help me get a job
- My parents believe that mathematics is important for my career
- My parents like mathematics.

How much do you agree with the following statements about maths? Items taken from PISA student questionnaire (OECD, 2013). Categories collapsed: SA/A – strongly agree and agree; N – neither agree nor disagree; D/SD – disagree and strongly disagree.

	Mixed attainment						Setting					
	SA/A		N		D/SD		SA/A		N		D/SD	
	N	%	N	%	N	%	N	%	N	%	N	%
Mathematics is an important subject for me because I need it for what I want to study later on.	2740	62.3	1064	24.2	597	13.6	6253	64.3	2254	23.2	1222	12.6
I will learn many things in mathematics that will help me get a job	2779	63.1	1055	24.0	569	12.9	6286	64.6	2218	22.8	1220	12.5

My parents believe that mathematics is important for my career	3053	69.4	1004	22.8	340	7.7	6760	69.5	2167	22.3	805	8.3
My parents like mathematics.	2031	46.3	1453	33.1	907	20.7	4546	46.8	3149	32.4	2010	20.7

How much do you agree with the following statements about maths? Items taken from PISA student questionnaire (OECD, 2013). Comparing responses from students in mixed attainment and setting schools, high, middle and low prior attainment. Categories collapsed: SA/A – strongly agree and agree; N – neither agree nor disagree; D/SD – disagree and strongly disagree.

Item	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Mathematics is an important subject for me because I need it for what I want to study later on.	Mixed	1061	70.2	289	19.1	161	10.7	906	60.2	392	26.0	207	13.8	687	55.0	346	27.7	215	17.2
	Setting	2269	69.7	632	19.4	353	10.8	2141	63.2	814	24.0	432	12.8	1585	58.3	728	26.8	406	14.9
I will learn many things in mathematics that will help me get a job	Mixed	1075	71.0	281	18.6	158	10.4	933	62.0	380	25.3	191	12.7	696	55.8	349	28.0	203	16.3
	Setting	2274	69.8	617	18.9	366	11.2	2137	63.1	827	24.4	425	12.5	1618	59.7	697	25.7	394	14.5
My parents believe that mathematics is important for my career	Mixed	1160	79.0	247	16.8	62	4.2	1043	69.1	356	23.6	111	7.4	758	61.0	340	27.4	145	11.7
	Setting	2483	76.2	592	18.2	184	5.6	2315	68.3	783	23.1	289	8.5	1696	62.4	719	26.5	302	11.1
My parents like mathematics.	Mixed	824	54.6	462	30.6	223	14.8	688	45.8	516	34.3	299	19.9	449	36.0	436	35.0	361	29.0

PISA items by FSM

Table A.14.15 Mathematics is an important subject for me because I need it for what I want to study later on

FSM	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	930	70.6	251	19.0	137	11.6	725	59.5	319	26.2	175	16.8	477	55.7	230	26.9	149	21.1
	Setting	2045	70.6	553	19.1	300	11.5	1759	63.4	666	24.0	350	14.4	1157	59.3	522	26.8	271	16.1
FSM	Mixed	131	67.9	38	19.7	24	14.2	181	63.3	73	25.5	32	12.6	210	53.6	116	29.6	66	20.2
	Setting	224	62.9	79	22.2	53	17.5	382	62.4	148	24.2	82	15.5	428	55.7	206	26.8	135	21.3

Table A.14.16 I will learn many things in mathematics that will help me get a job

FSM	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	942	71.3	243	18.4	136	11.5	747	61.2	310	25.4	163	15.4	461	54.0	244	28.6	149	21.1
	Setting	2049	70.7	539	18.6	312	12.1	1762	63.4	672	24.2	343	14.1	1170	60.3	494	25.5	277	16.6
FSM	Mixed	133	68.9	38	19.7	22	12.9	186	65.5	70	24.6	28	10.9	235	59.6	105	26.6	54	15.9
	Setting	225	63.0	78	21.8	54	17.8	375	61.3	155	25.3	82	15.5	448	58.3	203	26.4	117	18.0

Table A.14.17 My parents believe that mathematics is important for my career

FSM	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	1009	76.6	247	18.7	62	4.9	856	69.9	282	23.0	86	7.6	525	61.6	231	27.1	96	12.7
	Setting	2229	76.8	525	18.1	148	5.4	1934	69.7	624	22.5	218	8.5	1240	63.8	510	26.2	194	11.1
FSM	Mixed	151	*	*	*	*	*	187	65.4	74	25.9	25	9.6	233	59.6	109	27.9	49	14.3
	Setting	254	71.1	67	18.8	36	11.2	381	62.4	159	26.0	71	13.1	456	59.0	209	27.0	108	16.2

Table A.14.18 My parents like mathematics.

FSM	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
No FSM	Mixed	731	55.5	393	29.9	192	17.1	558	45.8	412	33.8	248	25.6	311	36.5	306	35.9	236	38.2
	Setting	1700	58.7	780	26.9	416	16.8	1294	46.7	906	32.7	568	25.8	708	36.6	746	38.5	483	33.2
FSM	Mixed	93	48.2	69	35.8	31	19.1	130	45.6	104	36.5	51	21.8	138	35.1	130	33.1	125	46.6
	Setting	160	44.8	107	30.0	90	33.7	226	36.9	225	36.8	161	35.7	269	35.0	263	34.2	236	44.4

PISA items by sex

Table A.14.19 Mathematics is an important subject for me because I need it for what I want to study later on

Sex	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	616	73.9	144	17.3	74	9.7	484	64.9	178	23.9	84	12.7	353	22.1	1141	71.6	100	6.7
	Setting	1316	75.0	262	14.9	176	11.2	1055	67.3	344	22.0	168	12.0	727	63.4	266	23.2	154	15.5

Girl	Mixed	445	65.7	145	21.4	87	14.7	422	55.6	214	28.2	123	19.3	334	51.1	205	31.3	115	21.3
	Setting	953	63.5	370	24.7	177	13.4	1086	59.7	470	25.8	264	17.0	858	54.6	462	29.4	252	19.1

Table A.14.20 I will learn many things in mathematics that will help me get a job

Sex	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	598	71.7	158	18.9	78	10.3	466	62.6	190	25.5	89	13.6	353	59.3	154	25.9	88	17.4
	Setting	1269	72.1	284	16.1	206	13.3	1034	65.9	357	22.8	178	12.8	716	62.8	260	22.8	165	16.9
Girl	Mixed	477	70.1	123	18.1	80	13.3	467	61.5	190	25.0	102	15.5	343	52.5	195	29.9	115	21.4
	Setting	1005	67.1	333	22.2	160	12.0	1103	60.6	470	25.8	247	15.7	902	57.5	437	27.9	229	17.1

Table A.14.21 My parents believe that mathematics is important for my career

Sex	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	654	78.6	146	17.5	32	4.0	526	70.3	171	22.9	51	7.3	378	63.9	150	25.3	64	12.1
	Setting	1357	77.1	313	17.8	89	5.3	1116	71.1	341	21.7	112	7.7	735	64.2	298	26.0	112	10.8
Girl	Mixed	506	74.6	134	19.8	38	5.9	517	67.8	185	24.3	60	8.5	380	58.4	190	29.2	81	14.2
	Setting	1126	75.1	279	18.6	95	6.8	1199	66.0	442	24.3	177	10.8	961	61.1	421	26.8	190	13.7

Table A.14.22 My parents like mathematics.

Sex	Group	High						Middle						Low					
		A/SA		N		SD/D		A/SA		N		SD/D		A/SA		N		SD/D	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Boy	Mixed	459	55.2	262	31.5	111	15.4	359	48.2	264	35.4	122	19.6	241	40.6	197	33.2	155	35.4
	Setting	1019	58.1	490	27.9	245	16.2	773	49.4	526	33.6	266	20.5	460	40.5	407	35.9	268	30.9
Girl	Mixed	365	53.9	200	29.5	112	19.8	329	43.4	252	33.2	177	30.5	208	31.9	239	36.6	206	46.1
	Setting	841	56.1	397	26.5	261	21.1	747	41.2	605	33.3	463	34.2	517	32.9	602	38.3	451	40.3

Appendix C.15: R code for impact analysis

Power calculations

```
library(PowerUpR)

# Study plan
# 30 mixed attainment schools in 20 matched groups and aiming to recruit average of 3 setted schools per mixed attainment school
# Calculations assume 3-level MLM: students clustered in schools and schools clustered in matched groups (average 6 schools per group)
# Assumptions: ICC (schools) = 0.15; pre/post-test correlations 0.75 at student-level & 0.38 at school-level
# Standard assumptions for power and alpha levels
mdes.plan <- mdes.bcra3f2(power=.80, alpha=.05, two.tailed=TRUE, rho2=0.15, p=.33, g2=1, r21=0.56, r22=0.14, n=100, J=6, K=20)

# FSM power calculations on the same basis and assuming average 30 FSM students per school
mdes.planFSM <- mdes.bcra3f2(power=.80, alpha=.05, two.tailed=TRUE, rho2=0.15, p=.33, g2=1, r21=0.56, r22=0.14, n=25, J=6, K=20)


# Recruited matched sample as per evaluation report
# 28 MA schools, 69 setted, in 28 groups, and average 3.5 in each group
# Average cohort size 190
mdes.rec <- mdes.bcra3f2(power=.80, alpha=.05, two.tailed=TRUE, rho2=0.15, p=.29, g2=1, r21=0.56, r22=0.14, n=190, J=3.5, K=28)
# FSM power calculations on the same basis and assuming average 40 FSM students per school
mdes.recFSM <- mdes.bcra3f2(power=.80, alpha=.05, two.tailed=TRUE, rho2=0.15, p=.29, g2=1, r21=0.56, r22=0.14, n=40, J=3.5, K=28)

# At analysis
# 27 mixed attainment schools and 62 setted schools in 27 matched groups (ie 3.3 schools per matched group)
# Assumptions as per planned sample except average student number per school = 167 based on recruitment
# Actual per-post test correlations: pupil-level: 0.73; school-level: 0.75;
# Actual ICC: 0.074
mdes.rec <- mdes.bcra3f2(power=.80, alpha=.05, two.tailed=TRUE, rho2=0.074, p=.3, g2=1, r21=0.53, r22=0.56, n=167, J=3.3, K=27)

# FSM power calculations on the same basis and average 32 FSM students per school
# Actual per-post test correlations for FSM: pupil-level: 0.71; school-level: 0.76;
# Actual ICC: 0.061
mdes.recFSM <- mdes.bcra3f2(power=.80, alpha=.05, two.tailed=TRUE, rho2=0.061, p=.3, g2=1, r21=0.58, r22=0.14, n=32, J=3.3, K=27)
```

Impact analysis

Variable name	Variable description
RU_PupilID	random unique pupil ID for NPD matching to be returned
RU_SchoolID	random unique school ID
RU_ClassID	random unique class ID
mixedAttain	Treatment allocation
Classabilitylevel	Mixed/Nurture or set level
setlevel	Fot setted schools, school reported set level
maxsetlevel	For setted schools, school reported number of set levels
hmlsetlevel	high-medium-low set level (1=high, 2 = medium, 3 = low)
N_pupils_byschool_schrep	Number of pupils in Y7 reported by school
N_pupils_byschool	Count of RU_PupilIDs (data rows) per school
N_pupils_byclass_schrep	School-reported number of pupils in maths class
N_pupils_byclass	Count of RU_PupilIDs (data rows) per school-class
P2SAS	Primary Outcome - GL SAS score based on Paper 2
flag_earlyGL	Flag for early GL test
flag_online	Flag for online GL test
rawSSCbase_scale	General self-confidence baseline measure
rawMSCbase_scale	Maths self-confidence baseline measure
rawSSCend_scale	Secondary outcome: General self-confidence endline measure
rawMSCend_scale	Secondary outcome: Maths self-confidence endline measure
meanTQ_scale	Teacher-quality mean score
meanOTL_scale	Opportunity to Learn mean score
academy	School academy status
tpup_2019	School cohort number of pupils
ks2aps_2019	2019 School average KS2 attainment of intake
ks2aps_2018	2018 School average KS2 attainment of intake
ks2aps_2017	2017 School average KS2 attainment of intake
ptpriorlo_2019	School low prior attainment proportion
ptpriorav_2019	School average prior attainment proportion
ptpriorhi_2019	School high prior attainment proportion
ptfsm6cla1a_2019	School FSM proportion
ptealgrp2_2019	School EAL proportion
gender	School intake: boys, girls or mixed
ofsted_rating	Ofsted grade
region	School geographical region
urbanrural	Urban rural classification
IDACI	School composition IDACI

subclass	Block identifier
EVERFSM_6_P_SP R24	Ever FSM variable, 0/1 factor recoding would be good
IDACIScore_19_SP R24	Individual IDACI score
EAL	EAL flag, factor 0/1 - recode?
flag_selfadmin	Numerical 0/1, 1 indicating a school that self-administered final testing
KS2_MATMRK_lo	Numerical 0/1, 1 indicating a pupil in the lower tertile of KS2_MATMRK scores
KS2_MATMRK_hi	Numerical 0/1, 1 indicating a pupil in the upper tertile of KS2_MATMRK scores
KS2_MATMRK_lmh	Factor low/med/hi indicating the tertile of the pupil's KS2_MATMRK score
ur_3cat	Factor of grouped urbal/rural classification of a pupil's school
stratum.mixed	Concatenation of subclass and mixedAttain variables
is_complete_att	Logical indicating whether all variables in primary regression are complete
is_complete_SSC	As above for SSC (general self-confidence) regression
is_complete_MSC	As above for MSC (maths self-confidence) regression
weight_att	weight of the data point for att(ainment) regressions
weight_SSC	weight of the data point for SSC regressions
weight_MSC	weight of the data point for MSC regressions
EVERFSM_loAtt	numeric 0/1, 1 indicating the pupil is 1 in both EVERFSM_6_P_SPR24 and KS2_MATMRK_lo
IDACIx (x=1,2,3,4,5)	numeric 0/1, 1 indicating the pupil is in IDACI quintile x
Ofsted_Outs	numeric 0/1, 1 indicating the pupil's school is Ofsted Outstanding
Ofsted_Good	numeric 0/1, 1 indicating the pupil's school is Ofsted Good
Ofsted_RI	numeric 0/1, 1 indicating the pupil's school is Ofsted Requires Improvement
Average_att	Mean of ks2aps_2017, ks2aps_2018 and ks2aps_2019
Urban	numeric 0/1, 1 indicating the pupil is in an urban setting

```
#Load packages
library(dplyr)
library(ggplot2)
library(readxl)
library(tidyverse)
library(janitor)
library(mgcv)
library(plotly)
library(quantreg)
library(miceadds)
library(sandwich)
library(modelsummary)
library(gtsummary)
library(broom.helpers)
library(jtools)
library(tictoc)
library(mediation)
library(effectsize)
```

```
library(TAM)
library(MatchIt)
library(sensitivitymv)
library(wCorr)
library(tidyr)
library(performance)
library(mice)
library(TH.data)
library(multcomp)

#Data set-up
#Importing and merging
#Import data
sgs_data <- read_excel("SGS_data_for_SRS_unencrypted.xlsx")
npd_data <- read.csv("Spring_Census_2024_KS2.csv")

#Remove duplicates
npd_dup <- npd_data$RU_PupilID[duplicated(npd_data$RU_PupilID)]
npd_data_nodup <- filter(npd_data, RU_PupilID != x [...])
npd_data_nodup$RU_PupilID[duplicated(npd_data_nodup$RU_PupilID)]

#Select variables for analysis from NPD data
npd_data_nodup_tidy <- dplyr::select(npd_data_nodup, RU_PupilID,
PupilMatchingRefAnonymous_SPR24, Sex_SPR24,
                YearOfBirth_SPR24:EVERFSM_6_P_SPR24, LanguageGroupMinor_SPR24,
starts_with("IDACI"),
                starts_with("KS2_MAT"))
npd_data_nodup_tidy <- dplyr::select(npd_data_nodup_tidy, -starts_with("KS2_MATPROG"), -
KS2_MATOUTCOME, -KS2_MATSCORE,
                -KS2_MATTAOUTCOME, -starts_with("KS2_MATEXP"), -
starts_with("KS2_MATHIGH"), -KS2_MATAT)

#Merge on RU_PupilID
project_data <- left_join(sgs_data, npd_data_nodup_tidy, by="RU_PupilID")

#Check no duplicates created
project_data$RU_PupilID[duplicated(project_data$RU_PupilID)]

#Tidying
#Tidy variables for analysis
#specifying categorical variables as factors
project_data$RU_PupilID <- as.factor(project_data$RU_PupilID)
project_data$RU_SchoolID <- as.factor(project_data$RU_SchoolID)
project_data$RU_ClassID <- as.factor(project_data$RU_ClassID)
project_data$mixedAttain <- as.factor(project_data$mixedAttain)
project_data$hmlsetlevel <- as.factor(project_data$hmlsetlevel)
project_data$flag_earlyGL <- as.factor(project_data$flag_earlyGL)
```

```
project_data$flag_online <- as.factor(project_data$flag_online)

#recode Sex, M = 0, F = 1 then make as factor
project_data$Sex_SPR24 <- recode(project_data$Sex_SPR24, M = 0, F = 1)
project_data$Sex_SPR24 <- as.factor(project_data$Sex_SPR24)

#Recode Language, ENB, ENG = 0; OTB, OTH = 1; everything else = NA, then make as factor
project_data$EAL <- recode(project_data$LanguageGroupMinor_SPR24, ENB = 0, ENG = 0, OTB
= 1, OTH = 1)
project_data$EAL <- as.factor(project_data$EAL)

#Recode FSM as factor
project_data$EVERFSM_6_P_SPR24 <- as.factor(project_data$EVERFSM_6_P_SPR24)

#More tidying
project_data <-
 project_data |>
 mutate(school.class = paste0(as.character(RU_SchoolID),'.',as.character(RU_ClassID)),
     subclass = factor(as.numeric(subclass)),
     N_pupils_byschool_schrep = as.numeric(N_pupils_byschool_schrep),
     N_pupils_byclass_schrep = as.numeric(N_pupils_byclass_schrep),
     hmlsetlevel = case_match(hmlsetlevel, '1' ~ 'H', '2' ~ 'M', '3' ~ 'L', .default = NA) |>
      factor(levels=c('L','M','H'), ordered=TRUE),
     IDACI = factor(IDACI, levels=c(1,2,3,4,5)))

#Importing additional 'self-admin' data
self_admin_data <- readxl::read_excel("SGS_test_self_admin_data_forSRS.xlsx")

#Join this self-admin data to the main project data
project_data <-
 project_data |>
 left_join(self_admin_data |>
       dplyr::select(RU_PupilID, flag_selfadmin) |>
       mutate(RU_PupilID = as.factor(RU_PupilID)),
      by = join_by(RU_PupilID))

#Importing new maths self-confidence data
MSC_data <- readxl::read_excel("SGS_new_scales_Sep2025.xlsx")

#Join this MSC data to the main project data
project_data <-
 project_data |>
 left_join(MSC_data |>
       dplyr::select(RU_PupilID, rawSSCbase_scale, rawMSCbase_scale, rawSSCend_scale,
rawMSCend_scale) |>
       mutate(RU_PupilID = as.factor(RU_PupilID)),
      by = join_by(RU_PupilID))
```

```
#Check SSC scores match and delete new column:
identical(project_data$rawSSCbase_scale.x, project_data$rawSSCbase_scale.y)
identical(project_data$rawSSCend_scale.x, project_data$rawSSCend_scale.y)
project_data$rawSSCbase_scale.y <- NULL
project_data$rawSSCend_scale.y <- NULL
project_data <- project_data %>% rename(rawSSCbase_scale = rawSSCbase_scale.x)
project_data <- project_data %>% rename(rawSSCend_scale = rawSSCend_scale.x)

#Check MSC scores DON'T match and delete old column:
identical(project_data$rawMSCbase_scale.x, project_data$rawMSCbase_scale.y)
identical(project_data$rawMSCend_scale.x, project_data$rawMSCend_scale.y)
project_data$rawMSCbase_scale.x <- NULL
project_data$rawMSCend_scale.x <- NULL
project_data <- project_data %>% rename(rawMSCbase_scale = rawMSCbase_scale.y)
project_data <- project_data %>% rename(rawMSCend_scale = rawMSCend_scale.y)

#Make new variables flagging individuals as lo/hi attainers
KS2_MATMRK_tertiles <-
 project_data |>
 summarise(as_tibble_row(quantile(KS2_MATMRK, probs=c(1/3,2/3), na.rm=TRUE),
            .name_repair = \(x) paste0('tertile', (round(parse_number(x)/100*3))))) |>
 as.vector()

project_data <-
 project_data |>
 mutate(KS2_MATMRK_lo = if_else(KS2_MATMRK<KS2_MATMRK_tertiles[1],1,0),
     KS2_MATMRK_hi = if_else(KS2_MATMRK>KS2_MATMRK_tertiles[2],1,0),
     KS2_MATMRK_lmh = case_when(KS2_MATMRK<KS2_MATMRK_tertiles[1] ~ 'low',
                 KS2_MATMRK>KS2_MATMRK_tertiles[2] ~ 'hi',
                 KS2_MATMRK>=KS2_MATMRK_tertiles[1] &
KS2_MATMRK<=KS2_MATMRK_tertiles[2] ~ 'med') |> factor(levels=c('low','med','hi')))

#Recode urbanrural
project_data <-
 project_data |>
 mutate(ur_3cat = case_match(urbanrural,
              "Urban city and town" ~ "Urban city/town",
              "Urban major conurbation" ~ "Urban conurbation",
              "Rural town" ~ "Rural",
              "Urban minor conurbation" ~ "Urban conurbation",
              "Rural hamlet" ~ "Rural",
              "Rural village" ~ "Rural") |>
     factor(levels=c("Urban conurbation", "Urban city/town", "Rural")))

#Recode Ofsted
project_data <-
```

```
 project_data |>
 mutate(ofsted_rating = case_match(ofsted_rating, 1 ~ 'Outstanding', 2 ~ 'Good', 3 ~ 'Requires
improvement', .default = NA) |>
      factor(levels=c('Outstanding','Good','Requires improvement'), ordered=FALSE))

#Calculating weights
project_data <-
 project_data |>
 mutate(stratum.mixed = paste0(as.character(subclass),
                  '.',
                  as.character(mixedAttain)
                  ) |>
               factor(),
     is_complete_att = !(is.na(KS2_MATMRK) |
                  is.na(P2SAS) |
                  is.na(EVERFSM_6_P_SPR24) |
                  is.na(IDACIScore_19_SPR24) |
                  is.na(EAL)),
     is_complete_SSC = !(is.na(rawSSCbase_scale) |
                  is.na(rawSSCend_scale) |
                  is.na(EVERFSM_6_P_SPR24) |
                  is.na(IDACIScore_19_SPR24) |
                  is.na(EAL)),
     is_complete_MSC = !(is.na(rawMSCbase_scale) |
                  is.na(rawMSCend_scale) |
                  is.na(EVERFSM_6_P_SPR24) |
                  is.na(IDACIScore_19_SPR24) |
                  is.na(EAL)))

#Now need to count these is_complete_ variables by stratum.mixed, i.e. within each group
defined by a type of school in a stratum.
weighting_counts <-
 project_data |>
 group_by(subclass,mixedAttain) |>
 summarise(#n_SM = n(),
      num_complete_att = sum(is_complete_att),
      num_complete_SSC = sum(is_complete_SSC),
      num_complete_MSC = sum(is_complete_MSC),
      .groups = "drop") |>
 dplyr::ungroup() |>
 pivot_wider(names_from = mixedAttain,
       values_from = c("num_complete_att", "num_complete_SSC", "num_complete_MSC"),
       names_prefix = "mix",
       values_fill = 0)

#Now join this back to the main dataset and use the numbers to calculate the weights
project_data <-
```

```
 project_data |>
 left_join(weighting_counts, by='subclass') |>
 mutate(
  weight_att = case_when(
   num_complete_att_mix1*num_complete_att_mix0 == 0 ~ 0,
   mixedAttain==0 ~ num_complete_att_mix1/num_complete_att_mix0,
   mixedAttain==1 ~ 1),
  weight_SSC = case_when(
   num_complete_SSC_mix1*num_complete_SSC_mix0 == 0 ~ 0,
   mixedAttain==0 ~ num_complete_SSC_mix1/num_complete_SSC_mix0,
   mixedAttain==1 ~ 1),
  weight_MSC = case_when(
   num_complete_MSC_mix1*num_complete_MSC_mix0 == 0 ~ 0,
   mixedAttain==0 ~ num_complete_MSC_mix1/num_complete_MSC_mix0,
   mixedAttain==1 ~ 1))

#Set globals
DVs <- c("P2SAS", "rawSSCend_scale", "rawMSCend_scale")
baselines <- c("KS2_MATMRK", "rawSSCbase_scale", "rawMSCbase_scale")
covariates <- c("ks2aps_2019",
        "ptpriorlo_2019", "ptpriorhi_2019",
        "N_pupils_byschool_schrep",
        "EVERFSM_6_P_SPR24", "ptfsm6cla1a_2019",
        "EAL", "ptealgrp2_2019",
        "academy",
        "IDACIScore_19_SPR24", "IDACI",
        "ofsted_rating", "ur_3cat")

#Participant flow including losses and exclusions
#All schools:
project_data_tidy |>
 filter(subclass!=x & [...]) |>
 summarise(n_pup = n(),
      n_sch = n_distinct(RU_SchoolID),
      .by = mixedAttain) |>
 adorn_totals("row") |>
 knitr::kable()

#Schools/pupils eligible for analysis:
project_data_tidy |>
 filter(subclass!=x & [...] & is_complete_att) |>
 summarise(n_pup = n(),
      n_sch = n_distinct(RU_SchoolID),
      .by = mixedAttain) |>
 adorn_totals("row") |>
 knitr::kable()
```

```
#Schools/pupils in eligible strata:
project_data_tidy |>
 filter(subclass!=x & [...] & weight_att>0) |>
 summarise(n_pup = n(),
        n_sch = n_distinct(RU_SchoolID),
        .by = mixedAttain) |>
 adorn_totals("row") |>
 knitr::kable()

#Schools/pupils in eligible strata and eligible for analysis (final sample):
project_data_tidy |>
 filter(subclass!=x & [...] & is_complete_att & weight_att>0) |>
 summarise(n_pup = n(),
        n_sch = n_distinct(RU_SchoolID),
        .by = mixedAttain) |>
 adorn_totals("row") |>
 knitr::kable()

#MDES
#Pre-/post-test correlations
#Pupil weighted:
project_data_tidy_2 <-
 project_data_tidy|>
 filter(weight_att>0 & is_complete_att)

weightedCorr(project_data_tidy_2$P2SAS, project_data_tidy_2$KS2_MATMRK, weights =
project_data_tidy_2$weight_att, method = "pearson")

#FSM pupil weighted:
project_data_tidy_2_FSM <-
 project_data_tidy|>
 filter(weight_att>0 & is_complete_att==1 & EVERFSM_6_P_SPR24==1)

weightedCorr(project_data_tidy_2_FSM$P2SAS, project_data_tidy_2_FSM$KS2_MATMRK,
weights = project_data_tidy_2_FSM$weight_att, method = "pearson")

#School weighted:
project_data_tidy_2_school <-
 project_data_tidy_2 |>
 filter(weight_att>0 & is_complete_att) |>
 group_by(RU_SchoolID) |>
 summarise(
  mean_KS2_MATMRK = mean(KS2_MATMRK),
  mean_P2SAS = mean(P2SAS),
  subclass = first(subclass),
  mixedAttain = first(mixedAttain)
 )
```

```
strata_counts <- project_data_tidy_2_school |>
 count(subclass, name = "n_schools_subclass") |>
 mutate(subclass_setted = n_schools_subclass - 1)

project_data_tidy_2_school <- project_data_tidy_2_school |>
 left_join(strata_counts, by="subclass")

project_data_tidy_2_school <-
 project_data_tidy_2_school |>
 mutate(
  weight_sch = case_when(
        mixedAttain==0 ~ 1 / subclass_setted,
        mixedAttain==1 ~ 1))

weightedCorr(project_data_tidy_2_school$mean_P2SAS,
project_data_tidy_2_school$mean_KS2_MATMRK, weights =
project_data_tidy_2_school$weight_sch, method = "pearson")

#FSM school weighted:
project_data_tidy_2_FSM_school <-
 project_data_tidy_2_FSM |>
 filter(weight_att>0 & is_complete_att) |>
 group_by(RU_SchoolID) |>
 summarise(
  mean_KS2_MATMRK = mean(KS2_MATMRK),
  mean_P2SAS = mean(P2SAS),
  subclass = first(subclass),
  mixedAttain = first(mixedAttain)
 )

strata_counts <- project_data_tidy_2_FSM_school |>
 count(subclass, name = "n_schools_subclass") |>
 mutate(subclass_setted = n_schools_subclass - 1)

project_data_tidy_2_FSM_school <- project_data_tidy_2_FSM_school |>
 left_join(strata_counts, by="subclass")

project_data_tidy_2_FSM_school <-
 project_data_tidy_2_FSM_school |>
 mutate(
  weight_sch = case_when(
        mixedAttain==0 ~ 1 / subclass_setted,
        mixedAttain==1 ~ 1))
```

```
weightedCorr(project_data_tidy_2_FSM_school$mean_P2SAS,
project_data_tidy_2_FSM_school$mean_KS2_MATMRK, weights =
project_data_tidy_2_FSM_school$weight_sch, method = "pearson")

#ICCs
#All:
lme.empty.primA <- lme4::lmer(P2SAS ~ 1 + (1|RU_SchoolID),
            data = project_data_tidy,
            subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)
summary(lme.empty.primA)
vc <- lme4::VarCorr(lme.empty.primA)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.primA <- random_effect_var/(residual_var+random_effect_var))

#For FSM group:
lme.empty.primA <- lme4::lmer(P2SAS ~ 1 + (1|RU_SchoolID),
            data = project_data_tidy,
            subset = subset = project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0 & EVERFSM_6_P_SPR24==1)
summary(lme.empty.primA)
vc <- lme4::VarCorr(lme.empty.primA)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.primA <- random_effect_var/(residual_var+random_effect_var))

#Mean pupils in schools
#All pupils in final analysis:
project_data_tidy |>
 filter(weight_att>0 & is_complete_att) |>
 count(RU_SchoolID) |>
 summarise(mean_pupils = mean(n))

#FSM pupils in final analysis
project_data_tidy |>
 filter(weight_att>0 & is_complete_att==1 & EVERFSM_6_P_SPR24==1) |>
 count(RU_SchoolID) |>
 summarise(mean_pupils = mean(n))

#Mean matched groups of schools
project_data_tidy |>
 filter(weight_att>0 & is_complete_att) |>
 group_by(subclass) |>
 summarise(nn=length(unique(RU_SchoolID))) |>
 adorn_totals()

#Pupil and school characteristics
```

```
#Need to calculate some variables.
project_data_tidy <-
  project_data_tidy |>
  mutate(Average_att =
rowMeans(across(c("ks2aps_2017","ks2aps_2018","ks2aps_2019")),na.rm=TRUE),
     IDACI1 = as.numeric(IDACI==1),
     IDACI2 = as.numeric(IDACI==2),
     IDACI3 = as.numeric(IDACI==3),
     IDACI4 = as.numeric(IDACI==4),
     IDACI5 = as.numeric(IDACI==5),
     Ofsted_Outs = as.numeric(ofsted_rating=='Outstanding'),
     Ofsted_Good = as.numeric(ofsted_rating=='Good'),
     Ofsted_RI = as.numeric(ofsted_rating=='Requires improvement'),
     Urban = as.numeric(ur_3cat=="Urban conurbation" | ur_3cat=="Urban city/town"),
     FSM1 = as.numeric(EVERFSM_6_P_SPR24==1),
     EAL1 = as.numeric(EAL==1))

project_data_tidy_2 <-
  project_data_tidy_2 |>
  mutate(Average_att =
rowMeans(across(c("ks2aps_2017","ks2aps_2018","ks2aps_2019")),na.rm=TRUE),
     IDACI1 = as.numeric(IDACI==1),
     IDACI2 = as.numeric(IDACI==2),
     IDACI3 = as.numeric(IDACI==3),
     IDACI4 = as.numeric(IDACI==4),
     IDACI5 = as.numeric(IDACI==5),
     Ofsted_Outs = as.numeric(ofsted_rating=='Outstanding'),
     Ofsted_Good = as.numeric(ofsted_rating=='Good'),
     Ofsted_RI = as.numeric(ofsted_rating=='Requires improvement'),
     Urban = as.numeric(ur_3cat=="Urban conurbation" | ur_3cat=="Urban city/town"),
     FSM1 = as.numeric(EVERFSM_6_P_SPR24==1),
     EAL1 = as.numeric(EAL==1))

#Set up a vector of variable names that we want to analyse.
vars_to_balance <- c(
  "ks2aps_2019",
  "ks2aps_2018",
  "ks2aps_2017",
  "Average_att",
  "ptpriorlo_2019",
  "ptpriorhi_2019",
  "N_pupils_byschool_schrep",
  "ptfsm6cla1a_2019",
  "ptealgrp2_2019",
  "academy",
  "IDACI1",
  "IDACI2",
```

```
 "IDACI3",
 "IDACI4",
 "IDACI5",
 "Ofsted_Outs",
 "Ofsted_Good",
 "Ofsted_RI",
 "Urban",
 "KS2_MATMRK",
 "KS2_MATMRK_lo",
 "KS2_MATMRK_hi",
 "rawSSCbase_scale",
 "rawMSCbase_scale",
 "FSM1",
 "EAL1",
 "IDACIScore_19_SPR24")

#N / missing
project_data_tidy_2 |>
 summarise(across(all_of(vars_to_balance),
          list(
           non_miss = ~sum(!is.na(.x)),
           miss = ~sum(is.na(.x)),
           mean = ~TAM::weighted_mean(.x, weight_att)
          ), .names = "{.col}__{.fn}"),
      .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
        names_to = c("var", "stat"),
        names_pattern = "(.*)__(.*)",
        values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
       names_sep = "",
       values_from = val) |>
 modelsummary::datasummary_df()

#Standardised differences - Standard deviation from all schools in sample:
means_by_mixed <-
project_data_tidy |>
 filter(is_complete_att & weight_att>0) |>
  summarise(across(all_of(vars_to_balance),
        ~TAM::weighted_mean(.x, weight_att),
          .names = "{.col}__mean"),
        .by = mixedAttain )

sds_all <-
project_data_tidy |>
 filter(is_complete_att & weight_att>0) |>
  summarise(across(all_of(vars_to_balance),
```

```
          ~TAM::weighted_sd(.x, weight_att),
            .names = "{.col}__sd_all"))

means_long <-
  means_by_mixed |>
  pivot_longer(cols = -mixedAttain,
          names_to = c("var", "stat"),
          names_sep = "__") |>
  pivot_wider(names_from = mixedAttain, values_from = value, names_prefix =
"mean_mixedAttain_")

sds_long <-
  sds_all |>
  pivot_longer(everything(),
          names_to = c("var", "stat"),
          names_sep = "__") |>
  dplyr::select(var, sd_all = value)

combined <- means_long |>
  left_join(sds_long, by = "var") |>
  mutate(delta = (mean_mixedAttain_1 - mean_mixedAttain_0) / sd_all)

modelsummary::datasummary_df(combined)

#Baseline balance at pupil-level (unstandardised differences for mean, sd and skewness):
project_data_tidy |>
  filter(is_complete_att & weight_att>0) |>
  summarise(across(all_of(vars_to_balance),
            list(
              mean = ~TAM::weighted_mean(.x, weight_att),
              sd = ~TAM::weighted_sd(.x, weight_att),
              skew = ~TAM::weighted_skewness(.x, weight_att)
            ), .names = "{.col}__{.fn}"),
        .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
          names_to = c("var", "stat"),
          names_pattern = "(.*)__(.*)",
          values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
         names_sep = "",
         values_from = val) |>
  mutate(
    mean_diff = mean1 - mean0,
    #median_diff = median1 - median0,
    sd_diff = sd1 - sd0,
    skew_diff = skew1 - skew0
    ) |>
```

```
  modelsummary::datasummary_df()

#Baseline balance at school-level (unstandardised differences for mean, sd and skewness):
project_data_tidy |>
  filter(is_complete_att & weight_att>0) |>
  summarise(across(all_of(vars_to_balance),
                list(
                  mean = ~TAM::weighted_mean(.x, 1/N_pupils_byschool),
                  sd = ~TAM::weighted_sd(.x, 1/N_pupils_byschool),
                  skew = ~TAM::weighted_skewness(.x, 1/N_pupils_byschool)
                ), .names = "{.col}__{.fn}"),
            .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
               names_to = c("var", "stat"),
               names_pattern = "(.*)__(.*)",
               values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
              names_sep = "",
              values_from = val) |>
  mutate(
    mean_diff = mean1 - mean0,
    sd_diff = sd1 - sd0,
    skew_diff = skew1 - skew0
    ) |>
  modelsummary::datasummary_df()

#Baseline balance at pupil-level for fSM sub-group (unstandardised differences for mean, sd
and skewness):
project_data_tidy |>
  filter(is_complete_att & weight_att>0) |>
  filter(EVERFSM_6_P_SPR24 == 1) |>
  summarise(across(all_of(vars_to_balance),
                list(
                  mean = ~TAM::weighted_mean(.x, weight_att),
                  #median = ~TAM::weighted_quantile(.x, weight_att, probs=0.5),
                  sd = ~TAM::weighted_sd(.x, weight_att),
                  skew = ~TAM::weighted_skewness(.x, weight_att)
                ), .names = "{.col}__{.fn}"),
            .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
               names_to = c("var", "stat"),
               names_pattern = "(.*)__(.*)",
               values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
              names_sep = "",
              values_from = val) |>
  mutate(
```

```
   mean_diff = mean1 - mean0,
   #median_diff = median1 - median0,
   sd_diff = sd1 - sd0,
   skew_diff = skew1 - skew0
   ) |>
  modelsummary::datasummary_df()

#Primary analysis
#All students
#Means
#Unadjusted mean:
project_data_tidy |>
  filter(weight_att>0 & is_complete_att) |>
  summarise(across(all_of("P2SAS"),
              list(
               non_miss = ~sum(!is.na(.x)),
               miss = ~sum(is.na(.x)),
               mean = ~TAM::weighted_mean(.x, weight_att)
              ), .names = "{.col}__{.fn}"),
          .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
            names_to = c("var", "stat"),
            names_pattern = "(.*)__(.*)",
            values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
            names_sep = "",
            values_from = val) |>
  modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(P2SAS ~ mixedAttain,
        data = project_data_tidy,
        subset = weight_att>0 & is_complete_att,
        weights = weight_att)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model
fm.prim <-
  lm(reformulate(c(baselines[1],"mixedAttain",covariates),
            response=DVs[1]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_att,
    subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)
```

```
coef_names <- fm.prim$coefficients |> names()

modelsummary::modelsummary(
 fm.prim,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
Re-run ICC:
lme.empty.prim <- lme4::lmer(P2SAS ~ 1 + (1|RU_SchoolID),
              data = project_data_tidy,
              subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)
summary(lme.empty.prim)
vc <- lme4::VarCorr(lme.empty.prim)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.prim <- random_effect_var/(residual_var+random_effect_var))

#First code this as a function to find the effect size from beta, s p, n, m.
hedgesG <- function(beta, s, p, n, m){
 h <- (n-2-2*(n/m - 1)*p)^2 / ( (n-2)*(1-p)^2 - n/m*(n-2*n/m)*p^2 + 2*(n-2*n/m)*p*(1-p) )
 lam <- 1 - ( 2*n*p / (m*(n-1)) )
 j <- 1 - 3/(4*h-1)

 return(j*sqrt(lam)*beta/s)
}

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
n_pup_att <-
 project_data_tidy |>
 filter(is_complete_att & weight_att>0) |>
 nrow()

n_clu_att <-
 project_data_tidy |>
 filter(is_complete_att & weight_att>0) |>
 pull(subclass) |>
 unique() |>
 length()

beta <- fm.prim$coefficients[c('mixedAttain1')] |> unname()
mixedAttain1_SE <- x

hedgesG(
```

```
 beta = beta,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
                 data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = beta + c(-1,1) * 1.96 * mixedAttain1_SE,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
                 data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Low attainers
#Means
#Unadjusted means:
project_data_tidy |>
 filter(weight_att>0 & is_complete_att==1 & KS2_MATMRK_lmh=="low") |>
 summarise(across(all_of("P2SAS"),
          list(
           non_miss = ~sum(!is.na(.x)),
           miss = ~sum(is.na(.x)),
           mean = ~TAM::weighted_mean(.x, weight_att)
          ), .names = "{.col}__{.fn}"),
      .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
        names_to = c("var", "stat"),
        names_pattern = "(.*)__(.*)",
        values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
       names_sep = "",
       values_from = val) |>
 modelsummary::datasummary_df()
Unadjusted mean from regression (to check):
mod <- lm(P2SAS ~ mixedAttain,
     data = project_data_tidy,
     subset = weight_att>0 & is_complete_att==1 & KS2_MATMRK_lmh=="low",
     weights = weight_att)
summary(mod)

#Confidence interval from regression:
```

```
confint(mod)

#Full regression model
hilo_att_mix <- "KS2_MATMRK_lo*mixedAttain + KS2_MATMRK_hi*mixedAttain"

fm.prim.att <-
 lm(reformulate(c(baselines[1],"mixedAttain",hilo_att_mix,covariates),
          response=DVs[1]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)

coef_names <- fm.prim.att$coefficients |> names()

modelsummary::modelsummary(#output='data.frame',
 fm.prim.att,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 coef_map = coef_names[c(seq(1,5),24,25,seq(6,23))],
 vcov=list('iid'='iid',
       'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
#Combine coefficients
fm.prim.att <-
 lm(reformulate(c(baselines[1],"mixedAttain",hilo_att_mix,covariates),
          response=DVs[1]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)

vcov_cluster <- sandwich::vcovCL(fm.prim.att, cluster=~RU_SchoolID)
att_lo_full <- glht(fm.prim.att, linfct = c("mixedAttain1 + mixedAttain1:KS2_MATMRK_lo = 0"),
vcov = vcov_cluster)
summary_att_lo_full <- summary(att_lo_full)
summary_att_lo_full
mixedAttain1_beta_low <- summary_att_lo_full$test$coefficients |> unname()
mixedAttain1_se_low <- summary_att_lo_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
 beta = mixedAttain1_beta_low,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
            data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
```

```
 m = n_clu_att
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = mixedAttain1_beta_low + c(-1,1) * 1.96 * mixedAttain1_se_low,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
                data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Medium attainers
#Means
#Unadjusted means:
project_data_tidy |>
 filter(weight_att>0 & is_complete_att==1 & KS2_MATMRK_lmh=="med") |>
 summarise(across(all_of("P2SAS"),
          list(
           non_miss = ~sum(!is.na(.x)),
           miss = ~sum(is.na(.x)),
           mean = ~TAM::weighted_mean(.x, weight_att)
          ), .names = "{.col}__{.fn}"),
       .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
        names_to = c("var", "stat"),
        names_pattern = "(.*)__(.*)",
        values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
       names_sep = "",
       values_from = val) |>
 modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(P2SAS ~ mixedAttain,
     data = project_data_tidy,
     subset = weight_att>0 & is_complete_att==1 & KS2_MATMRK_lmh=="med",
     weights = weight_att)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Effect size
att_mid_full <- glht(fm.prim.att, linfct = c("mixedAttain1 = 0"), vcov = vcov_cluster)
```

```
summary_att_mid_full <- summary(att_mid_full)
summary_att_mid_full
mixedAttain1_beta_mid <- summary_att_mid_full$test$coefficients |> unname()
mixedAttain1_se_mid <- summary_att_mid_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
 beta = mixedAttain1_beta_mid,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
              data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = mixedAttain1_beta_mid + c(-1,1) * 1.96 * mixedAttain1_se_mid,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
              data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#High attainers
#Means
#Unadjusted means:
project_data_tidy |>
 filter(weight_att>0 & is_complete_att==1 & KS2_MATMRK_lmh=="hi") |>
 summarise(across(all_of("P2SAS"),
         list(
          non_miss = ~sum(!is.na(.x)),
          miss = ~sum(is.na(.x)),
          mean = ~TAM::weighted_mean(.x, weight_att)
         ), .names = "{.col}__{.fn}"),
      .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
       names_to = c("var", "stat"),
       names_pattern = "(.*)__(.*)",
       values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
       names_sep = "",
       values_from = val) |>
 modelsummary::datasummary_df()
```

```
#Unadjusted mean from regression (to check):
mod <- lm(P2SAS ~ mixedAttain,
     data = project_data_tidy,
     subset = weight_att>0 & is_complete_att==1 & KS2_MATMRK_lmh=="hi",
     weights = weight_att)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Effect size
att_hi_full <- glht(fm.prim.att, linfct = c("mixedAttain1 + mixedAttain1:KS2_MATMRK_hi = 0"),
vcov = vcov_cluster)
summary_att_hi_full <- summary(att_hi_full)
summary_att_hi_full
mixedAttain1_beta_high <- summary_att_hi_full$test$coefficients |> unname()
mixedAttain1_se_high <- summary_att_hi_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
 beta = mixedAttain1_beta_high,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
              data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = mixedAttain1_beta_high + c(-1,1) * 1.96 * mixedAttain1_se_high,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
              data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Low attainment - highattainment gap
att_diff <- glht(fm.prim.att, linfct = c("mixedAttain1:KS2_MATMRK_hi -
mixedAttain1:KS2_MATMRK_lo = 0"), vcov = vcov_cluster)
summary(att_diff)

#Updating the beta, everything else stays the same.
beta <- coef(att_diff) |> unname()
```

```
#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
 beta = beta,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
                   data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = beta + c(-1,1) * 1.96 * mixedAttain1_SE,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
                   data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#FSM
#Means
#Unadjusted means:
project_data_tidy |>
 filter(weight_att>0 & is_complete_att==1 & EVERFSM_6_P_SPR24==1) |>
 summarise(across(all_of("P2SAS"),
          list(
           non_miss = ~sum(!is.na(.x)),
           miss = ~sum(is.na(.x)),
           mean = ~TAM::weighted_mean(.x, weight_att)
          ), .names = "{.col}__{.fn}"),
       .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
        names_to = c("var", "stat"),
        names_pattern = "(.*)__(.*)",
        values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
        names_sep = "",
        values_from = val) |>
 modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(P2SAS ~ mixedAttain,
     data = project_data_tidy,
     subset = weight_att>0 & is_complete_att==1 & EVERFSM_6_P_SPR24==1,
     weights = weight_att)
```

```
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model subsample
fm.prim.fsm.subset <-
 lm(reformulate(c(baselines[1],"mixedAttain",setdiff(covariates,"EVERFSM_6_P_SPR24")),
          response=DVs[1]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att &
        project_data_tidy$weight_att>0 &
        project_data_tidy$EVERFSM_6_P_SPR24 == 1)

modelsummary::modelsummary(
 fm.prim.fsm.subset,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
       'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Full regression model interacted
fsm_att_mix <- "EVERFSM_6_P_SPR24*mixedAttain"

fm.prim.fsm <-
 lm(reformulate(c(baselines[1],"mixedAttain",fsm_att_mix,covariates),
          response=DVs[1]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)

coef_names <- fm.prim.fsm$coefficients |> names()

modelsummary::modelsummary(#output='data.frame',
 fm.prim.fsm,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
       'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
fm.prim.fsm <-
 lm(reformulate(c(baselines[1],"mixedAttain",fsm_att_mix,covariates),
          response=DVs[1]),
```

```
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)

vcov_cluster <- sandwich::vcovCL(fm.prim.fsm, cluster=~RU_SchoolID)
att_fsm_full <- glht(fm.prim.fsm, linfct = c("mixedAttain1 +
mixedAttain1:EVERFSM_6_P_SPR241 = 0"), vcov = vcov_cluster)
summary_att_fsm_full <- summary(att_fsm_full)
summary_att_fsm_full
mixedAttain1_beta_fsm <- summary_att_fsm_full$test$coefficients |> unname()
mixedAttain1_se_fsm <- summary_att_fsm_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
 beta = mixedAttain1_beta_fsm,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
             data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = mixedAttain1_beta_fsm + c(-1,1) * 1.96 * mixedAttain1_se_fsm,
 s = effectsize::sd_pooled(P2SAS ~ mixedAttain,
             data = project_data_tidy |> filter(is_complete_att & weight_att>0)),
 p = icc.prim,
 n = n_pup_att,
 m = n_clu_att
)

#Secondary analysis: General self-confidence
#All students
#Means
#Unadjusted mean:
project_data_tidy |>
 filter(weight_SSC>0 & is_complete_SSC) |>
 summarise(across(all_of("rawSSCend_scale"),
          list(
           non_miss = ~sum(!is.na(.x)),
           miss = ~sum(is.na(.x)),
           mean = ~TAM::weighted_mean(.x, weight_SSC)
          ), .names = "{.col}__{.fn}"),
      .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
```

```
           names_to = c("var", "stat"),
           names_pattern = "(.*)__(.*)",
           values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
          names_sep = "",
          values_from = val) |>
  modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(rawSSCend_scale ~ mixedAttain,
       data = project_data_tidy,
       subset = weight_SSC>0 & is_complete_SSC==1,
       weights = weight_SSC)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model
fm.sec1 <-
  lm(reformulate(c(baselines[2],"mixedAttain",covariates),
            response=DVs[2]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_SSC,
    subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)

modelsummary::modelsummary(
  fm.sec1,
  fmt=modelsummary::fmt_significant(3),
  statistic=c('std.error','conf.int','p.value'),
  gof_map=c("nobs","std.error.type","se_type","vcov.type"),
  vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
#ICC:
lme.empty.sec1 <- lme4::lmer(rawSSCend_scale ~ 1 + (1|RU_SchoolID),
              data = project_data_tidy,
              subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)
summary(lme.empty.sec1)
vc <- lme4::VarCorr(lme.empty.sec1)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.sec1 <- random_effect_var/(residual_var+random_effect_var))
n_pup_SSC <-
  project_data_tidy |>
  filter(is_complete_SSC & weight_SSC>0) |>
```

```
  nrow()

n_clu_SSC <-
  project_data_tidy |>
  filter(is_complete_SSC & weight_SSC>0) |>
  pull(subclass) |>
  unique() |>
  length()

beta <- fm.sec1$coefficients[c('mixedAttain1')] |> unname()
mixedAttain1_SE <- x

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG <- function(beta, s, p, n, m){
  h <- (n-2-2*(n/m - 1)*p)^2 / ( (n-2)*(1-p)^2 - n/m*(n-2*n/m)*p^2 + 2*(n-2*n/m)*p*(1-p) )
  lam <- 1 - ( 2*n*p / (m*(n-1)) )
  j <- 1 - 3/(4*h-1)

  return(j*sqrt(lam)*beta/s)
}

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
  beta = beta,
  s = effectsize::sd_pooled(rawSSCend_scale ~ mixedAttain,
                    data = project_data_tidy |> filter(is_complete_SSC & weight_SSC>0)),
  p = icc.sec1,
  n = n_pup_SSC,
  m = n_clu_SSC
)

hedgesG(
  beta = beta + c(-1,1) * 1.96 * mixedAttain1_SE,
  s = effectsize::sd_pooled(rawSSCend_scale ~ mixedAttain,
                    data = project_data_tidy |> filter(is_complete_SSC & weight_SSC>0)),
  p = icc.sec1,
  n = n_pup_SSC,
  m = n_clu_SSC
)

#Low attainers
#Means
#Unadjusted mean:
project_data_tidy |>
  filter(weight_SSC>0 & is_complete_SSC==1 & KS2_MATMRK_lmh=="low") |>
  summarise(across(all_of("rawSSCend_scale"),
```

```
              list(
                non_miss = ~sum(!is.na(.x)),
                miss = ~sum(is.na(.x)),
                mean = ~TAM::weighted_mean(.x, weight_SSC)
              ), .names = "{.col}__{.fn}"),
         .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
            names_to = c("var", "stat"),
            names_pattern = "(.*)__(.*)",
            values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
           names_sep = "",
           values_from = val) |>
  modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(rawSSCend_scale ~ mixedAttain,
      data = project_data_tidy,
      subset = weight_SSC>0 & is_complete_SSC==1 & KS2_MATMRK_lmh=="low",
      weights = weight_SSC)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model (with interactions)
hilo_SSC_mix <- "KS2_MATMRK_lo*mixedAttain + KS2_MATMRK_hi*mixedAttain"

fm.sec1.SSC <-
  lm(reformulate(c(baselines[2],"mixedAttain",hilo_SSC_mix,covariates),
            response=DVs[2]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_SSC,
    subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)

coef_names <- fm.sec1.SSC$coefficients |> names()

modelsummary::modelsummary(#output='data.frame',
  fm.sec1.SSC,
  fmt=modelsummary::fmt_significant(3),
  statistic=c('std.error','conf.int','p.value'),
  gof_map=c("nobs","std.error.type","se_type","vcov.type"),
  coef_map = coef_names[c(seq(1,5),24,25,seq(6,23))],
  vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
```

```
#Combine coefficients:
fm.sec1.SSC <-
 lm(reformulate(c(baselines[2],"mixedAttain",hilo_SSC_mix,covariates),
        response=DVs[2]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_SSC,
   subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)

vcov_cluster <- sandwich::vcovCL(fm.sec1.SSC, cluster=~RU_SchoolID)
library(TH.data)
library(multcomp)

SSC_lo_full <- glht(fm.sec1.SSC, linfct = c("mixedAttain1 + mixedAttain1:KS2_MATMRK_lo = 0"),
vcov = vcov_cluster)
summary_SSC_lo_full <- summary(SSC_lo_full)
summary_SSC_lo_full
mixedAttain1_beta_low <- summary_SSC_lo_full$test$coefficients |> unname()
mixedAttain1_se_low <- summary_SSC_lo_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
 beta = mixedAttain1_beta_low,
 s = effectsize::sd_pooled(rawSSCend_scale ~ mixedAttain,
          data = project_data_tidy |> filter(is_complete_SSC & weight_SSC>0)),
 p = icc.sec1,
 n = n_pup_SSC,
 m = n_clu_SSC
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
 beta = mixedAttain1_beta_low + c(-1,1) * 1.96 * mixedAttain1_se_low,
 s = effectsize::sd_pooled(rawSSCend_scale ~ mixedAttain,
          data = project_data_tidy |> filter(is_complete_SSC & weight_SSC>0)),
 p = icc.sec1,
 n = n_pup_SSC,
 m = n_clu_SSC
)

#FSM
#Means
#Unadjusted mean:
project_data_tidy |>
 filter(weight_SSC>0 & is_complete_SSC==1 & EVERFSM_6_P_SPR24==1) |>
 summarise(across(all_of("rawSSCend_scale"),
        list(
```

```
              non_miss = ~sum(!is.na(.x)),
              miss = ~sum(is.na(.x)),
              mean = ~TAM::weighted_mean(.x, weight_SSC)
            ), .names = "{.col}__{.fn}"),
        .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
           names_to = c("var", "stat"),
           names_pattern = "(.*)__(.*)",
           values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
           names_sep = "",
           values_from = val) |>
  modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(rawSSCend_scale ~ mixedAttain,
        data = project_data_tidy,
        subset = weight_SSC>0 & is_complete_SSC==1 & EVERFSM_6_P_SPR24==1,
        weights = weight_SSC)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model
fsm_SSC_mix <- "EVERFSM_6_P_SPR24*mixedAttain"

fm.sec1.fsm <-
  lm(reformulate(c(baselines[2],"mixedAttain",fsm_SSC_mix,covariates),
            response=DVs[2]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_SSC,
    subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)

coef_names <- fm.sec1.fsm$coefficients |> names()

modelsummary::modelsummary(
  fm.sec1.fsm,
  fmt=modelsummary::fmt_significant(3),
  statistic=c('std.error','conf.int','p.value'),
  gof_map=c("nobs","std.error.type","se_type","vcov.type"),
  vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
fm.sec1.fsm <-
  lm(reformulate(c(baselines[2],"mixedAttain",fsm_SSC_mix,covariates),
```

```
            response=DVs[2]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_SSC,
    subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)

vcov_cluster <- sandwich::vcovCL(fm.sec1.fsm, cluster=~RU_SchoolID)

#Combine coefficients:
SSC_fsm_full <- glht(fm.sec1.fsm, linfct = c("mixedAttain1 +
mixedAttain1:EVERFSM_6_P_SPR241 = 0"), vcov = vcov_cluster)
summary_SSC_fsm_full <- summary(SSC_fsm_full)
summary_SSC_fsm_full
mixedAttain1_beta_fsm <- summary_SSC_fsm_full$test$coefficients |> unname()
mixedAttain1_se_fsm <- summary_SSC_fsm_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
  beta = mixedAttain1_beta_fsm,
  s = effectsize::sd_pooled(rawSSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_SSC & weight_SSC>0)),
  p = icc.sec1,
  n = n_pup_SSC,
  m = n_clu_SSC
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̄_1-x ̄_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
  beta = mixedAttain1_beta_fsm + c(-1,1) * 1.96 * mixedAttain1_se_fsm,
  s = effectsize::sd_pooled(rawSSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_SSC & weight_SSC>0)),
  p = icc.sec1,
  n = n_pup_SSC,
  m = n_clu_SSC
)

#Secondary analysis: Maths self-confidence
#All students
#Means
#Unadjusted mean:
project_data_tidy |>
  filter(weight_MSC>0 & is_complete_MSC) |>
  summarise(across(all_of("rawMSCend_scale"),
          list(
            non_miss = ~sum(!is.na(.x)),
            miss = ~sum(is.na(.x)),
            mean = ~TAM::weighted_mean(.x, weight_MSC)
```

```
            ), .names = "{.col}__{.fn}"),
        .by = mixedAttain) |>
  pivot_longer(cols = !mixedAttain,
          names_to = c("var", "stat"),
          names_pattern = "(.*)__(.*)",
          values_to = "val") |>
  pivot_wider(names_from = c(stat,mixedAttain),
          names_sep = "",
          values_from = val) |>
  modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(rawMSCend_scale ~ mixedAttain,
      data = project_data_tidy,
      subset = weight_MSC>0 & is_complete_MSC==1,
      weights = weight_MSC)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model
fm.sec2 <-
  lm(reformulate(c(baselines[3],"mixedAttain",covariates),
            response=DVs[3]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_MSC,
    subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)

modelsummary::modelsummary(
  fm.sec2,
  fmt=modelsummary::fmt_significant(3),
  statistic=c('std.error','conf.int','p.value'),
  gof_map=c("nobs","std.error.type","se_type","vcov.type"),
  vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
#ICC:
lme.empty.sec2 <- lme4::lmer(rawMSCend_scale ~ 1 + (1|RU_SchoolID),
              data = project_data_tidy,
              subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)
summary(lme.empty.sec2)
vc <- lme4::VarCorr(lme.empty.sec2)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.sec2 <- random_effect_var/(residual_var+random_effect_var))
```

```
n_pup_MSC <-
  project_data_tidy |>
  filter(is_complete_MSC & weight_MSC>0) |>
  nrow()

n_clu_MSC <-
  project_data_tidy |>
  filter(is_complete_MSC & weight_MSC>0) |>
  pull(subclass) |>
  unique() |>
  length()

beta <- fm.sec2$coefficients[c('mixedAttain1')] |> unname()
mixedAttain1_SE <- x

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG <- function(beta, s, p, n, m){
  h <- (n-2-2*(n/m - 1)*p)^2 / ( (n-2)*(1-p)^2 - n/m*(n-2*n/m)*p^2 + 2*(n-2*n/m)*p*(1-p) )
  lam <- 1 - ( 2*n*p / (m*(n-1)) )
  j <- 1 - 3/(4*h-1)

  return(j*sqrt(lam)*beta/s)
}

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̄_1-x ̄_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
  beta = beta,
  s = effectsize::sd_pooled(rawMSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_MSC & weight_MSC>0)),
  p = icc.sec2,
  n = n_pup_MSC,
  m = n_clu_MSC
)

hedgesG(
  beta = beta + c(-1,1) * 1.96 * mixedAttain1_SE,
  s = effectsize::sd_pooled(rawMSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_MSC & weight_MSC>0)),
  p = icc.sec2,
  n = n_pup_MSC,
  m = n_clu_MSC
)

#Low attainers
#Means
#Unadjusted mean:
```

```
project_data_tidy |>
 filter(weight_MSC>0 & is_complete_MSC==1 & KS2_MATMRK_lmh=="low") |>
 summarise(across(all_of("rawMSCend_scale"),
          list(
           non_miss = ~sum(!is.na(.x)),
           miss = ~sum(is.na(.x)),
           mean = ~TAM::weighted_mean(.x, weight_MSC)
          ), .names = "{.col}__{.fn}"),
       .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
        names_to = c("var", "stat"),
        names_pattern = "(.*)__(.*)",
        values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
        names_sep = "",
        values_from = val) |>
 modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(rawMSCend_scale ~ mixedAttain,
     data = project_data_tidy,
     subset = weight_MSC>0 & is_complete_MSC==1 & KS2_MATMRK_lmh=="low",
     weights = weight_MSC)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model (with interactions)
hilo_MSC_mix <- "KS2_MATMRK_lo*mixedAttain + KS2_MATMRK_hi*mixedAttain"

fm.sec2.MSC <-
 lm(reformulate(c(baselines[3],"mixedAttain",hilo_MSC_mix,covariates),
         response=DVs[3]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_MSC,
   subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)

coef_names <- fm.sec2.MSC$coefficients |> names()

modelsummary::modelsummary(#output='data.frame',
 fm.sec2.MSC,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 coef_map = coef_names[c(seq(1,5),24,25,seq(6,23))],
 vcov=list('iid'='iid',
```

```
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
#Combine coefficients:
fm.sec2.MSC <-
  lm(reformulate(c(baselines[3],"mixedAttain",hilo_MSC_mix,covariates),
           response=DVs[3]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_MSC,
    subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)

vcov_cluster <- sandwich::vcovCL(fm.sec2.MSC, cluster=~RU_SchoolID)
library(TH.data)
library(multcomp)

MSC_lo_full <- glht(fm.sec2.MSC, linfct = c("mixedAttain1 + mixedAttain1:KS2_MATMRK_lo =
0"), vcov = vcov_cluster)
summary_MSC_lo_full <- summary(MSC_lo_full)
summary_MSC_lo_full
mixedAttain1_beta_low <- summary_MSC_lo_full$test$coefficients |> unname()
mixedAttain1_se_low <- summary_MSC_lo_full$test$sigma |> unname()
Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
  beta = mixedAttain1_beta_low,
  s = effectsize::sd_pooled(rawMSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_MSC & weight_MSC>0)),
  p = icc.sec2,
  n = n_pup_MSC,
  m = n_clu_MSC
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
  beta = mixedAttain1_beta_low + c(-1,1) * 1.96 * mixedAttain1_se_low,
  s = effectsize::sd_pooled(rawMSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_MSC & weight_MSC>0)),
  p = icc.sec2,
  n = n_pup_MSC,
  m = n_clu_MSC
)

#FSM
#Means
#Unadjusted mean:
project_data_tidy |>
  filter(weight_MSC>0 & is_complete_MSC==1 & EVERFSM_6_P_SPR24==1) |>
```

```
 summarise(across(all_of("rawMSCend_scale"),
             list(
               non_miss = ~sum(!is.na(.x)),
               miss = ~sum(is.na(.x)),
               mean = ~TAM::weighted_mean(.x, weight_MSC)
             ), .names = "{.col}__{.fn}"),
        .by = mixedAttain) |>
 pivot_longer(cols = !mixedAttain,
          names_to = c("var", "stat"),
          names_pattern = "(.*)__(.*)",
          values_to = "val") |>
 pivot_wider(names_from = c(stat,mixedAttain),
          names_sep = "",
          values_from = val) |>
 modelsummary::datasummary_df()

#Unadjusted mean from regression (to check):
mod <- lm(rawMSCend_scale ~ mixedAttain,
      data = project_data_tidy,
      subset = weight_MSC>0 & is_complete_MSC==1 & EVERFSM_6_P_SPR24==1,
      weights = weight_MSC)
summary(mod)

#Confidence interval from regression:
confint(mod)

#Full regression model
fsm_MSC_mix <- "EVERFSM_6_P_SPR24*mixedAttain"

fm.sec2.fsm <-
 lm(reformulate(c(baselines[3],"mixedAttain",fsm_MSC_mix,covariates),
            response=DVs[3]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_MSC,
   subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)

coef_names <- fm.sec2.fsm$coefficients |> names()

modelsummary::modelsummary(
 fm.sec2.fsm,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Effect size
```

```
fm.sec2.fsm <-
  lm(reformulate(c(baselines[3],"mixedAttain",fsm_MSC_mix,covariates),
             response=DVs[3]),
    data = project_data_tidy,
    weights = project_data_tidy$weight_MSC,
    subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)

vcov_cluster <- sandwich::vcovCL(fm.sec2.fsm, cluster=~RU_SchoolID)

#Combine coefficients:
MSC_fsm_full <- glht(fm.sec2.fsm, linfct = c("mixedAttain1 +
mixedAttain1:EVERFSM_6_P_SPR241 = 0"), vcov = vcov_cluster)
summary_MSC_fsm_full <- summary(MSC_fsm_full)
summary_MSC_fsm_full
mixedAttain1_beta_fsm <- summary_MSC_fsm_full$test$coefficients |> unname()
mixedAttain1_se_fsm <- summary_MSC_fsm_full$test$sigma |> unname()

#Now calculate using the pooled SD for a genuine Hedges g as in the SAP:
hedgesG(
  beta = mixedAttain1_beta_fsm,
  s = effectsize::sd_pooled(rawMSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_MSC & weight_MSC>0)),
  p = icc.sec2,
  n = n_pup_MSC,
  m = n_clu_MSC
)

#Now here are the endpoints of a CI for effect size, calculated by putting endpoints of the CI for
β_1=x ̅_1-x ̅_2 (i.e. estimate ± 1.96 SE) into the same formula
hedgesG(
  beta = mixedAttain1_beta_fsm + c(-1,1) * 1.96 * mixedAttain1_se_fsm,
  s = effectsize::sd_pooled(rawMSCend_scale ~ mixedAttain,
                  data = project_data_tidy |> filter(is_complete_MSC & weight_MSC>0)),
  p = icc.sec2,
  n = n_pup_MSC,
  m = n_clu_MSC
)

#Missing data analysis
#Describing and subsetting data net dropped schools (Primary analysis only)
#Dropping the entirely dropped schools. Both because of missing data and because of weights
calculations (strata).
project_data_tidy_2 <-
  subset(project_data_tidy, !(RU_SchoolID %in% c(x, [...])))

project_data_tidy_2_P2SAS <-
  subset(project_data_tidy_2, !(RU_SchoolID %in% c(x, [...])))
```

```
#Predict missingness
#Make a binary variable indicating missing P2SAS:
project_data_tidy_2_P2SAS <-
 project_data_tidy_2_P2SAS %>%
 mutate(miss_P2SAS=as.integer(is.na(P2SAS)))

#Predict missingness using variables from the imbalance testing.
project_data_tidy_2_P2SAS <-
 project_data_tidy_2_P2SAS |>
 mutate(Average_att =
rowMeans(across(c("ks2aps_2017","ks2aps_2018","ks2aps_2019")),na.rm=TRUE),
     Urban = as.numeric(ur_3cat=="Urban conurbation" | ur_3cat=="Urban city/town"))

miss_P2SAS_model <- miss_P2SAS ~
 ks2aps_2019 + ks2aps_2018 + ks2aps_2017 + Average_att + ptpriorlo_2019 + ptpriorhi_2019 +
 N_pupils_byschool_schrep + ptfsm6cla1a_2019 + ptealgrp2_2019 +
 academy + factor(IDACI) + factor(ofsted_rating) + Urban +
 KS2_MATMRK + EVERFSM_6_P_SPR24 + EAL +
 IDACIScore_19_SPR24

miss_P2SAS_results <-
 glm(miss_P2SAS_model,
   data = project_data_tidy_2_P2SAS,
   family = binomial(link="logit"))

modelsummary::modelsummary(
 miss_P2SAS_results,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
      'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#There is some predictability from some variables.

#Re-do the weights
#Make a new P2SAS variable with no NAs.
project_data_tidy_2_P2SAS <-
 project_data_tidy_2_P2SAS %>%
 mutate(P2SAS_all = P2SAS,
     P2SAS_all = ifelse(is.na(P2SAS_all), 100, P2SAS_all))

#Find new complete cases.
project_data_tidy_2_P2SAS <-
 project_data_tidy_2_P2SAS |>
 mutate(stratum.mixed = paste0(as.character(subclass),
```

```
                 '',
                 as.character(mixedAttain)
                 ) |>
               factor(),
      is_complete_att_P2SAS = !(is.na(KS2_MATMRK) |
                 is.na(P2SAS_all) |
                 is.na(EVERFSM_6_P_SPR24) |
                 is.na(IDACIScore_19_SPR24) |
                 is.na(EAL)))

project_data_tidy_2_P2SAS |> tabyl(is_complete_att_P2SAS)

#Make weights.
weighting_counts <-
 project_data_tidy_2_P2SAS |>
 group_by(subclass,mixedAttain) |>
 summarise(#n_SM = n(),
      num_complete_att_P2SAS = sum(is_complete_att_P2SAS)) |>
 dplyr::ungroup() |>
 pivot_wider(names_from = mixedAttain,
       values_from = c("num_complete_att_P2SAS"),
       names_prefix = "mix",
       values_fill = 0)

#Add weights to data.
project_data_tidy_2_P2SAS <-
 project_data_tidy_2_P2SAS |>
 left_join(weighting_counts, by='subclass') |>
 mutate(
  weight_att_P2SAS = case_when(
   mix1*mix0 == 0 ~ 0,
   mixedAttain==0 ~ mix1/mix0,
   mixedAttain==1 ~ 1))

#Implement MICE
#Keep only relevant columns.
project_data_tidy_2_P2SAS_mi <- project_data_tidy_2_P2SAS %>%
 dplyr::select(RU_SchoolID, mixedAttain, P2SAS, KS2_MATMRK,
        tpup_2019, N_pupils_byschool_schrep, academy,
        ks2aps_2019, ks2aps_2018, ks2aps_2017, Average_att,
        ptpriorlo_2019, ptpriorav_2019, ptpriorhi_2019,
        ptfsm6cla1a_2019, ptealgrp2_2019,
        ofsted_rating, Urban, ur_3cat,
        IDACI, rawSSCbase_scale, rawMSCbase_scale,
        EVERFSM_6_P_SPR24, EAL, IDACIScore_19_SPR24,
        KS2_MATMRK, KS2_MATMRK_lo, KS2_MATMRK_hi,
        is_complete_att_P2SAS, weight_att_P2SAS)
```

```
#Only predicting P2SAS, and using only the variables from the imbalance testing.
init = mice(project_data_tidy_2_P2SAS_mi, maxit=0)
meth = init$method
pred = init$predictionMatrix

pred <- make.predictorMatrix(project_data_tidy_2_P2SAS_mi)

#pred
pred[,] <- 0
#pred
pred["P2SAS",
c("ks2aps_2019","ks2aps_2018","ks2aps_2017","Average_att","ptpriorlo_2019","ptpriorhi_2019","
N_pupils_byschool_schrep","ptfsm6cla1a_2019","ptealgrp2_2019","academy","IDACI","ofsted_rat
ing","Urban","KS2_MATMRK","rawSSCbase_scale","rawMSCbase_scale","EVERFSM_6_P_SPR24","
EAL","IDACIScore_19_SPR24","KS2_MATMRK_lo","KS2_MATMRK_hi")] <- 1
#pred

meth[] <- ""
meth[3] <- "pmm"

#Run imputation using 20 imputed datasets, maxit = 5 (iterations for each imputed dataset).
imputed_data <- mice(project_data_tidy_2_P2SAS_mi, m=20, seed=123, predictorMatrix=pred,
method=meth)

#Combine datasets, run on our main Primary analysis model, and summerise results.
with.imp <- with(imputed_data, lm(reformulate(c(baselines[1],"mixedAttain",covariates),
           response=DVs[1]),
    weights = imputed_data$weight_att_P2SAS,
    subset = imputed_data$is_complete_att_P2SAS))

pooled <- pool(with.imp)
summary(pooled)

#Now with clustering:
datalist <- miceadds::mids2datlist(imputed_data)

cluster_mod <- lapply(datalist, FUN = function(project_data_tidy_2_P2SAS_mi){
  miceadds::lm.cluster(data = project_data_tidy_2_P2SAS_mi,
                formula = reformulate(c(baselines[1],"mixedAttain",covariates),
           response=DVs[1]),
               weights = imputed_data$weight_att_P2SAS,
               subset = imputed_data$is_complete_att_P2SAS,
                cluster = project_data_tidy_2_P2SAS_mi$RU_SchoolID)})

betas <- lapply(cluster_mod, FUN=function(rr){ coef(rr)})
vars <- lapply(cluster_mod, FUN=function(rr){ vcov(rr)})
```

```
summary(miceadds::pool_mi(qhat=betas,u=vars))

#Additional analyses: mediation analysis
#For all pupils
#A subset of the data to use:
project_data_tidy_mediation_subset <-
  project_data_tidy |>
  filter(is_complete_att,
       !is.na(meanOTL_scale),
       !is.na(meanTQ_scale)) |>
   filter(weight_att!=0)

#Then fit the 3 linear models
fm.otl <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates),
           response="meanOTL_scale"),
   data = project_data_tidy_mediation_subset,
   weights = project_data_tidy_mediation_subset$weight_att)

fm.tq <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates),
           response="meanTQ_scale"),
   data = project_data_tidy_mediation_subset,
   weights = project_data_tidy_mediation_subset$weight_att)

fm.outcome <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates,"meanOTL_scale","meanTQ_scale"),
           response=DVs[1]),
   data = project_data_tidy_mediation_subset,
   weights = project_data_tidy_mediation_subset$weight_att)

#Then run the `clustered_mediation.R' function.
source("clustered_mediation.R")
clustered_mediation <- function(
   mediator_model, # lm/lmer model predicting mediator from X (and covariates)
   outcome_model, # lm/lmer model predicting outcome from mediator, X (and covariates)
   x_var, # name of independent variable
   med_var, # name of mediator variable
   conf_level = 0.95, # confidence level
   bootstrap = FALSE, # whether to use bootstrapping for SEs
   boot_samples = 1000, # number of bootstrap samples
   cluster_var = NULL, # cluster variable name (for bootstrapping)
   data = NULL # original dataset (required for bootstrapping)
) {
   # Extract coefficients and standard errors
   # a path (X -> Mediator)
```

```
a_coef <- coef(summary(mediator_model))[x_var, "Estimate"]
a_se <- coef(summary(mediator_model))[x_var, "Std. Error"]

# b path (Mediator -> Y, controlling for X)
b_coef <- coef(summary(outcome_model))[med_var, "Estimate"]
b_se <- coef(summary(outcome_model))[med_var, "Std. Error"]

# c' path (direct effect: X -> Y, controlling for Mediator)
cp_coef <- coef(summary(outcome_model))[x_var, "Estimate"]
cp_se <- coef(summary(outcome_model))[x_var, "Std. Error"]

# Calculate indirect effect (a*b)
indirect_effect <- a_coef * b_coef

if (!bootstrap) {
  # Calculate standard error for indirect effect using delta method
  indirect_se <- sqrt(b_coef^2 * a_se^2 + a_coef^2 * b_se^2)
} else {
  if (is.null(data)) {
    stop("Data must be provided for bootstrapping")
  }

  # Bootstrap for standard errors
  boot_results <- bootstrap_mediation(
    data = data,
    x_var = x_var,
    med_var = med_var,
    mediator_formula = formula(mediator_model),
    outcome_formula = formula(outcome_model),
    cluster_var = cluster_var,
    n_boot = boot_samples
  )

  indirect_se <- sd(boot_results$indirect_effects, na.rm = TRUE)
  a_se <- sd(boot_results$a_effects, na.rm = TRUE)
  b_se <- sd(boot_results$b_effects, na.rm = TRUE)
  cp_se <- sd(boot_results$direct_effects, na.rm = TRUE)
}

# Total effect
total_effect <- cp_coef + indirect_effect

# Z values
z_value <- qnorm((1 + conf_level) / 2)

# Results table
results <- data.frame(
```

```
    Effect = c(
      "a (X→M)",
      "b (M→Y)",
      "c' (Direct)",
      "a*b (Indirect)",
      "Total"
    ),
    Estimate = c(a_coef, b_coef, cp_coef, indirect_effect, total_effect),
    SE = c(a_se, b_se, cp_se, indirect_se, sqrt(cp_se^2 + indirect_se^2)),
    Lower_CI = rep(NA, 5),
    Upper_CI = rep(NA, 5),
    p_value = rep(NA, 5)
  )

  # Compute CIs and p-values
  for (i in seq_len(nrow(results))) {
    results$Lower_CI[i] <- results$Estimate[i] - z_value * results$SE[i]
    results$Upper_CI[i] <- results$Estimate[i] + z_value * results$SE[i]
    results$p_value[i] <- 2 *
      pnorm(-abs(results$Estimate[i] / results$SE[i]))
  }

  # Add proportion mediated
  prop_mediated <- indirect_effect / total_effect

  return(list(
    results = results,
    prop_mediated = prop_mediated,
    conf_level = conf_level,
    bootstrap = bootstrap,
    boot_samples = if (bootstrap) boot_samples else NULL
  ))
}

# Helper function for bootstrapping
bootstrap_mediation <- function(
  data,
  x_var,
  med_var,
  mediator_formula,
  outcome_formula,
  cluster_var = NULL,
  n_boot = 1000
) {
  n <- nrow(data)
  a_effects <- numeric(n_boot)
  b_effects <- numeric(n_boot)
```

```
direct_effects <- numeric(n_boot)
indirect_effects <- numeric(n_boot)

# Function to resample clusters
resample_clusters <- function(data, cluster_var) {
  clusters <- unique(data[[cluster_var]])
  sampled_clusters <- sample(clusters, replace = TRUE)

  resampled_data <- do.call(
    rbind,
    lapply(sampled_clusters, function(cl) {
      data[data[[cluster_var]] == cl, ]
    })
  )

  return(resampled_data)
}

for (i in 1:n_boot) {
  # Resample with replacement
  if (is.null(cluster_var)) {
    boot_data <- data[sample(1:n, n, replace = TRUE), ]
  } else {
    boot_data <- resample_clusters(data, cluster_var)
  }

  # Fit models on bootstrap sample
  tryCatch(
    {
      med_model <- lm(mediator_formula, data = boot_data)
      out_model <- lm(outcome_formula, data = boot_data)

      # Extract coefficients
      a_effects[i] <- coef(med_model)[x_var]
      b_effects[i] <- coef(out_model)[med_var]
      direct_effects[i] <- coef(out_model)[x_var]
      indirect_effects[i] <- a_effects[i] * b_effects[i]
    },
    error = function(e) {
      a_effects[i] <- NA
      b_effects[i] <- NA
      direct_effects[i] <- NA
      indirect_effects[i] <- NA
    }
  )
}
```

```
  return(list(
    a_effects = a_effects,
    b_effects = b_effects,
    direct_effects = direct_effects,
    indirect_effects = indirect_effects
  ))
}

#Then implement.
set.seed(392)

mediate.otl.cluster <-
 clustered_mediation(fm.otl, fm.outcome,
          bootstrap = TRUE,
          boot_samples = 1000,
          x = "mixedAttain1",
          med_var = "meanOTL_scale",
          cluster_var = "RU_SchoolID",
          data = project_data_tidy_mediation_subset
          )
mediate.otl.cluster$results
mediate.tq.cluster <-
 clustered_mediation(fm.tq, fm.outcome,
          bootstrap = TRUE,
          boot_samples = 1000,
          x = "mixedAttain1",
          med_var = "meanTQ_scale",
          cluster_var = "RU_SchoolID",
          data = project_data_tidy_mediation_subset
          )
mediate.tq.cluster$results

#For low-attaining pupils
#First restrict the existing subset to only low-attaining pupils.
project_data_tidy_mediation_subset_lo <-
 project_data_tidy |>
 filter(is_complete_att,
     !is.na(meanOTL_scale),
     !is.na(meanTQ_scale)) |>
  filter(weight_att!=0) |>
 filter(KS2_MATMRK_lo == 1)

#Then fit the 3 linear models
fm.otl <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates),
        response="meanOTL_scale"),
  data = project_data_tidy_mediation_subset_lo,
```

```
   weights = project_data_tidy_mediation_subset_lo$weight_att)

fm.tq <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates),
           response="meanTQ_scale"),
   data = project_data_tidy_mediation_subset_lo,
   weights = project_data_tidy_mediation_subset_lo$weight_att)

fm.outcome <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates,"meanOTL_scale","meanTQ_scale"),
           response=DVs[1]),
   data = project_data_tidy_mediation_subset_lo,
   weights = project_data_tidy_mediation_subset_lo$weight_att)

##OTL:
set.seed(392)

mediate.otl.cluster <-
  clustered_mediation(fm.otl, fm.outcome,
              bootstrap = TRUE,
              boot_samples = 1000,
              x = "mixedAttain1",
              med_var = "meanOTL_scale",
              cluster_var = "RU_SchoolID",
              data = project_data_tidy_mediation_subset_lo
              )
mediate.otl.cluster$results
#TQ:
set.seed(392)

mediate.tq.cluster <-
  clustered_mediation(fm.tq, fm.outcome,
              bootstrap = TRUE,
              boot_samples = 1000,
              x = "mixedAttain1",
              med_var = "meanTQ_scale",
              cluster_var = "RU_SchoolID",
              data = project_data_tidy_mediation_subset_lo
              )
mediate.tq.cluster$results

##Additional analyses: quantile analysis
#10th
#Model:
fm.prim.quartile10 <-
  project_data_tidy |>
  quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
```

```
                    response=DVs[1]),
          #tau=c(0.25,0.5,0.75),
          tau=0.1,
          weights = weight_att,
          subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile10 |>
 modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
                   statistic=c('std.error','conf.int','p.value'),
                   gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
                   coef_map = coef_names[seq(1,3)],
                   se='boot',
                   bsmethod='cluster',
                   cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
                   R=1000)
#20th
#Model:
fm.prim.quartile20 <-
 project_data_tidy |>
 quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                    response=DVs[1]),
          #tau=c(0.25,0.5,0.75),
          tau=0.2,
          weights = weight_att,
          subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile20 |>
 modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
                   statistic=c('std.error','conf.int','p.value'),
                   gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
                   coef_map = coef_names[seq(1,3)],
                   se='boot',
                   bsmethod='cluster',
                   cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
                   R=1000)
#25th
#Model:
fm.prim.quartile25 <-
 project_data_tidy |>
```

```
 quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                   response=DVs[1]),
       #tau=c(0.25,0.5,0.75),
       tau=0.25,
       weights = weight_att,
       subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile25 |>
 modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
               statistic=c('std.error','conf.int','p.value'),
               gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
               coef_map = coef_names[seq(1,3)],
               se='boot',
               bsmethod='cluster',
               cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
               R=1000)
#30th
#Model:
fm.prim.quartile30 <-
 project_data_tidy |>
 quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                   response=DVs[1]),
       #tau=c(0.25,0.5,0.75),
       tau=0.3,
       weights = weight_att,
       subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile30 |>
 modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
               statistic=c('std.error','conf.int','p.value'),
               gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
               coef_map = coef_names[seq(1,3)],
               se='boot',
               bsmethod='cluster',
               cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
               R=1000)
#40th
#Model:
fm.prim.quartile40 <-
```

```
 project_data_tidy |>
 quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                        response=DVs[1]),
        #tau=c(0.25,0.5,0.75),
        tau=0.4,
        weights = weight_att,
        subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile40 |>
 modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
               statistic=c('std.error','conf.int','p.value'),
               gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
               coef_map = coef_names[seq(1,3)],
               se='boot',
               bsmethod='cluster',
               cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
               R=1000)
#50th
#Model:
fm.prim.quartile50 <-
 project_data_tidy |>
 quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                        response=DVs[1]),
        #tau=c(0.25,0.5,0.75),
        tau=0.5,
        weights = weight_att,
        subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile50 |>
 modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
               statistic=c('std.error','conf.int','p.value'),
               gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
               coef_map = coef_names[seq(1,3)],
               se='boot',
               bsmethod='cluster',
               cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
               R=1000)
#60th
#Model:
```

```
fm.prim.quartile60 <-
  project_data_tidy |>
  quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                          response=DVs[1]),
          #tau=c(0.25,0.5,0.75),
          tau=0.6,
          weights = weight_att,
          subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile60 |>
  modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
                    statistic=c('std.error','conf.int','p.value'),
                    gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
                    coef_map = coef_names[seq(1,3)],
                    se='boot',
                    bsmethod='cluster',
                    cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
                    R=1000)
#70th
#Model:
fm.prim.quartile70 <-
  project_data_tidy |>
  quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                          response=DVs[1]),
          #tau=c(0.25,0.5,0.75),
          tau=0.7,
          weights = weight_att,
          subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile70 |>
  modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
                    statistic=c('std.error','conf.int','p.value'),
                    gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
                    coef_map = coef_names[seq(1,3)],
                    se='boot',
                    bsmethod='cluster',
                    cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
                    R=1000)
#75th
```

```
#Model:
fm.prim.quartile75 <-
  project_data_tidy |>
  quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                        response=DVs[1]),
        #tau=c(0.25,0.5,0.75),
        tau=0.75,
        weights = weight_att,
        subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile75 |>
  modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
              statistic=c('std.error','conf.int','p.value'),
              gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
              coef_map = coef_names[seq(1,3)],
              se='boot',
              bsmethod='cluster',
              cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
              R=1000)
#80th
#Model:
fm.prim.quartile80 <-
  project_data_tidy |>
  quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                        response=DVs[1]),
        #tau=c(0.25,0.5,0.75),
        tau=0.8,
        weights = weight_att,
        subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile80 |>
  modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
              statistic=c('std.error','conf.int','p.value'),
              gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
              coef_map = coef_names[seq(1,3)],
              se='boot',
              bsmethod='cluster',
              cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
              R=1000)
```

```
#90th
#Model:
fm.prim.quartile90 <-
  project_data_tidy |>
  quantreg::rq(formula = reformulate(c(baselines[1],"mixedAttain",covariates),
                      response=DVs[1]),
         #tau=c(0.25,0.5,0.75),
         tau=0.9,
         weights = weight_att,
         subset = is_complete_att & weight_att>0)

#Bootstrapped SEs:
set.seed(1788)

fm.prim.quartile90 |>
  modelsummary::modelsummary(fmt=modelsummary::fmt_significant(3),
               statistic=c('std.error','conf.int','p.value'),
               gof_map=c("nobs","r.squared","std.error.type","se_type","vcov.type"),
               coef_map = coef_names[seq(1,3)],
               se='boot',
               bsmethod='cluster',
               cluster=project_data_tidy$RU_SchoolID[project_data_tidy$is_complete_att &
project_data_tidy$weight_att>0],
               R=1000)

#Pupil-matched analysis
#Create a new dataset and predict and plot propensity scores.
modelvars <- c("RU_SchoolID", "mixedAttain", "P2SAS",
         "tpup_2019", "N_pupils_byschool_schrep", "academy",
         "ks2aps_2019", "ks2aps_2018", "ks2aps_2017",
         "ptpriorlo_2019", "ptpriorav_2019", "ptpriorhi_2019",
         "ptfsm6cla1a_2019", "ptealgrp2_2019",
         "ofsted_rating",
         "region", "urbanrural", "ur_3cat",
         "IDACI",
         "Sex_SPR24",
         "EVERFSM_6_P_SPR24", "EAL", "IDACIScore_19_SPR24",
         "KS2_MATMRK","KS2_MATMRK_lo", "KS2_MATMRK_lmh", "KS2_MATMRK_hi")

project_data_tidy_pupil_matching <- project_data_tidy |>
         dplyr::select(all_of(modelvars)) |>
         na.omit(object = _)

matchModel <- mixedAttain ~ ptfsm6cla1a_2019 +
         ks2aps_2019 + ks2aps_2018 + ks2aps_2017 +
         ptpriorlo_2019 + ptpriorhi_2019 + tpup_2019 +
         factor(IDACI) + factor(ofsted_rating) +
```

```
          factor(Sex_SPR24) + factor(EVERFSM_6_P_SPR24) +
          IDACIScore_19_SPR24 + KS2_MATMRK

pscoremodel <- project_data_tidy_pupil_matching |>
          glm(matchModel, family = binomial(link = "logit"), data = _)
project_data_tidy_pupil_matching$pscore <- predict(pscoremodel, type = "response")

par(ps=10)
plot(density(project_data_tidy_pupil_matching$pscore[project_data_tidy_pupil_matching$mixe
dAttain==1], na.rm=TRUE), col="green",
   main="",
   xlab="Propensity Score", xlim=c(0,1),
   ylab="Density", ylim=c(0,5)
  )
lines(density(project_data_tidy_pupil_matching$pscore[project_data_tidy_pupil_matching$mix
edAttain==0], na.rm=TRUE), col="red")

#Match.
set.seed(392)

pupil_match <- matchit(matchModel,
           method = "nearest", distance = "logit",
           data = project_data_tidy_pupil_matching, discard = "none",
           #caliiper = 0.2,
           ratio = 1)
summary(pupil_match)
project_data_tidy_pupil_matched <- pupil_match |>
                     match.data()

par(ps=10)
plot(density(project_data_tidy_pupil_matched$pscore[project_data_tidy_pupil_matched$mixe
dAttain==1], na.rm=TRUE), col="green",
   main="",
   xlab="Propensity Score", xlim=c(0,1),
   ylab="Density", ylim=c(0,5)
  )
lines(density(project_data_tidy_pupil_matched$pscore[project_data_tidy_pupil_matched$mixe
dAttain==0], na.rm=TRUE), col="red")

#Main analysis:
fm.prim.pupil <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates),
           response=DVs[1]),
   data = project_data_tidy_pupil_matched)

modelsummary::modelsummary(
 fm.prim.pupil,
```

```
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Now add the indicators for low attainers & high attainers and their interaction with mixed status.
hilo_att_mix <- "KS2_MATMRK_lo*mixedAttain + KS2_MATMRK_hi*mixedAttain"

fm.prim.pupil <-
 lm(reformulate(c(baselines[1],"mixedAttain",hilo_att_mix,covariates),
            response=DVs[1]),
   data = project_data_tidy_pupil_matched)

coef_names <- fm.prim.pupil$coefficients |> names()


modelsummary::modelsummary(
 fm.prim.pupil,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 coef_map = coef_names[c(seq(1,5),24,25,seq(6,23))],
 vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#Additional analyses: self-administration
#First: model as for the primary analysis, but on the subset of pupils at schools that self-administered the test.
fm.prim.selfadmin.subset <-
 lm(reformulate(c(baselines[1],"mixedAttain",covariates),
            response=DVs[1]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att &
         project_data_tidy$weight_att>0 &
         project_data_tidy$flag_selfadmin == 1)

modelsummary::modelsummary(#output='data.frame',
 fm.prim.selfadmin.subset,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))
```

```
#Second: on the full dataset but with indicator for selfAdminister and
selfAdminister:mixedAttain interaction.
selfadmin_att_mix <- "flag_selfadmin + flag_selfadmin:mixedAttain"

fm.prim.selfadmin.addvar <-
 lm(reformulate(c(baselines[1],"mixedAttain",selfadmin_att_mix,covariates),
          response=DVs[1]),
   data = project_data_tidy,
   weights = project_data_tidy$weight_att,
   subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)

coef_names <- fm.prim.selfadmin.addvar$coefficients |> names()

modelsummary::modelsummary(#output='data.frame',
 fm.prim.selfadmin.addvar,
 fmt=modelsummary::fmt_significant(3),
 statistic=c('std.error','conf.int','p.value'),
 gof_map=c("nobs","std.error.type","se_type","vcov.type"),
 coef_map = coef_names[c(seq(1,4),23,seq(5,22))],
 vcov=list('iid'='iid',
        'CL'=function(x) sandwich::vcovCL(x,cluster=~RU_SchoolID)))

#ICCs
#PTM13
lme.empty.primA <- lme4::lmer(P2SAS ~ 1 + (1|RU_SchoolID),
              data = project_data_tidy,
              subset = project_data_tidy$is_complete_att & project_data_tidy$weight_att>0)
summary(lme.empty.primA)
vc <- lme4::VarCorr(lme.empty.primA)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.primA <- random_effect_var/(residual_var+random_effect_var))

#General self-confidence
lme.empty.sec1A <- lme4::lmer(rawSSCend_scale ~ 1 + (1|RU_SchoolID),
              data = project_data_tidy,
              subset = project_data_tidy$is_complete_SSC & project_data_tidy$weight_SSC>0)
summary(lme.empty.sec1A)
vc <- lme4::VarCorr(lme.empty.sec1A)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.sec1A <- random_effect_var/(residual_var+random_effect_var))

#Maths self-confidence
lme.empty.sec2A <- lme4::lmer(rawMSCend_scale ~ 1 + (1|RU_SchoolID),
              data = project_data_tidy,
              subset = project_data_tidy$is_complete_MSC & project_data_tidy$weight_MSC>0)
```

```
summary(lme.empty.sec2A)
vc <- lme4::VarCorr(lme.empty.sec2A)
(residual_var <- attr(vc, "sc")^2)
(random_effect_var <- vc$RU_SchoolID[1,1])
(icc.sec2A <- random_effect_var/(residual_var+random_effect_var))
```

Appendix C.16: Summary of analyses

Table C.16.1: Primary Analysis (Mathematics Attainment Outcome: PTM13): Impact of mixed attainment for all pupils

Table C.16.2: Primary Outcome Analysis with Interactions for Low & High Prior Attaining Pupils (Mathematics Attainment Outcome: PTM13): Distributional change analysis, impact of mixed attainment on low, middle and high prior attaining pupils

Table C.16.3: Secondary Analysis (General Self-Confidence): Impact of mixed attainment for all pupils

Table C.16.4: Secondary Analysis with Interactions for Low & High Prior Attaining Pupils (General Self-Confidence): Distributional change analysis, impact of mixed attainment on low, middle and high prior attaining pupils

Table C.16.5: Secondary Analysis (Mathematics Self-Confidence): Impact of mixed attainment for all pupils

Table C.16.6: Secondary Analysis with Interactions for Low & High Prior Attaining Pupils (Mathematics Self-Confidence): Distributional change analysis, impact of mixed attainment on low, middle and high prior attaining pupils

Table C.16.7: FSM Sub-Group Analysis (Mathematics Attainment Outcome: PTM13)

Table C.16.8: Analysis with FSM and Mixed Attainment Interaction on full dataset (Mathematics Attainment Outcome: PTM13)

Table C.16.9: Sensitivity Analysis Primary Outcome: Matched Pupil Sample

Table C.16.10: Sensitivity Analysis Primary Outcome with Interactions for Low & High Prior Attaining Pupils: Matched Pupil Sample

Table C.16.11: Summary of Quantile analysis: Coefficients, SEs & p-values for treatment effect

Table C.16.12: Summary of analysis examining school self-administration of outcome testing

Table C.16.1: Primary Analysis (Mathematics Attainment Outcome: PTM13). N=14877.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		16.2	37.8	-57.90, 90.30	0.668
KS2_MATMRK	KS2 Maths score	0.3749	0.00704	0.361, 0.389	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**-0.670**	**0.561**	**-1.769, 0.428**	**0.232**
ks2aps_2019	2019 School average KS2 attainment of intake	2.198	1.39	-0.519, 4.916	0.113
ptpriorlo_2019	School low prior attainment proportion	0.125	0.152	-0.174, 0.423	0.413
ptpriorhi_2019	School high prior attainment proportion	-0.16011	0.0860	-0.32876, 0.00853	0.063
N_pupils_byschool_schrep	Number of pupils in Y7	0.00287	0.00521	-0.00733, 0.01308	0.581
EVERFSM_6_P_SPR241	Pupil Ever FSM	-1.325	0.215	-1.747, -0.903	<0.001
ptfsm6cla1a_2019	School FSM proportion	-0.0201	0.0452	-0.1087, 0.0685	0.656
EAL1	Pupil EAL flag	1.027	0.361	0.321, 1.734	0.004
ptealgrp2_2019	School EAL proportion	0.0386	0.0153	0.00861, 0.06864	0.012
academy	School academy status	-0.661	0.577	-1.792, 0.471	0.252
IDACIScore_19_SPR24	Individual IDACI score	-5.22	1.14	-7.450, -3.000	<0.001
IDACI2	School IDACI quintile 2	-0.461	0.743	-1.919, 0.996	0.535
IDACI3	School IDACI quintile 3	-1	0.811	-2.589, 0.590	0.218
IDACI4	School IDACI quintile 4	0.315	0.918	-1.485, 2.114	0.732
IDACI5	School IDACI quintile 5	-3.97	1.30	-6.530, -1.410	0.002
ofsted_ratingGood	School Ofsted Good	0.153	0.728	-1.274, 1.579	0.834
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.876	1.33	-1.735, 3.487	0.511

ur_3catUrban city/town	Urban classification: city/town	1.438	0.718	0.0310, 2.8450	0.045
ur_3catRural	Urban classification: rural	1.798	0.963	-0.08940, 3.6860	0.062

Table C.16.2: Primary Outcome Analysis with Interactions (Mathematics Attainment Outcome: PTM13). N=14877.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		14.3	37.7	-59.60, 88.20	0.705
KS2_MATMRK	KS2 Maths score	0.32383	0.00813	0.30790, 0.33976	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**-0.571**	**0.590**	**-1.728, 0.587**	**0.334**
KS2_MATMRK_lo	**Pupil in the lower tertile of KS2 Maths scores**	**0.847**	**0.384**	**0.09360, 1.5995**	**0.028**
KS2_MATMRK_hi	**Pupil in the upper tertile of KS2 Maths scores**	**6.357**	**0.316**	**5.737, 6.978**	**<0.001**
mixedAttain1:KS2_MATMRK_lo	**Interaction: Treatment x Pupil in Lower Tertile KS2 Maths**	**1.149**	**0.448**	**0.271, 2.026**	**0.010**
mixedAttain1:KS2_MATMRK_hi	**Interaction: Treatment x Pupil in Upper Tertile KS2 Maths**	**-1.224**	**0.577**	**-2.3545, -0.09360**	**0.034**
ks2aps_2019	2019 School average KS2 attainment of intake	2.357	1.38	-0.349, 5.064	0.088
ptpriorlo_2019	School low prior attainment proportion	0.125	0.150	-0.169, 0.419	0.404
ptpriorhi_2019	School high prior attainment proportion	-0.1878	0.0848	-0.3540, -0.0216	0.027
N_pupils_byschool_schrep	Number of pupils in Y7	0.0026	0.00513	-0.00745, 0.01266	0.612
EVERFSM_6_P_SPR241	Pupil Ever FSM	-1.239	0.224	-1.679, -0.799	<0.001
ptfsm6cla1a_2019	School FSM proportion	-0.0209	0.0441	-0.1074, 0.0656	0.636
EAL1	Pupil EAL flag	0.956	0.348	0.274, 1.637	0.006
ptealgrp2_2019	School EAL proportion	0.03517	0.0152	0.00535, 0.0650	0.021
academy	School academy status	-0.666	0.573	-1.788, 0.456	0.245
IDACIScore_19_SPR24	Individual IDACI score	-5.08	1.16	-7.34, -2.81	<0.001
IDACI2	School IDACI quintile 2	-0.354	0.723	-1.771, 1.063	0.625
IDACI3	School IDACI quintile 3	-0.84	0.796	-2.400, 0.721	0.292

IDACI4	School IDACI quintile 4	0.43	0.879	-1.293, 2.153	0.625
IDACI5	School IDACI quintile 5	-3.89	1.28	-6.40, -1.38	0.002
ofsted_ratingGood	School Ofsted Good	0.21	0.729	-1.218, 1.638	0.773
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.841	1.28	-1.661, 3.343	0.510
ur_3catUrban city/town	Urban classification: city/town	1.466	0.703	0.0883, 2.845	0.037
ur_3catRural	Urban classification: rural	1.866	0.932	0.0395, 3.693	0.045

Table C.16.3: Secondary Analysis (General Self-Confidence). N=10269.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		20.68	16.2	-11.10, 52.50	0.203
rawSSCbase_scale	General self-confidence baseline	0.6018	0.0115	0.579, 0.624	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**0.3716**	**0.197**	**-0.0144, 0.757**	**0.059**
ks2aps_2019	2019 School average KS2 attainment of intake	-0.531	0.591	-1.689, 0.626	0.368
ptpriorlo_2019	School low prior attainment proportion	-0.00456	0.0604	-0.12298, 0.11386	0.940
ptpriorhi_2019	School high prior attainment proportion	0.0319	0.0351	-0.0369, 0.1007	0.363
N_pupils_byschool_schrep	Number of pupils in Y7	0.00165	0.00182	-0.00191, 0.00522	0.364
EVERFSM_6_P_SPR241	Pupil Ever FSM	0.759	0.180	0.406, 1.111	<0.001
ptfsm6cla1a_2019	School FSM proportion	-0.0244	0.0198	-0.0632, 0.0144	0.217
EAL1	Pupil EAL flag	-0.4145	0.177	-0.7620, -0.0669	0.019
ptealgrp2_2019	School EAL proportion	-0.01804	0.00651	-0.03079, -0.00528	0.006
academy	School academy status	0.234	0.228	-0.213, 0.680	0.305
IDACIScore_19_SPR24	Individual IDACI score	1.611	0.719	0.202, 3.020	0.025
IDACI2	School IDACI quintile 2	-0.107	0.317	-0.727, 0.514	0.736
IDACI3	School IDACI quintile 3	-0.116	0.262	-0.630, 0.398	0.658
IDACI4	School IDACI quintile 4	-0.374	0.418	-1.193, 0.445	0.371
IDACI5	School IDACI quintile 5	0.0792	0.668	-1.230, 1.388	0.906
ofsted_ratingGood	School Ofsted Good	0.616	0.246	0.134, 1.099	0.012
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.446	0.441	-0.419, 1.311	0.312

ur_3catUrban city/town	Urban classification: city/town	-0.271	0.222	-0.705, 0.164	0.222
ur_3catRural	Urban classification: rural	-0.6332	0.322	-1.265, -0.00196	0.049

Table C.16.4: Secondary Analysis with Interactions (General Self-Confidence). N=10269.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		17.2	15.7	-13.5, 48.0	0.272
rawSSCbase_scale	General self-confidence baseline	0.5316	0.0132	0.5058, 0.5575	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**0.4409**	**0.226**	**-0.0025, 0.884**	**0.051**
KS2_MATMRK_lo	**Pupil in the lower tertile of KS2 Maths scores**	**1.002**	**0.186**	**0.637, 1.37**	**<0.001**
KS2_MATMRK_hi	**Pupil in the upper tertile of KS2 Maths scores**	**-1.427**	**0.163**	**-1.746, -1.108**	**<0.001**
mixedAttain1:KS2_MATMRK_lo	**Interaction: Treatment x Pupil in Lower Tertile KS2 Maths**	**0.298**	**0.250**	**-0.191, 0.787**	**0.233**
mixedAttain1:KS2_MATMRK_hi	**Interaction: Treatment x Pupil in Upper Tertile KS2 Maths**	**-0.4864**	**0.229**	**-0.935, -0.0378**	**0.034**
ks2aps_2019	2019 School average KS2 attainment of intake	-0.366	0.572	-1.487, 0.756	0.523
ptpriorlo_2019	School low prior attainment proportion	0.0085	0.0597	-0.1085, 0.126	0.887
ptpriorhi_2019	School high prior attainment proportion	0.0343	0.0333	-0.0310, 0.0996	0.303
N_pupils_byschool_schrep	Number of pupils in Y7	0.00186	0.00190	-0.00186, 0.00558	0.326
EVERFSM_6_P_SPR241	Pupil Ever FSM	0.454	0.171	0.119, 0.790	0.008
ptfsm6cla1a_2019	School FSM proportion	-0.021	0.0188	-0.0579, 0.0160	0.266
EAL1	Pupil EAL flag	-0.48	0.167	-0.808, -0.151	0.004
ptealgrp2_2019	School EAL proportion	-0.01934	0.00653	-0.0322, -0.00654	0.003
academy	School academy status	0.224	0.228	-0.223, 0.671	0.326
IDACIScore_19_SPR24	Individual IDACI score	0.691	0.725	-0.729, 2.11	0.340
IDACI2	School IDACI quintile 2	-0.000741	0.300	-0.588, 0.587	0.998
IDACI3	School IDACI quintile 3	-0.0798	0.262	-0.594, 0.434	0.761

IDACI4	School IDACI quintile 4	-0.351	0.424	-1.182, 0.480	0.408
IDACI5	School IDACI quintile 5	0.0367	0.658	-1.253, 1.326	0.956
ofsted_ratingGood	School Ofsted Good	0.528	0.227	0.083, 0.973	0.020
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.282	0.407	-0.515, 1.08	0.488
ur_3catUrban city/town	Urban classification: city/town	-0.4656	0.227	-0.910, -0.0209	0.040
ur_3catRural	Urban classification: rural	-0.853	0.311	-1.462, -0.244	0.006

Table C.16.5: Secondary Analysis (Mathematics Self-Confidence). N=9076.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		19.8	18.6	-16.7, 56.3	0.287
rawSSCbase_scale	General self-confidence baseline	0.658	0.0151	0.628, 0.687	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**0.438**	**0.228**	**-0.00878, 0.884**	**0.055**
ks2aps_2019	2019 School average KS2 attainment of intake	-0.483	0.686	-1.83, 0.862	0.482
ptpriorlo_2019	School low prior attainment proportion	0.0118	0.0742	-0.134, 0.157	0.874
ptpriorhi_2019	School high prior attainment proportion	0.0482	0.0415	-0.0331, 0.130	0.245
N_pupils_byschool_schrep	Number of pupils in Y7	-0.00234	0.00259	-0.00742, 0.00273	0.366
EVERFSM_6_P_SPR241	Pupil Ever FSM	0.995	0.245	0.515, 1.48	<0.001
ptfsm6cla1a_2019	School FSM proportion	-0.0373	0.0248	-0.0859, 0.0113	0.133
EAL1	Pupil EAL flag	-0.359	0.186	-0.724, 0.00544	0.054
ptealgrp2_2019	School EAL proportion	-0.0159	0.00901	-0.0335, 0.00178	0.078
academy	School academy status	0.297	0.281	-0.253, 0.847	0.290
IDACIScore_19_SPR24	Individual IDACI score	2.21	0.758	0.718, 3.69	0.004
IDACI2	School IDACI quintile 2	-0.076	0.372	-0.806, 0.654	0.838
IDACI3	School IDACI quintile 3	0.182	0.336	-0.477, 0.841	0.588
IDACI4	School IDACI quintile 4	-0.497	0.527	-1.53, 0.536	0.346
IDACI5	School IDACI quintile 5	0.374	0.871	-1.33, 2.08	0.668
ofsted_ratingGood	School Ofsted Good	0.682	0.271	0.151, 1.21	0.012
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.468	0.455	-0.423, 1.36	0.304

ur_3catUrban city/town	Urban classification: city/town	-0.574	0.292	-1.146, -0.00147	0.049
ur_3catRural	Urban classification: rural	-1.06	0.392	-1.83, -0.292	0.007

Table C.16.6: Secondary Analysis with Interactions (Mathematics Self-Confidence). N=8807.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		13.7	19.2	-24.1, 51.4	0.478
rawSSCbase_scale	General self-confidence baseline	0.558	0.0159	0.527, 0.589	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**0.484**	**0.294**	**-0.0924, 1.06**	**0.100**
KS2_MATMRK_lo	**Pupil in the lower tertile of KS2 Maths scores**	**0.978**	**0.247**	**0.495, 1.46**	**<0.001**
KS2_MATMRK_hi	**Pupil in the upper tertile of KS2 Maths scores**	**-1.99**	**0.200**	**-2.39, -1.60**	**<0.001**
mixedAttain1:KS2_MATMRK_lo	**Interaction: Treatment x Pupil in Lower Tertile KS2 Maths**	**0.779**	**0.302**	**0.186, 1.37**	**0.010**
mixedAttain1:KS2_MATMRK_hi	**Interaction: Treatment x Pupil in Upper Tertile KS2 Maths**	**-0.515**	**0.267**	**-1.038, 0.0078**	**0.054**
ks2aps_2019	2019 School average KS2 attainment of intake	-0.190	0.708	-1.58, 1.20	0.788
ptpriorlo_2019	School low prior attainment proportion	0.0345	0.0748	-0.112, 0.181	0.644
ptpriorhi_2019	School high prior attainment proportion	0.0450	0.0419	-0.0372, 0.127	0.283
N_pupils_byschool_schrep	Number of pupils in Y7	-0.00129	0.00278	-0.00674, 0.00416	0.643
EVERFSM_6_P_SPR241	Pupil Ever FSM	0.547	0.240	0.0771, 1.02	0.022
ptfsm6cla1a_2019	School FSM proportion	-0.0419	0.0246	-0.0901, 0.00619	0.088
EAL1	Pupil EAL flag	-0.563	0.179	-0.913, -0.212	0.002
ptealgrp2_2019	School EAL proportion	-0.01699	0.00948	-0.0356, 0.00160	0.073
academy	School academy status	0.350	0.298	-0.234, 0.933	0.240
IDACIScore_19_SPR24	Individual IDACI score	1.54	0.795	-0.0169, 3.10	0.053
IDACI2	School IDACI quintile 2	0.103	0.374	-0.630, 0.836	0.783
IDACI3	School IDACI quintile 3	0.315	0.353	-0.376, 1.01	0.372

IDACI4	School IDACI quintile 4	-0.346	0.560	-1.44, 0.752	0.537
IDACI5	School IDACI quintile 5	0.566	0.875	-1.15, 2.28	0.518
ofsted_ratingGood	School Ofsted Good	0.581	0.263	0.0659, 1.10	0.027
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.177	0.424	-0.654, 1.01	0.676
ur_3catUrban city/town	Urban classification: city/town	-0.905	0.296	-1.49, -0.324	0.002
ur_3catRural	Urban classification: rural	-1.39	0.390	-2.16, -0.626	<0.001

Table C.16.7: FSM Sub-Group Analysis (Mathematics Attainment Outcome: PTM13). N=2883.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		28.5	28.9	-28.2, 85.3	0.324
KS2_MATMRK	KS2 Maths score	0.31957	0.00862	0.303, 0.336	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**0.474**	**0.577**	**-0.658, 1.61**	**0.412**
ks2aps_2019	2019 School average KS2 attainment of intake	1.9813	1.08	-0.129, 4.09	0.066
ptpriorlo_2019	School low prior attainment proportion	0.0412	0.133	-0.220, 0.303	0.757
ptpriorhi_2019	School high prior attainment proportion	-0.2151	0.0809	-0.374, -0.056	0.008
N_pupils_byschool_schrep	Number of pupils in Y7	-0.00017	0.00427	-0.00855, 0.00821	0.968
ptfsm6cla1a_2019	School FSM proportion	-0.0405	0.0417	-0.122, 0.041	0.332
EAL1	Pupil EAL flag	1.51	0.421	0.684, 2.34	<0.001
ptealgrp2_2019	School EAL proportion	0.02199	0.0168	-0.0110, 0.0550	0.191
academy	School academy status	-0.799	0.613	-2.00, 0.403	0.193
IDACIScore_19_SPR24	Individual IDACI score	-1.25	2.02	-5.22, 2.72	0.537
IDACI2	School IDACI quintile 2	-0.646	1.11	-2.81, 1.52	0.559
IDACI3	School IDACI quintile 3	-0.172	0.939	-2.01, 1.67	0.854
IDACI4	School IDACI quintile 4	0.141	0.936	-1.70, 1.98	0.881
IDACI5	School IDACI quintile 5	-3.09	1.24	-5.53, -0.652	0.013
ofsted_ratingGood	School Ofsted Good	-0.534	0.790	-2.08, 1.02	0.499
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	-0.786	1.77	-4.26, 2.69	0.657
ur_3catUrban city/town	Urban classification: city/town	0.87	0.843	-0.782, 2.52	0.302

ur_3catRural	Urban classification: rural	2.68	1.09	0.557, 4.81	0.013

Table C.16.8: Analysis with FSM and Mixed Attainment Interaction on full dataset (Mathematics Attainment Outcome: PTM13). N=14877.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		16.5	37.5	-57.0, 89.9	0.661
KS2_MATMRK	KS2 Maths score	0.37488	0.00702	0.361, 0.389	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**-0.862**	**0.583**	**-2.00, 0.280**	**0.139**
EVERFSM_6_P_SPR241	**Pupil Ever FSM**	**-1.832**	**0.318**	**-2.456, -1.209**	**<0.001**
mixedAttain1:EVERFSM_6_P_SPR241	**Interaction: Treatment x Pupil Ever FSM**	**0.969**	**0.473**	**0.0425, 1.895**	**0.040**
ks2aps_2019	2019 School average KS2 attainment of intake	2.195	1.374	-0.497, 4.888	0.110
ptpriorlo_2019	School low prior attainment proportion	0.124	0.152	-0.173, 0.422	0.413
ptpriorhi_2019	School high prior attainment proportion	-0.16172	0.0852	-0.329, 0.00532	0.058
N_pupils_byschool_schrep	Number of pupils in Y7	0.00287	0.00517	-0.00727, 0.0130	0.579
ptfsm6cla1a_2019	School FSM proportion	-0.0188	0.0450	-0.107, 0.0695	0.677
EAL1	Pupil EAL flag	1.032	0.360	0.326, 1.738	0.004
ptealgrp2_2019	School EAL proportion	0.03771	0.0153	0.00769, 0.0677	0.014
academy	School academy status	-0.633	0.573	-1.757, 0.491	0.270
IDACIScore_19_SPR24	Individual IDACI score	-5.18	1.13	-7.41, -2.96	<0.001
IDACI2	School IDACI quintile 2	-0.461	0.739	-1.909, 0.987	0.533
IDACI3	School IDACI quintile 3	-0.997	0.808	-2.580, 0.586	0.217
IDACI4	School IDACI quintile 4	0.298	0.918	-1.502, 2.097	0.746

IDACI5	School IDACI quintile 5	-4	1.29	-6.53, -1.46	0.002
ofsted_ratingGood	School Ofsted Good	0.141	0.722	-1.275, 1.556	0.846
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	0.859	1.327	-1.743, 3.461	0.518
ur_3catUrban city/town	Urban classification: city/town	1.4144	0.717	0.00972, 2.819	0.048
ur_3catRural	Urban classification: rural	1.7966	0.959	-0.0832, 3.676	0.061

Table C.16.9: Sensitivity Analysis Primary Outcome: Matched Pupil Sample. N=9898.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		55.4	50.4	-43.5, 154.3	0.272
KS2_MATMRK	KS2 Maths score	0.36649	0.00585	0.35502, 0.37796	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**-0.215**	**0.539**	**-1.271, 0.841**	**0.69**
ks2aps_2019	2019 School average KS2 attainment of intake	0.798	1.85	-2.828, 4.425	0.666
ptpriorlo_2019	School low prior attainment proportion	-0.119	0.176	-0.464, 0.226	0.498
ptpriorhi_2019	School high prior attainment proportion	-0.0554	0.1077	-0.2665, 0.1557	0.607
N_pupils_byschool_schrep	Number of pupils in Y7	-0.0017	0.00553	-0.01255, 0.00914	0.758
EVERFSM_6_P_SPR241	Pupil Ever FSM	-1.629	0.204	-2.028, -1.230	<0.001
ptfsm6cla1a_2019	School FSM proportion	0.00991	0.04384	-0.07603, 0.09585	0.821
EAL1	Pupil EAL flag	1.081	0.345	0.405, 1.757	0.002
ptealgrp2_2019	School EAL proportion	0.0136	0.0155	-0.0168, 0.0441	0.38
academy	School academy status	-0.762	0.584	-1.906, 0.383	0.192
IDACIScore_19_SPR24	Individual IDACI score	-4.2	1.14	-6.44, -1.95	<0.001
IDACI2	School IDACI quintile 2	-0.255	0.847	-1.916, 1.405	0.763
IDACI3	School IDACI quintile 3	-0.471	0.845	-2.128, 1.186	0.577
IDACI4	School IDACI quintile 4	0.379	0.92	-1.424, 2.183	0.68
IDACI5	School IDACI quintile 5	-0.453	1.72	-3.825, 2.919	0.792
ofsted_ratingGood	School Ofsted Good	-0.131	0.633	-1.373, 1.110	0.836
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	-0.204	0.998	-2.159, 1.752	0.838

ur_3catUrban city/town	Urban classification: city/town	1.129	0.642	-0.130, 2.387	0.079
ur_3catRural	Urban classification: rural	1.699	0.962	-0.187, 3.585	0.077

Table C.16.10: Sensitivity Analysis Primary Outcome with Interactions: Matched Pupil Sample. N=9898.

Variable	Description	Coef.	SE	95% CI	p-value
Intercept		53	52	-48.8, 154.9	0.307
KS2_MATMRK	KS2 Maths score	0.31964	0.00759	0.30477, 0.33451	<0.001
mixedAttain1	**Treatment allocation (Mixed Attainment)**	**-0.452**	**0.633**	**-1.693, 0.790**	**0.476**
KS2_MATMRK_lo	**Pupil in the lower tertile of KS2 Maths scores**	**0.681**	**0.369**	**-0.043, 1.405**	**0.065**
KS2_MATMRK_hi	**Pupil in the upper tertile of KS2 Maths scores**	**5.645**	**0.347**	**4.964, 6.326**	**<0.001**
mixedAttain1:KS2_MATMRK_lo	**Interaction: Treatment x Pupil in Lower Tertile KS2 Maths**	**1.22**	**0.49**	**0.259, 2.180**	**0.013**
mixedAttain1:KS2_MATMRK_hi	**Interaction: Treatment x Pupil in Upper Tertile KS2 Maths**	**-0.463**	**0.535**	**-1.511, 0.586**	**0.387**
ks2aps_2019	2019 School average KS2 attainment of intake	0.966	1.907	-2.773, 4.704	0.613
ptpriorlo_2019	School low prior attainment proportion	-0.122	0.18	-0.474, 0.231	0.500
ptpriorhi_2019	School high prior attainment proportion	-0.0803	0.1107	-0.2972, 0.1367	0.468
N_pupils_byschool_schrep	Number of pupils in Y7	-0.00167	0.00555	-0.01254, 0.00921	0.764
EVERFSM_6_P_SPR241	Pupil Ever FSM	-1.52	0.212	-1.936, -1.104	<0.001
ptfsm6cla1a_2019	School FSM proportion	0.00665	0.04367	-0.07894, 0.09225	0.879
EAL1	Pupil EAL flag	0.993	0.317	0.371, 1.615	0.002
ptealgrp2_2019	School EAL proportion	0.0124	0.0157	-0.0184, 0.0433	0.430
academy	School academy status	-0.723	0.589	-1.877, 0.430	0.219
IDACIScore_19_SPR24	Individual IDACI score	-4.26	1.13	-6.48, -2.03	<0.001
IDACI2	School IDACI quintile 2	-0.255	0.83	-1.883, 1.373	0.759
IDACI3	School IDACI quintile 3	-0.36	0.843	-2.011, 1.292	0.669

IDACI4	School IDACI quintile 4	0.444	0.909	-1.338, 2.227	0.625
IDACI5	School IDACI quintile 5	-0.305	1.722	-3.681, 3.071	0.859
ofsted_ratingGood	School Ofsted Good	-0.0753	0.6339	-1.3179, 1.1672	0.905
ofsted_ratingRequires improvement	School Ofsted Requires Improvement	-0.121	0.956	-1.995, 1.752	0.899
ur_3catUrban city/town	Urban classification: city/town	1.1916	0.634	-0.0512, 2.4343	0.060
ur_3catRural	Urban classification: rural	1.7479	0.9379	-0.0906, 3.5864	0.062

Table C.16.11: Summary of Quantile analysis: Coefficients, SEs & p-values for treatment effect

Percentile	Coef.	SE	p-value
10^{th}	-0.667	0.827	0.420
20^{th}	-0.363	0.658	0.581
25^{th}	-0.506	0.626	0.419
30^{th}	-0.377	0.632	0.551
40^{th}	-0.399	0.545	0.464
50^{th}	-0.409	0.514	0.426
60^{th}	-0.472	0.481	0.326
70^{th}	-0.508	0.517	0.326
75^{th}	-0.513	0.531	0.334
80^{th}	-0.552	0.586	0.346
90^{th}	-0.644	0.521	0.217

Table C.16.12: Summary of analysis examining school self-administration of outcome testing

Variable	Description	Coef.	SE	95% CI	p-value
Sub-Group Analysis for School Self-Administering Outcome Testing: N=4728 (29 schools)					
Intercept		-136.6	56.0	-246.3, -26.8	.015
KS2_MATMRK	KS2 Maths score	0.389	0.010	0.369, 0.409	<.001
mixedAttain1	**Treatment allocation Mixed Attainment**	**0.101**	**0.873**	**-1.61, 1.81**	**.907**
Interaction Analysis on full dataset: N=14877					
Intercept		12.0	41.4	-69.2, 93.2	.772
KS2_MATMRK	KS2 Maths score	0.375	0.007	0.361, 0.389	<.001
mixedAttain1	**Treatment allocation Mixed Attainment**	**-0.423**	**0.687**	**-1.769, 0.923**	**.538**
flag_selfadmin	**Self-Administered Flag**	**0.302**	**0.837**	**-1.34, 1.94**	**.718**
mixedAttain1:flag_selfadmin	**Interaction: Treatment x Self-Administered**	**-0.672**	**1.28**	**-3.17, 1.83**	**.598**

Appendix C.17: Mediation Analysis

For all pupils

For the opportunity to learn model (OTL), the mediation analysis examined whether the effect of mixed classes on test scores was mediated through OTL. (See Table F.1.) The results show that mixed classes are associated with greater opportunity to learn by 0.05 units (CI 0.00, 0.09), but that opportunity to learn was associated with 1.6 units lower test scores (CI -2.2, -1.1) when controlling for mixed/setting. The total effect of mixed attainment classes in this model was -0.244 (CI -1.4, 0.9), which is insignificant as in the earlier models but with a lower point estimate. Of this, the indirect effect is -0.077 (CI -0.167, 0.014): 36% of the (negative) effect of mixed classes on scores operates through OTL, but this is not statistically insignificant. This provides some weak evidence to suggest that mixed classes may indirectly reduce test scores by increasing OTL, but the overall effect is too small for this finding to provide conclusive evidence of a mediation effect.

Table C.17.1: Mediation analysis summary by Opportunity to Learn (OTL) for all pupils (N=11248)

Effect	**Estimate**	**SE**	**Upper 95% CI**	**Lower 95% CI**	***p*-value**
a: Mixed ® OTL	0.047	0.022	0.003	0.091	0.035
b: OTL ® PTM13	-1.626	0.277	-2.169	-1.083	<0.0001
c: Direct: Mixed ® PTM13	0.168	0.569	-1.283	0.948	0.768
a*b: Indirect	-0.077	0.046	-0.167	0.014	0.097
Total effect	-0.244	0.571	-1.364	0.875	0.669

For the teacher quality model (TQ), the mediation analysis examined whether the effect of mixed classes on test scores was mediated through TQ. (See Table F.2.) The results show that mixed classes are associated with greater teacher quality by 0.08 units (CI -0.05, 0.2), but that teacher quality is associated with 1.4 units lower test scores (CI -1.6, -1.1) when controlling for mixed/setting. The total effect of mixed attainment classes in this model was -0.271 (CI -1.4, 0.9), which is insignificant as in the earlier models but with a lower point estimate. Of this, the indirect effect is -0.103 (CI -0.26, 0.05): 38% of the (negative) effect of mixed classes on scores operates through TQ, but this is not statistically insignificant. This provides some weak evidence to suggest that mixed classes may indirectly reduce test scores by being associated with higher TQ, but the overall effect is too small for this finding to provide conclusive evidence of a mediation effect.

Table C.17.2: Mediation analysis summary by Teacher Quality (TQ) for all pupils (N=11248)

Effect	**Estimate**	**SE**	**Upper 95% CI**	**Lower 95% CI**	***p*-value**
a: Mixed ® TQ	0.076	0.063	-0.047	0.198	0.228
b: TQ ® PTM13	-1.368	0.124	-1.610	-1.126	<0.0001

c: Direct: Mixed ® PTM13	-0.168	0.588	-1.320	0.984	0.775
a*b: Indirect	-0.103	0.080	-0.260	0.054	0.197
Total effect	-0.271	0.593	-1.434	0.892	0.648

For low-attaining pupils

For the opportunity to learn model (OTL), the mediation analysis examined whether the effect of mixed classes on test scores was mediated through OTL. (See Table F.3.) The results show that mixed classes are associated with greater opportunity to learn by 0.02 units (CI = -0.03, 0.08), but that opportunity to learn was associated with 0.6 units lower test scores (CI -1.2, -0.09) when controlling for mixed/setting. The total effect of mixed attainment classes in this model was 1.11 (CI -0.03, 2.25), which is positive and just significant as in the earlier models. Of this, the indirect effect is -0.01 (CI -0.07, 0.05): <1% of the (positive) effect of mixed classes on scores operates through OTL. This suggests that OTL has no part to play in why mixed classes may increase test scores for lower attainers.

Table C.17.3: Mediation analysis summary by Opportunity to Learn (OTL) for pupils with low prior attainment (N=3127)

Effect	**Estimate**	**SE**	**Upper 95% CI**	**Lower 95% CI**	***p*-value**
a: Mixed ® OTL	0.024	0.030	-0.034	0.082	0.422
b: OTL ® PTM13	-0.570	0.337	-1.231	0.091	0.091
c: Direct: Mixed ® PTM13	1.119	0.581	-0.020	2.257	0.054
a*b: Indirect	-0.014	0.031	-0.073	0.046	0.655
Total effect	1.105	0.582	-0.035	2.245	0.057

For the teacher quality model (TQ), the mediation analysis examined whether the effect of mixed classes on test scores was mediated through TQ. (See Table F.4.) The results show that mixed classes are associated with greater teacher quality by 0.10 units (CI -0.04, 0.25), but that teacher quality is associated with 0.7 units lower test scores (CI -1.13, -0.38) when controlling for mixed/setting. The total effect of mixed attainment classes in this model was 1.04 (CI -0.09, 2.17), which is positive and just significant as in the earlier models. Of this, the indirect effect is -0.08 (CI -0.19, 0.03): 7.7% of the (negative) effect of mixed classes on scores operates through TQ, but this is not statistically significant.

Table C.17.4: Mediation analysis summary by Teacher Quality (TQ) for pupils with low prior attainment (N=3127)

Effect	**Estimate**	**SE**	**Upper 95% CI**	**Lower 95% CI**	***p*-value**
a: Mixed ® TQ	0.103	0.075	-0.044	0.250	0.170
b: TQ ® PTM13	-0.756	0.191	-1.131	-0.382	0.000
c: Direct: Mixed ® PTM13	1.119	0.574	-0.006	2.244	0.051
a*b: Indirect	-0.078	0.055	-0.186	0.030	0.155
Total effect	1.041	0.577	-0.089	2.171	0.071

Appendix C.18: Missing Data

Table C.18.1: Additional analyses: Imputation

Outcome	Total n	Main effect coefficient (95% CI)	p-value
Primary analysis	14877	-0.670 (-1.769, 0.428)	0.232
Imputation (imputing only main outcome: PTM13)	17726	-0.596 (-1.558, 0.366)	0.225

Describing: all original data

Note: zeros are structural because they indicate no missingness.

P2SAS: 26% missing. This includes 8/104 whole schools.

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606  5824     0.258

## [1] 8

## N.Obs: 22606
```

rawSSCend_scale: 34% missing. This includes 9/104 whole schools.

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606  7711     0.341

## [1] 9

## N.Obs: 22606
```

rawMSCend_scale: 34% missing. This includes 9/104 whole schools.

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606  7751     0.343

## [1] 9

## N.Obs: 22606
```

Models:

Primary: 6649 observations deleted due to missingness (29%)

```
##   (6649 observations deleted due to missingness)

## N.Obs: 22606
```

Secondary SSC: 11141 observations deleted due to missingness (49%)

```
##   (11141 observations deleted due to missingness)

## N.Obs: 22606
```

Secondary MSC: 12497 observations deleted due to missingness (55%)

```
##   (12497 observations deleted due to missingness)

## N.Obs: 22606
```

Covariates separately:

KS2_MATMRK:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606  1438    0.0636

## N.Obs: 22606
```

rawSSCbase_scale:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606  6863     0.304

## N.Obs: 22606
```

rawMSCbase_scale:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606  8636     0.382

## N.Obs: 22606
```

ks2aps_2019:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606     0         0

## N.Obs: 22606
```

ptpriorlo_2019:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606     0         0

## N.Obs: 22606
```

ptpriorhi_2019:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606     0         0

## N.Obs: 22606
```

N_pupils_byschool_schrep:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606     0         0

## N.Obs: 22606
```

ptfsm6cla1a_2019:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606     0         0

## N.Obs: 22606
```

ptealgrp2_2019:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606     0         0

## N.Obs: 22606
```

IDACIscore_19_SPR24:

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 22606   545    0.0241

## N.Obs: 22606
```

EAL:

```
## # A tibble: 1 × 3
##   total missing missing_pct
```

```
##   <int>  <int>      <dbl>
## 1 22606    703       3.11

## N.Obs: 22606
```

academy:

```
## # A tibble: 1 × 3
##   total missing missing_pct
##   <int>  <int>      <dbl>
## 1 22606      0          0

## N.Obs: 22606
```

IDACI:

```
## # A tibble: 1 × 3
##   total missing missing_pct
##   <int>  <int>      <dbl>
## 1 22606      0          0

## N.Obs: 22606
```

ofsted_rating:

```
## # A tibble: 1 × 3
##   total missing missing_pct
##   <int>  <int>      <dbl>
## 1 22606      0          0

## N.Obs: 22606
```

ur_3cat:

```
## # A tibble: 1 × 3
##   total missing missing_pct
##   <int>  <int>      <dbl>
## 1 22606      0          0

## N.Obs: 22606
```

EVERFSM_6_P_SPR24:

```
## # A tibble: 1 × 3
##   total missing missing_pct
##   <int>  <int>      <dbl>
## 1 22606    478       2.11

## N.Obs: 22606
```

In sum, everything 0% apart from:

KS2_MATMRK: 6%
rawSSCbase_scale: 30%
rawMSCbase_scale: 38%

IDACIScore_19_SPR24: 2%
EAL: 3%
EVERFSM_6_P_SPR24: 2%

Missingness work for Primary analysis (P2SAS)

Describing and subsetting data net dropped schools

Dropping the dropped schools. Both because of missing data and because of weights calculations.

Dropped because of weights: 9 schools.
Remaining sample: 5773 + 14877 = 20650
P2SAS: 24% still missing.

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 20650  5005     0.242

## N.Obs: 20650
```

Dropped because of missing P2SAS: 6 schools.
Remaining sample: 19179
P2SAS: 18% still missing not including entirely dropped/missing schools.

```
## # A tibble: 1 × 10
##       n  miss miss_rate
##   <int> <int>     <dbl>
## 1 19179  3534     0.184

## N.Obs: 19179
```

(Note: the pattern for the secondary analyses is very similar.)

Predict missingness

Make a binary variable indicating missing P2SAS. 3534 observations.

Use logistic regression to predict missingness using variables from the imbalance testing. (Note: Average_att dropped due to colinearity with ks2aps_2019 / ks2_aps_2018 /ks2aps_2017. KS2_MATMRK_lo/hi not included as already including KS2_MATMRK.)

Sample is smaller due to a small amount of missingness in the covariates.

	(1)	(2)
(Intercept)	13.79	13.79
	(3.75)	(6.13)
	[6.43, 21.14]	[1.78, 25.79]
	(<0.001)	(0.024)

	(1)	(2)
ks2aps_2019	-0.487	-0.4875
	(0.150)	(0.2238)
	[-0.781, -0.194]	[-0.9262, -0.0488]
	(0.001)	(0.029)
ks2aps_2018	-0.00698	-0.00698
	(0.05555)	(0.08641)
	[-0.11580, 0.10199]	[-0.17633, 0.16237]
	(0.900)	(0.936)
ks2aps_2017	-0.0354	-0.0354
	(0.0436)	(0.0619)
	[-0.1210, 0.0499]	[-0.1567, 0.0858]
	(0.416)	(0.567)
ptpriorlo_2019	-0.0545	-0.05450
	(0.0145)	(0.02305)
	[-0.0831, -0.0260]	[-0.09969, -0.00931]
	(<0.001)	(0.018)
ptpriorhi_2019	0.01259	0.0126
	(0.00852)	(0.0130)
	[-0.00412, 0.02929]	[-0.0130, 0.0381]
	(0.140)	(0.334)
N_pupils_byschool_schrep	0.002621	0.002621
	(0.000484)	(0.000699)
	[0.001674, 0.003572]	[0.001250, 0.003991]
	(<0.001)	(<0.001)
ptfsm6cla1a_2019	-0.01091	-0.01091
	(0.00484)	(0.00678)
	[-0.02041, -0.00143]	[-0.02420, 0.00237]
	(0.024)	(0.107)
ptealgrp2_2019	-0.00155	-0.00155
	(0.00153)	(0.00242)
	[-0.00456, 0.00143]	[-0.00630, 0.00320]
	(0.310)	(0.522)
academy	-0.0640	-0.0640
	(0.0519)	(0.0710)

	(1)	(2)
	[-0.1654, 0.0380]	[-0.2032, 0.0752]
	(0.218)	(0.368)
factor(IDACI)2	0.0373	0.0373
	(0.0760)	(0.1298)
	[-0.1118, 0.1862]	[-0.2172, 0.2917]
	(0.624)	(0.774)
factor(IDACI)3	-0.0163	-0.0163
	(0.0779)	(0.1353)
	[-0.1692, 0.1364]	[-0.2815, 0.2488]
	(0.834)	(0.904)
factor(IDACI)4	0.0463	0.0463
	(0.0967)	(0.1345)
	[-0.1435, 0.2357]	[-0.2173, 0.3099]
	(0.632)	(0.731)
factor(IDACI)5	0.32228	0.322
	(0.16313)	(0.226)
	[0.00193, 0.64149]	[-0.121, 0.766]
	(0.048)	(0.154)
factor(ofsted_rating)Good	0.00072	0.00072
	(0.05871)	(0.10900)
	[-0.11412, 0.11604]	[-0.21292, 0.21436]
	(0.990)	(0.995)
factor(ofsted_rating)Requires improvement	-0.2648	-0.26481
	(0.1011)	(0.13930)
	[-0.4639, -0.0675]	[-0.53783, 0.00821]
	(0.009)	(0.057)
Urban	0.16067	0.161
	(0.07913)	(0.117)
	[0.00671, 0.31698]	[-0.069, 0.390]
	(0.042)	(0.170)
KS2_MATMRK	-0.012439	-0.012439
	(0.000791)	(0.000932)
	[-0.013989, -0.010887]	[-0.014266, -0.010613]
	(<0.001)	(<0.001)
EVERFSM_6_P_SPR241	0.6955	0.6955

	(1)	(2)
	(0.0482)	(0.0518)
	[0.6009, 0.7899]	[0.5939, 0.7971]
	(<0.001)	(<0.001)
EAL1	-0.4229	-0.423
	(0.0697)	(0.080)
	[-0.5609, -0.2875]	[-0.580, -0.266]
	(<0.001)	(<0.001)
IDACIScore_19_SPR24	0.603	0.603
	(0.224)	(0.220)
	[0.161, 1.041]	[0.171, 1.034]
	(0.007)	(0.006)
Num.Obs.	17726	17726
Std.Errors	iid	CL

There is some predictability from some variables.

Now find the r-squared.

```
## # R2 for Logistic Regression
##   Tjur's R2: 0.056

##N.Obs: 17726
```

0.056

Re-do the weights

Make a new P2SAS variable with no NAs.

Find new complete cases.

New sample N=17726, meaning 1453 dropped because of covariates (not P2SAS) missing.

Make weights.

Add weights to data.

Implement MICE

Only predicting P2SAS, and using only the variables from the imbalance testing.

Run imputation using 20 imputed datasets, maxit = 5, method is pmm.

Combine datasets, run on our main Primary analysis model, and summarise results (with clustered standard errors):

```
## Multiple imputation results:
## Call: miceadds::pool_mi(qhat = betas, u = vars)
##                              results        se        t
```

```
## (Intercept)                    32.1500873093 43.850624322  0.73317285
## KS2_MATMRK                      0.3762546019  0.005629840 66.83220396
## mixedAttain1                   -0.6098061784  0.492987735 -1.23696014
## ks2aps_2019                     1.6258830975  1.626656103  0.99952479
## ptpriorlo_2019                  0.0511630647  0.162001233  0.31581898
## ptpriorhi_2019                 -0.1186062813  0.100193453 -1.18377276
## N_pupils_byschool_schrep        0.0007617408  0.004977022  0.15305152
## EVERFSM_6_P_SPR241             -1.5412943677  0.188186336 -8.19025653
## ptfsm6cla1a_2019                0.0184548893  0.043776086  0.42157468
## EAL1                            1.2197376289  0.306023496  3.98576463
## ptealgrp2_2019                  0.0271053566  0.013492395  2.00893597
## academy                        -0.7665553276  0.560828375 -1.36682693
## IDACIScore_19_SPR24            -5.3816233648  0.960215957 -5.60459689
## IDACI2                         -0.2510165582  0.791192831 -0.31726344
## IDACI3                         -1.4261249445  0.868552913 -1.64195517
## IDACI4                         -0.1055315055  0.883290553 -0.11947542
## IDACI5                         -3.8453659947  1.426521642 -2.69562401
## ofsted_ratingGood              -0.0685779848  0.556047419 -0.12333118
## ofsted_ratingRequires improvement -0.0604197880  1.088124732 -0.05552653
## ur_3catUrban city/town          1.1023220138  0.703498106  1.56691540
## ur_3catRural                    1.1270503982  0.993150171  1.13482375
##                                            p        (lower       upper)
## (Intercept)                    4.634540e-01 -5.379617e+01 118.09634887
## KS2_MATMRK                     0.000000e+00  3.652197e-01   0.38728945
## mixedAttain1                   2.161037e-01 -1.576052e+00   0.35643921
## ks2aps_2019                    3.175419e-01 -1.562327e+00   4.81409299
## ptpriorlo_2019                 7.521401e-01 -2.663544e-01   0.36868057
## ptpriorhi_2019                 2.365051e-01 -3.149836e-01   0.07777106
## N_pupils_byschool_schrep       8.783578e-01 -8.993100e-03   0.01051658
## EVERFSM_6_P_SPR241             5.943850e-16 -1.910461e+00  -1.17212762
## ptfsm6cla1a_2019               6.733371e-01 -6.734646e-02   0.10425624
## EAL1                           6.757028e-05  6.198959e-01   1.81957936
## ptealgrp2_2019                 4.454615e-02  6.604838e-04   0.05355023
## academy                        1.716800e-01 -1.865761e+00   0.33265006
## IDACIScore_19_SPR24            2.262521e-08 -7.264317e+00  -3.49893002
## IDACI2                         7.510440e-01 -1.801734e+00   1.29970098
## IDACI3                         1.006002e-01 -3.128463e+00   0.27621325
## IDACI4                         9.048989e-01 -1.836769e+00   1.62570598
## IDACI5                         7.029263e-03 -6.641400e+00  -1.04933209
## ofsted_ratingGood              9.018449e-01 -1.158413e+00   1.02125712
## ofsted_ratingRequires improvement 9.557190e-01 -2.193110e+00   2.07227051
## ur_3catUrban city/town         1.171354e-01 -2.765136e-01   2.48115763
## ur_3catRural                   2.564504e-01 -8.194987e-01   3.07359954

## N.Obs: 17726
```

Overall, the effect is slightly attentuated. This makes sense as lower-attainers are more likely to have missing P2SAS and have improved outcomes. But still non-significant.

Now with interactions:

```
## Multiple imputation results:
## Call: miceadds::pool_mi(qhat = betas, u = vars)
##                                   results          se           t
## (Intercept)                 30.0968940974 44.054165457  0.683179304
## KS2_MATMRK                   0.3228172471  0.006847932 47.140840508
## mixedAttain1                -0.6731110561  0.544471269 -1.236265519
## KS2_MATMRK_lo                0.8482227431  0.299171931  2.835235043
## KS2_MATMRK_hi                6.0374148348  0.271391296 22.246162383
## ks2aps_2019                  1.7843072750  1.633849581  1.092087849
## ptpriorlo_2019               0.0556672298  0.161138917  0.345461114
## ptpriorhi_2019              -0.1384674325  0.100575697 -1.376748429
## N_pupils_byschool_schrep     0.0004522463  0.004974654  0.090910109
## EVERFSM_6_P_SPR241          -1.4397293595  0.190721360 -7.548862691
## ptfsm6cla1a_2019             0.0178403272  0.042991440  0.414973936
## EAL1                         1.1373903972  0.294695130  3.859549346
## ptealgrp2_2019               0.0245812896  0.013817924  1.778942316
## academy                     -0.7786343639  0.553277031 -1.407313733
## IDACIScore_19_SPR24         -5.2844953800  0.939131034 -5.627005376
## IDACI2                      -0.1732091371  0.756098530 -0.229082759
## IDACI3                      -1.2950270405  0.849713988 -1.524074051
## IDACI4                       0.0010965114  0.842203867  0.001301955
## IDACI5                      -3.7396548940  1.416447573 -2.640164709
## ofsted_ratingGood            0.0486643986  0.555120654  0.087664543
## ofsted_ratingRequires improvement  0.0455809995  1.057791123  0.043090737
## ur_3catUrban city/town       1.1934953096  0.695440476  1.716171764
## ur_3catRural                 1.1995082265  0.966411984  1.241197591
## mixedAttain1:KS2_MATMRK_lo   1.1133077829  0.389487205  2.858393723
## mixedAttain1:KS2_MATMRK_hi  -0.8308034269  0.510345716 -1.627922801
##                                     p       (lower      upper)
## (Intercept)                 4.944945e-01 -56.248313255 116.44210145
## KS2_MATMRK                  0.000000e+00   0.309393948   0.33624055
## mixedAttain1                2.163619e-01  -1.740264105   0.39404199
## KS2_MATMRK_lo               4.609100e-03   0.261625732   1.43481975
## KS2_MATMRK_hi              1.573317e-102   5.505309937   6.56951973
## ks2aps_2019                 2.747960e-01  -1.418001400   4.98661595
## ptpriorlo_2019              7.297479e-01  -0.260160202   0.37149466
## ptpriorhi_2019              1.685923e-01  -0.335593934   0.05865907
## N_pupils_byschool_schrep    9.275641e-01  -0.009297953   0.01020245
## EVERFSM_6_P_SPR241          7.758692e-14  -1.813851229  -1.06560749
## ptfsm6cla1a_2019            6.781627e-01  -0.066423279   0.10210393
## EAL1                        1.141295e-04   0.559746350   1.71503444
## ptealgrp2_2019              7.525162e-02  -0.002501603   0.05166418
## academy                     1.593349e-01  -1.863039474   0.30577075
## IDACIScore_19_SPR24         2.000080e-08  -7.125890540  -3.44310022
## IDACI2                      8.188048e-01  -1.655143684   1.30872541
## IDACI3                      1.274912e-01  -2.960441938   0.37038786
## IDACI4                      9.989612e-01  -1.649614795   1.65180782
## IDACI5                      8.290590e-03  -6.515946020  -0.96336377
## ofsted_ratingGood           9.301433e-01  -1.039354276   1.13668307
## ofsted_ratingRequires improvement  9.656292e-01  -2.027656743   2.11881874
## ur_3catUrban city/town      8.613152e-02  -0.169547817   2.55653844
```

```
## ur_3catRural                   2.145342e-01  -0.694635749   3.09365220
## mixedAttain1:KS2_MATMRK_lo       4.265223e-03   0.349850157   1.87676541
## mixedAttain1:KS2_MATMRK_hi       1.035632e-01  -1.831146904   0.16954005

## N.Obs: 17726
```

Missingness work for Primary analysis: fully imputed dataset

Make weights for project_data_tidy_2_P2SAS assuming everything is a complete case.

Same method as before.

Primary analysis results:

```
## Multiple imputation results:
## Call: miceadds::pool_mi(qhat = betas, u = vars)
##                                  results          se           t
## (Intercept)                  61.046295693 49.745752746  1.22716599
## KS2_MATMRK                    0.295523971  0.007686768 38.44580553
## mixedAttain1                 -0.439216461  0.627365273 -0.70009687
## ks2aps_2019                   0.781891865  1.849478150  0.42276350
## ptpriorlo_2019               -0.039271550  0.177436495 -0.22132735
## ptpriorhi_2019               -0.058981389  0.122951691 -0.47971190
## N_pupils_byschool_schrep      0.002960165  0.005968800  0.49593967
## EVERFSM_6_P_SPR241           -0.470206750  0.273927088 -1.71653980
## ptfsm6cla1a_2019              0.009370868  0.056958176  0.16452191
## EAL1                          0.592226669  0.363832312  1.62774622
## ptealgrp2_2019                0.027314187  0.017111136  1.59628128
## academy                      -0.966991221  0.621615630 -1.55560957
## IDACIScore_19_SPR24          -4.168187250  1.192077644 -3.49657363
## IDACI2                       -0.209300742  0.892845063 -0.23442000
## IDACI3                       -1.343237057  0.961843937 -1.39652287
## IDACI4                       -0.071285209  1.079158650 -0.06605628
## IDACI5                       -4.177406077  1.744197318 -2.39503067
## ofsted_ratingGood            -0.415615225  0.693844953 -0.59900302
## ofsted_ratingRequires improvement -0.841968975  1.289935555 -0.65272174
## ur_3catUrban city/town        1.319030359  0.768705666  1.71591081
## ur_3catRural                  0.848769923  1.105262617  0.76793507
##                                     p       (lower       upper)
## (Intercept)                  2.203744e-01 -36.705263066 158.79785445
## KS2_MATMRK                   2.949704e-167  0.280429133   0.31061881
## mixedAttain1                 4.846837e-01  -1.676346283   0.79791336
## ks2aps_2019                  6.726581e-01  -2.852228263   4.41601199
## ptpriorlo_2019               8.248961e-01  -0.387587357   0.30904426
## ptpriorhi_2019               6.317112e-01  -0.300740644   0.18277787
## N_pupils_byschool_schrep     6.203657e-01  -0.008794461   0.01471479
## EVERFSM_6_P_SPR241           8.773043e-02  -1.010615162   0.07020166
## ptfsm6cla1a_2019             8.694726e-01  -0.102887533   0.12162927
## EAL1                         1.040118e-01  -0.122060166   1.30651350
## ptealgrp2_2019               1.118526e-01  -0.006407390   0.06103576
## academy                      1.200174e-01  -2.186345044   0.25236260
## IDACIScore_19_SPR24          5.258585e-04  -6.511936006  -1.82443849
## IDACI2                       8.147030e-01  -1.961190919   1.54258943
```

```
## IDACI3                      1.628487e-01  -3.230565732   0.54409162
## IDACI4                      9.473601e-01  -2.191669549   2.04909913
## IDACI5                      1.702464e-02  -7.605137841  -0.74967431
## ofsted_ratingGood            5.495918e-01  -1.780653503   0.94942305
## ofsted_ratingRequires improvement  5.146090e-01  -3.384012903   1.70007495
## ur_3catUrban city/town        8.657683e-02  -0.189954649   2.82801537
## ur_3catRural                 4.427512e-01  -1.320770426   3.01831027

## N.Obs: 19179
```

Even more attenuated but still non-significant.

Now with interactions:

```
## Multiple imputation results:
## Call: miceadds::pool_mi(qhat = betas, u = vars)
##                              results          se          t
## (Intercept)                 60.996167954 49.626342939  1.22910866
## KS2_MATMRK                   0.224923511  0.010636134 21.14711191
## mixedAttain1                -0.535272739  0.655945256 -0.81603264
## KS2_MATMRK_lo               -0.070097355  0.410946660 -0.17057531
## KS2_MATMRK_hi                6.634222378  0.327310971 20.26886649
## ks2aps_2019                  0.921093866  1.843593477  0.49961875
## ptpriorlo_2019              -0.035881436  0.176276742 -0.20355173
## ptpriorhi_2019              -0.079031369  0.122066780 -0.64744371
## N_pupils_byschool_schrep     0.002601231  0.005955051  0.43681085
## EVERFSM_6_P_SPR241          -0.391105453  0.275155781 -1.42139646
## ptfsm6cla1a_2019             0.009904620  0.056287178  0.17596583
## EAL1                         0.510658000  0.353843839  1.44317336
## ptealgrp2_2019               0.024144529  0.017291527  1.39632138
## academy                     -0.995695148  0.611715719 -1.62770894
## IDACIScore_19_SPR24         -4.164404661  1.169515118 -3.56079592
## IDACI2                      -0.136993516  0.861500074 -0.15901742
## IDACI3                      -1.244835535  0.943652499 -1.31916732
## IDACI4                       0.018883994  1.043584781  0.01809531
## IDACI5                      -4.079528405  1.734709291 -2.35170724
## ofsted_ratingGood           -0.298738335  0.690056546 -0.43291863
## ofsted_ratingRequires improvement -0.697068287  1.262324570 -0.55221003
## ur_3catUrban city/town       1.423111117  0.764155821  1.86233106
## ur_3catRural                 0.934376996  1.082287404  0.86333537
## mixedAttain1:KS2_MATMRK_lo   1.150721152  0.582890601  1.97416316
## mixedAttain1:KS2_MATMRK_hi  -0.961145152  0.592498317 -1.62219052
##                                 p        (lower       upper)
## (Intercept)                 2.196373e-01 -36.517218870 158.50955478
## KS2_MATMRK                  2.076555e-71   0.204026569   0.24582045
## mixedAttain1                4.153518e-01  -1.827913128   0.75736765
## KS2_MATMRK_lo               8.646301e-01  -0.877589224   0.73739451
## KS2_MATMRK_hi               6.953295e-75   5.991794341   7.27665041
## ks2aps_2019                 6.175703e-01  -2.701329289   4.54351702
## ptpriorlo_2019              8.387589e-01  -0.381934470   0.31017160
## ptpriorhi_2019              5.177381e-01  -0.319046821   0.16098408
## N_pupils_byschool_schrep    6.626241e-01  -0.009127018   0.01432948
```

```
## EVERFSM_6_P_SPR241           1.568529e-01  -0.933883623  0.15167272
## ptfsm6cla1a_2019             8.604919e-01  -0.101061545  0.12087079
## EAL1                         1.494653e-01  -0.184189089  1.20550509
## ptealgrp2_2019               1.640055e-01  -0.009931104  0.05822016
## academy                      1.038182e-01  -2.195703871  0.20431357
## IDACIScore_19_SPR24          4.183671e-04  -6.464193534 -1.86461579
## IDACI2                       8.736887e-01  -1.827641240  1.55365421
## IDACI3                       1.874214e-01  -3.096648241  0.60697717
## IDACI4                       9.855713e-01  -2.032358306  2.07012629
## IDACI5                       1.912367e-02  -7.488793055 -0.67026375
## ofsted_ratingGood            6.653676e-01  -1.656396516  1.05891985
## ofsted_ratingRequires improvement 5.813994e-01  -3.185709164  1.79157259
## ur_3catUrban city/town       6.294941e-02  -0.077040868  2.92326310
## ur_3catRural                 3.882345e-01  -1.190363209  3.05911720
## mixedAttain1:KS2_MATMRK_lo   4.893251e-02   0.005406591  2.29603571
## mixedAttain1:KS2_MATMRK_hi   1.050386e-01  -2.123652846  0.20136254

## N.Obs: 19179
```

Conclusion: some differences in significance levels but generally definitely confirms the main findings, suggesting missingness is not introducing bias.

Appendix C.19: Balance tables

Table C.19.1: Imbalance for sample weighted for number of matched comparator schools and the number of pupils in each school at planning, recruitment and analysis stages

Characteristic	Full sample (Unmatched)	All matched schools	Recruited and primary data matched	Analysed sample
KS2 2019	0.051	-0.025	-0.023	-0.09
KS2 2018	0.051	-0.025	-0.023	-0.11
KS2 2017	0.037	-0.029	-0.028	-0.28
Low Prior Attain Prop.	-0.096	0.022	0.006	-0.02
High Prior Attain Prop.	0.000	-0.084	-0.132	-0.19
Average (Attainment)	***0.009***	***-0.028***	***-0.040***	***-0.17***
Number of pupils on school roll[7]	0.235	-0.129	-0.270	0.17
FSM Prop.	-0.113	0.150	0.198	0.27
EAL Prop.	0.443	0.629	0.383	0.30
Academy status	-0.230	-0.298	-0.326	-0.20
IDACI1	0.053	-0.118	-0.165	0.00
IDACI2	-0.046	-0.057	-0.104	-0.12
IDACI4	-0.005	0.018	0.156	0.06
IDACI5	-0.123	0.075	-0.009	-0.11
Ofsted Outstanding	0.174	0.024	-0.266	-0.19
Ofsted Good	0.052	0.016	0.061	-0.06
Urban	-0.192	-0.105	-0.214	-0.24

[7] Weightings have been applied for the number of pupils in each school. Nevertheless, this weighting has a minimal effect on the imbalance in the the size of schools.

Table C.19.2: Baseline balance at pupil-level (unstandardised differences for mean, standard deviation and skewness)

Variable	mean_diff	sd_diff	skew_diff	num.obs
KS2_MATMRK	-0.16	0.15	-0.02	14877
KS2_MATMRK_lo	0.00	0.00	-0.01	4378
KS2_MATMRK_hi	-0.01	0.00	0.02	5139
rawSSCbase_scale	0.33	0.06	-0.10	14877
rawMSCbase_scale	-0.01	0.23	0.00	14877
FSM1	0.02	0.01	-0.15	2883
EAL1	0.02	0.02	-0.20	2281
IDACIScore_19_SPR24	-0.01	-0.01	-0.02	14877

Table C.19.3; Baseline balance at school-level (unstandardised differences for mean, standard deviation and skewness)

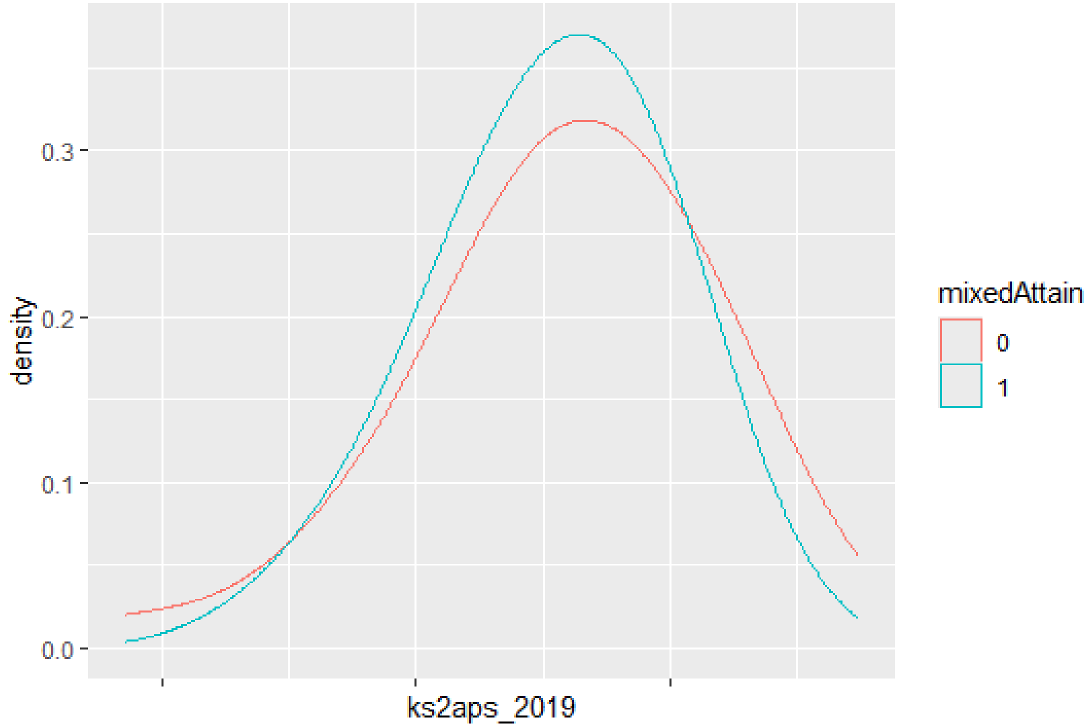
var	mean_diff	sd_diff	skew_diff	Num.obs
KS2_MATMRK	-0.14	0.45	-0.01	14877
KS2_MATMRK_lo	0.00	0.00	-0.02	4378
KS2_MATMRK_hi	0.00	0.00	-0.02	5139
rawSSCbase_scale	0.27	0.04	-0.08	14877
rawMSCbase_scale	0.02	0.25	-0.01	14877
FSM1	0.02	0.02	-0.19	2883
EAL1	0.03	0.03	-0.30	2281
IDACIScore_19_SPR24	0.00	-0.01	-0.20	14877

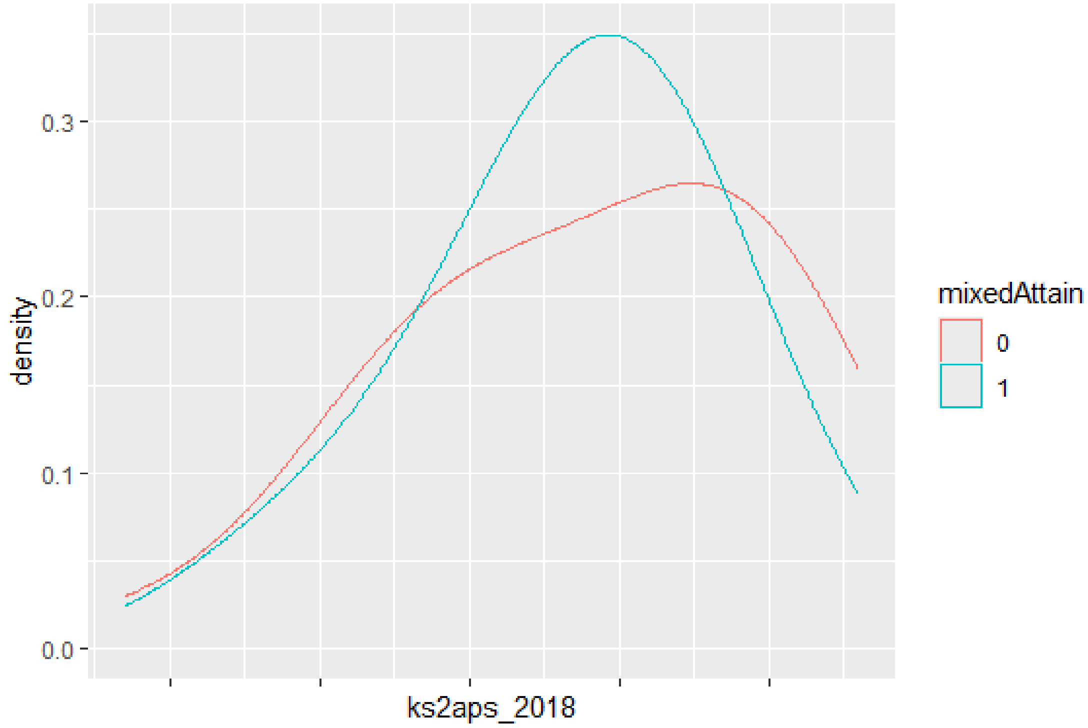
Table C.19.4: Baseline balance at pupil-level for FSM sub-group (unstandardised differences for mean, standard deviation and skewness)

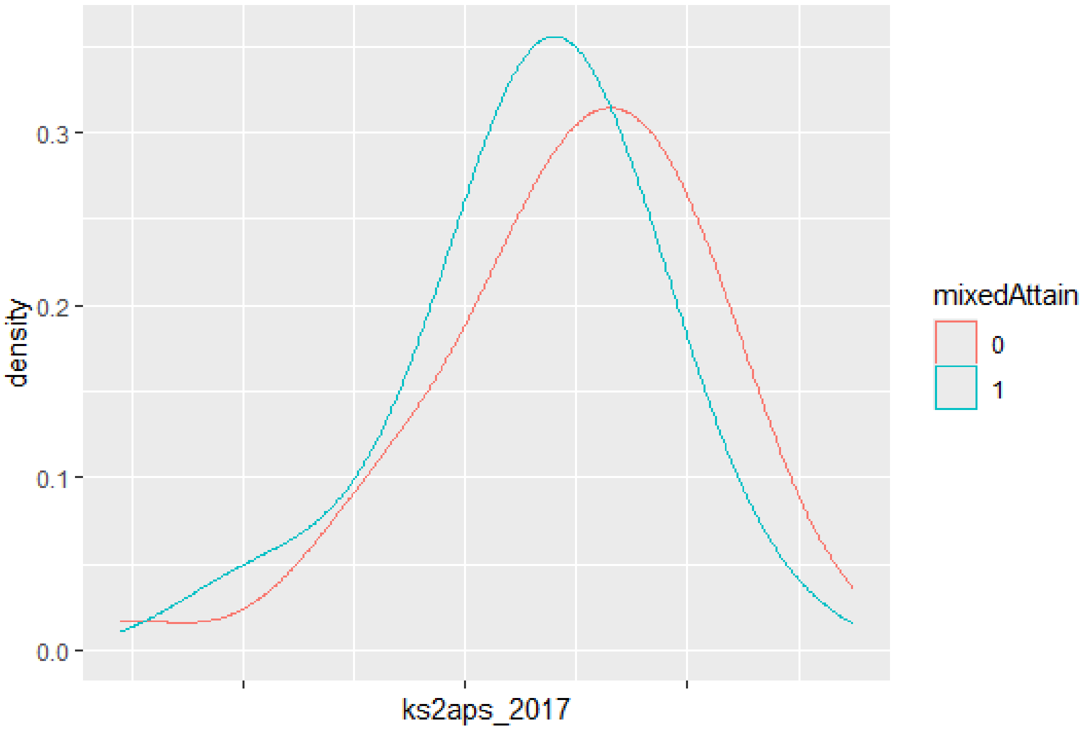
var	mean_diff	sd_diff	skew_diff	Num.obs
KS2_MATMRK	1.72	0.61	-0.05	2883
KS2_MATMRK_lo	-0.03	0.00	0.13	1299
KS2_MATMRK_hi	0.03	0.03	-0.27	568
rawSSCbase_scale	-0.28	-0.07	0.05	2883
rawMSCbase_scale	-0.33	0.06	0.03	2883
FSM1	0.00	0.00		568
EAL1	0.01	0.01	-0.11	562
IDACIScore_19_SPR24	-0.02	-0.01	-0.12	2883

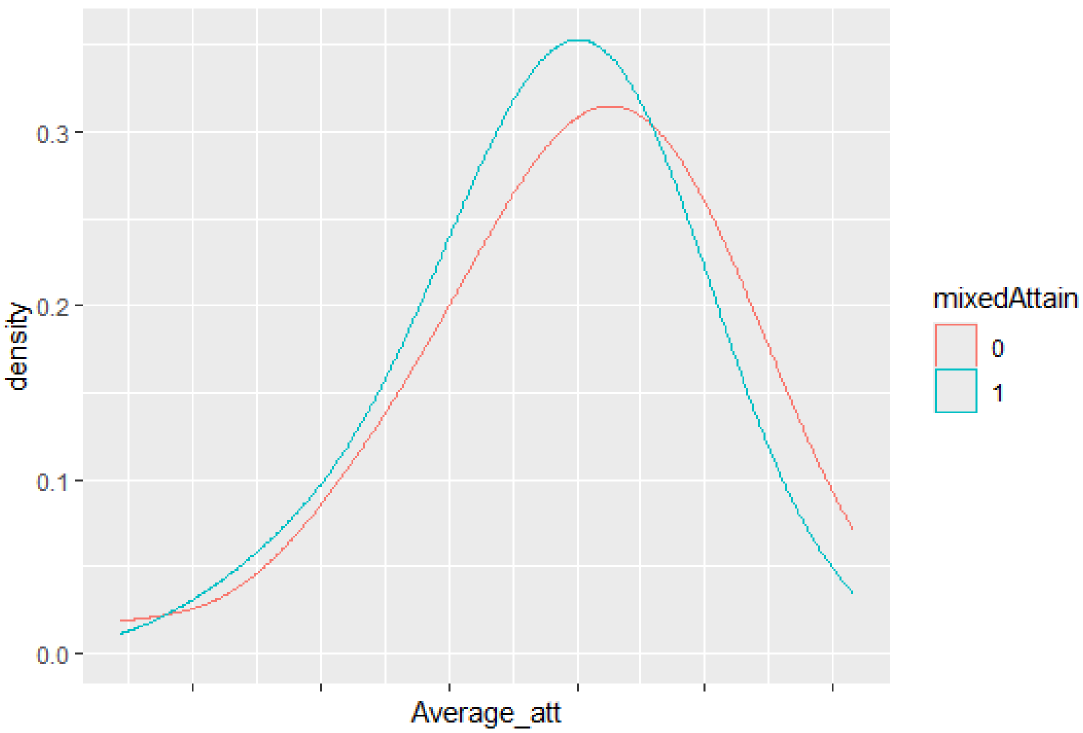
Overlapping kernel density plots

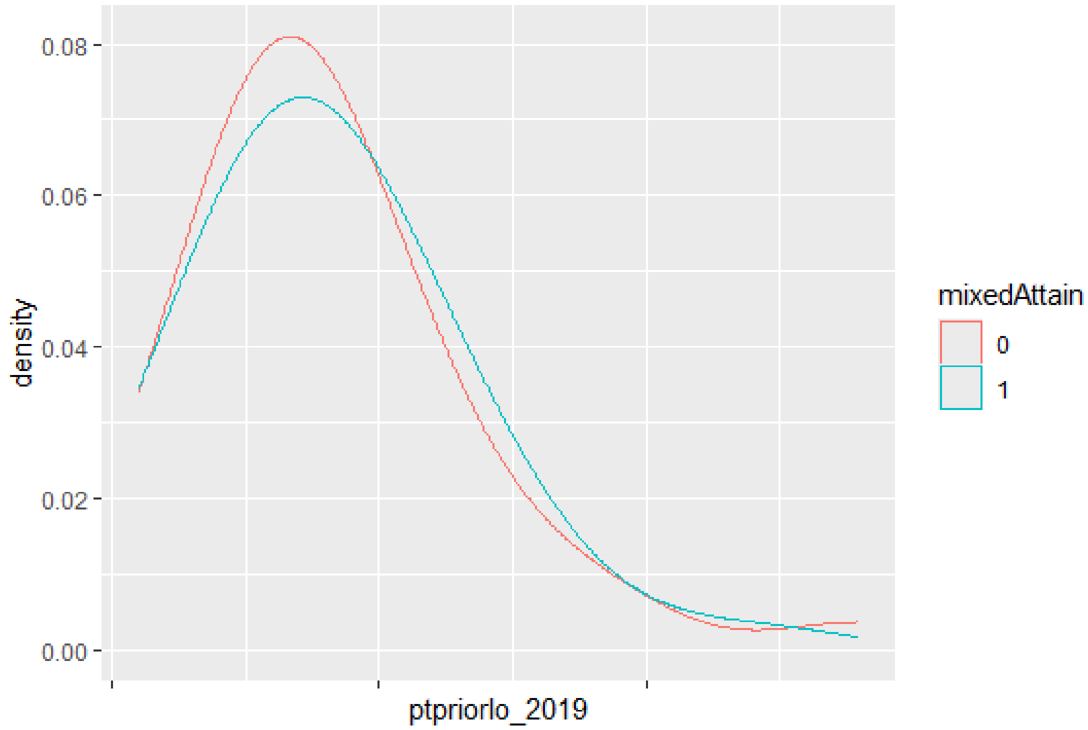
All at pupil level and using weight_att.
Num.Obs.mixedAttain.O: 10412. Num.Obs.mixedAttain.1: 4465.
X-axis value labels removed for statistical disclosure control purposes.

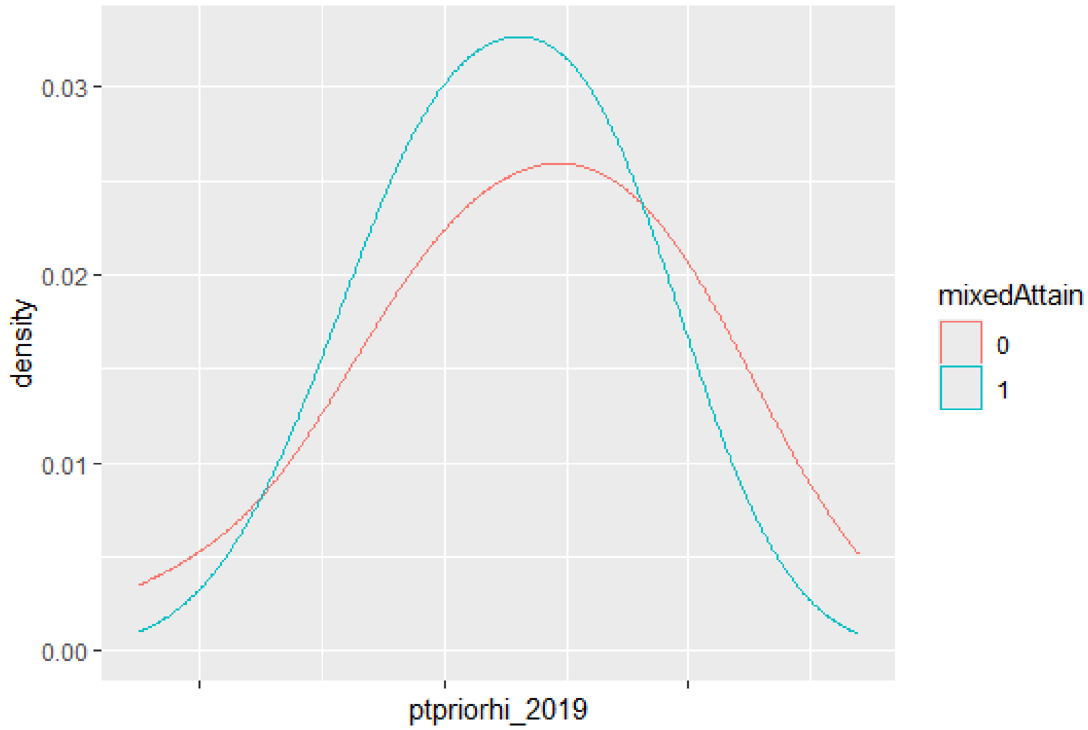


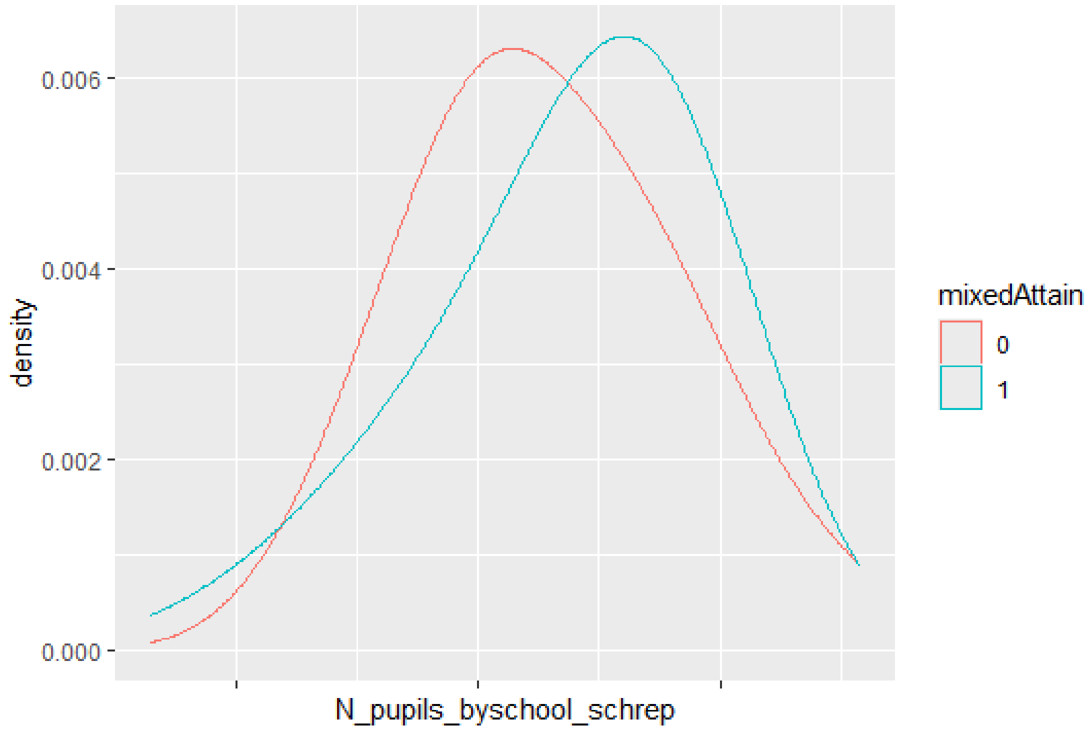


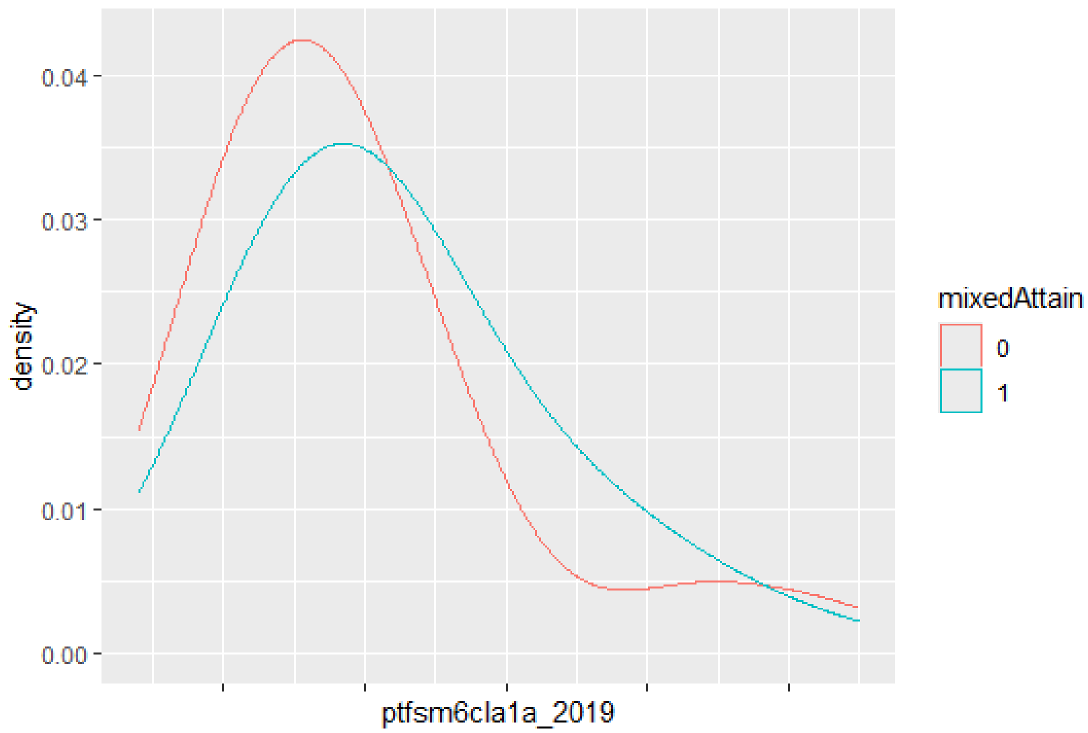

density
0.0
0.1
0.2
0.3
ks2aps_2017
mixedAttain
0
1


density
0.0
0.1
0.2
0.3
Average_att
mixedAttain
0
1


0.08
0.06
0.04
0.02
0.00
density
ptpriorlo_2019
mixedAttain
0
1

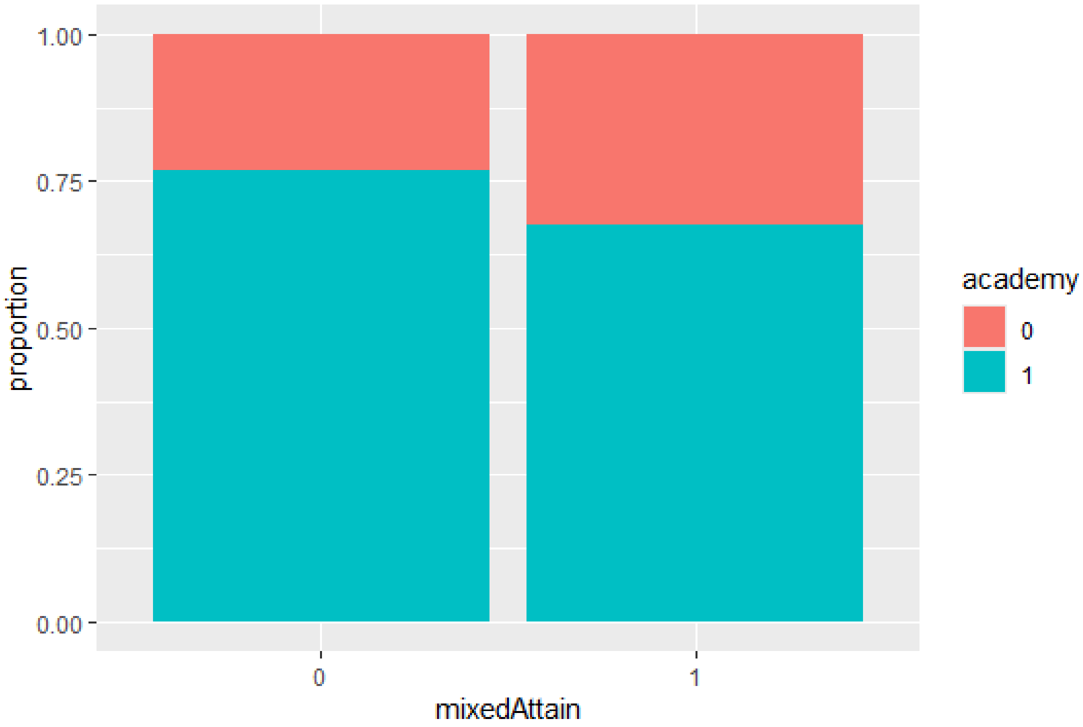

0.03
0.02
0.01
0.00
density
ptpriorhi_2019
mixedAttain
0
1



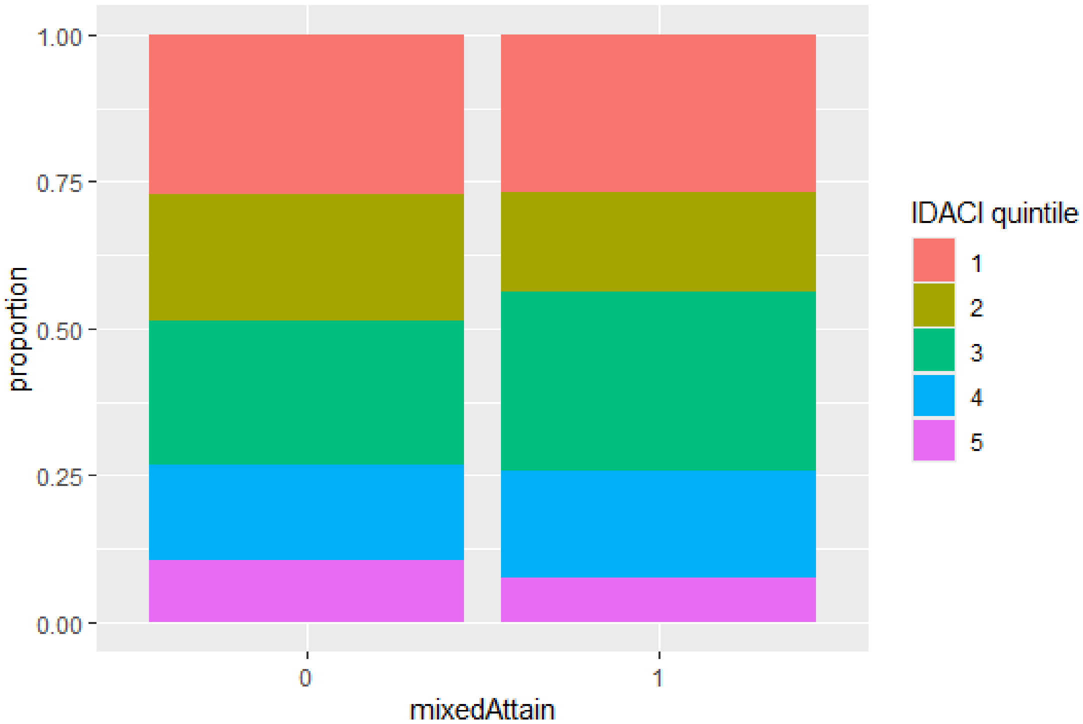


Stacked bar charts for categorical variables

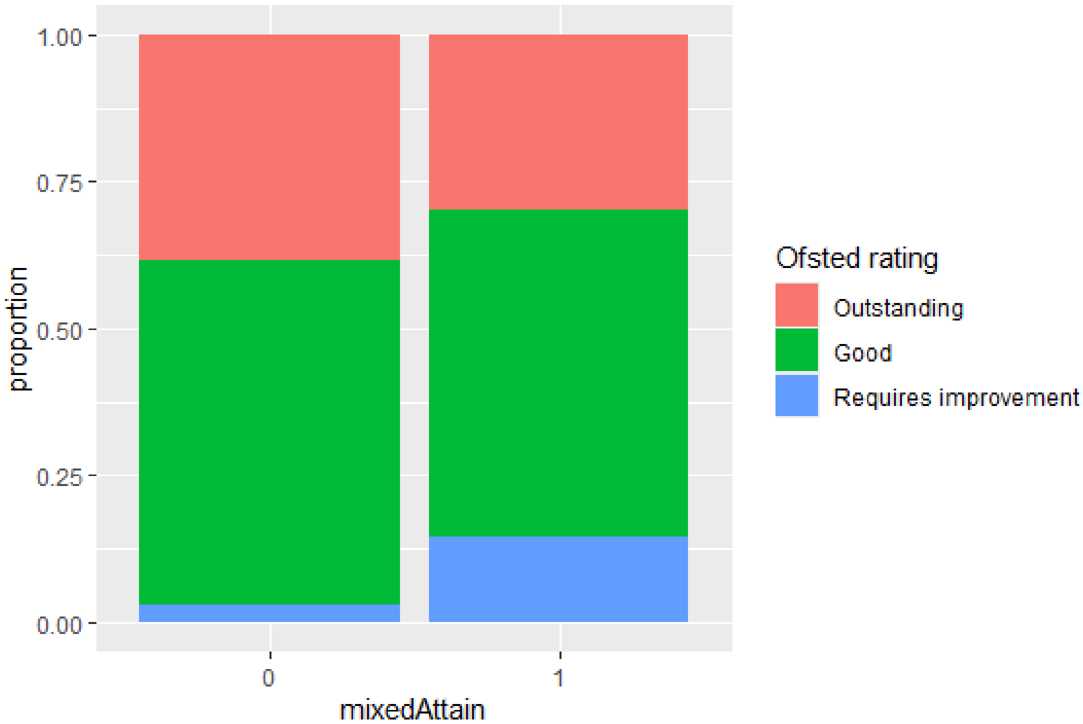
Note all at pupil level and using weight_att.
Num.Obs.mixedAttainO: 10412. Num.Obs.mixedAttain1: 4465.



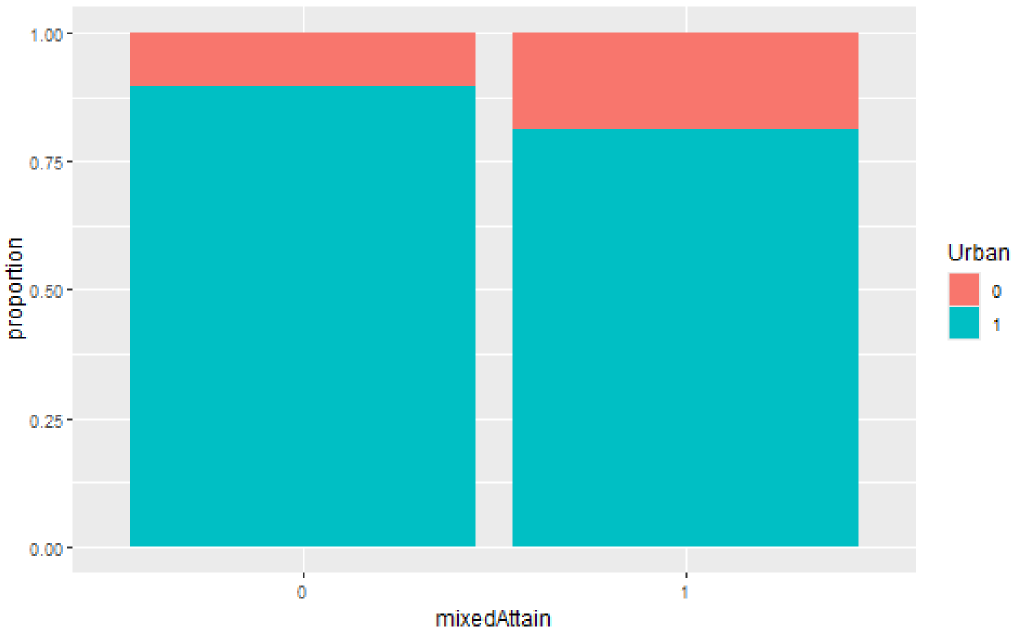
Cell counts	mixedAttain.0	mixedAttain.1	Total
Academy.0	Pupils: 2543 Schools: 15	Pupils: 1446 Schools: 9	Pupils: 3989 Schools: 24
Academy.1	Pupils: 7869 Schools: 47	Pupils: 3019 Schools: 18	Pupils: 10888 Schools: 65
Total	Pupils: 10412 Schools: 62	Pupils: 4465 Schools: 27	Pupils: 14877 Schools: 89



Cell counts	mixedAttain.0	mixedAttain.1	Total
IDACI1	Pupils: 2659 Schools: 17	Pupils: 1205 Schools: 7	Pupils: 3864 Schools: 24
IDACI2	Pupils: 2453 Schools: 13	Pupils: 752 Schools: 4	Pupils: 3205 Schools: 17
IDACI3	Pupils: 2722 Schools: 16	Pupils: 1351 Schools: 8	Pupils: 4073 Schools: 24
IDACI4	Pupils: 1665 Schools: 10	Pupils: 826 Schools: 5	Pupils: 2491 Schools: 15
IDACI5	Pupils: 913 Schools: 6	Pupils: 331 Schools: 3	Pupils: 1244 Schools: 9
Total	Pupils: 10412 Schools: 62	Pupils: 4465 Schools: 27	Pupils: 14877 Schools: 89



Cell counts	mixedAttain.0	mixedAttain.1	Total
Outstanding	Pupils: 3707 Schools: 20	Pupils: 1328 Schools: 7	Pupils: 5035 Schools: 27
Good	Pupils: 6096 Schools: 38	Pupils: 2487 Schools: 16	Pupils: 8583 Schools: 54
Requires improvement	Pupils: 609 Schools: 4	Pupils: 650 Schools: 4	Pupils: 1259 Schools: 8
Total	Pupils: 10412 Schools: 62	Pupils: 4465 Schools: 27	Pupils: 14877 Schools: 89



Cell counts	mixedAttain.0	mixedAttain.1	Total
Urban.0	Pupils: 1157 Schools: 8	Pupils: 3625 Schools: 22	Pupils: 4782 Schools: 30
Urban.1	Pupils: 9255 Schools: 54	Pupils: 840 Schools: 5	Pupils: 10095 Schools: 59
Total	Pupils: 10412 Schools: 62	Pupils: 4465 Schools: 27	Pupils: 14877 Schools: 89

Appendix C.20: Final matching process adopted

This appendix details the final matching process adopted for the Student Grouping Study. First, we imported the full list of schools, where NPD KS4 data were available for 2019, that we constructed and merged together elsewhere, including contextual information about schools and school-level averages about the pupils in these schools. The dataset also includes an identification of whether the school is one of the recruited mixed attainment schools; other known mixed attainment schools (notably those in earlier phases of this project) have been removed from the sample to prevent them being identified as potential comparators.

These are the variables that must be present for schools to be included in our sample (* indicates identified as a potential variable for matching, see Appendix C.3):

- URN
- Mixed attainment identifier
- Average KS2 performance of 2016, 2017 and 2018 GCSE cohorts (observations are carried backwards where necessary)*
- Proportion of 2018 GCSE cohort identified as low, average and high prior attainers*
- Proportion of FSM in school*
- Proportion of EAL in school*
- IDACI quintile*
- Number of pupils in KS4 cohort in 2018
- Most recent Ofsted judgement*
- Academy status*
- Urban/rural identifier*
- Single sex school
- Region

Matching is carried out using the MatchIt software with the aim of identifying 25 setted schools as potential matches for each mixed attainment school. The propensity score is estimated using the following logistic regression model fit using school-level data:

$$MixedAttain = FSMProp + KS2_{2019} + KS2_{2018} + KS2_{2017} + PriorLo + PriorHi + PupilNo. + IDACI' + Ofsted' + \varepsilon$$

Appendix C.21: Matching and Simulated Response Exercise

To support our decisions regarding matching approach, we have conducted a matching and simulated response exercise to guide our approach set out in the study plan.

Propensity score estimation

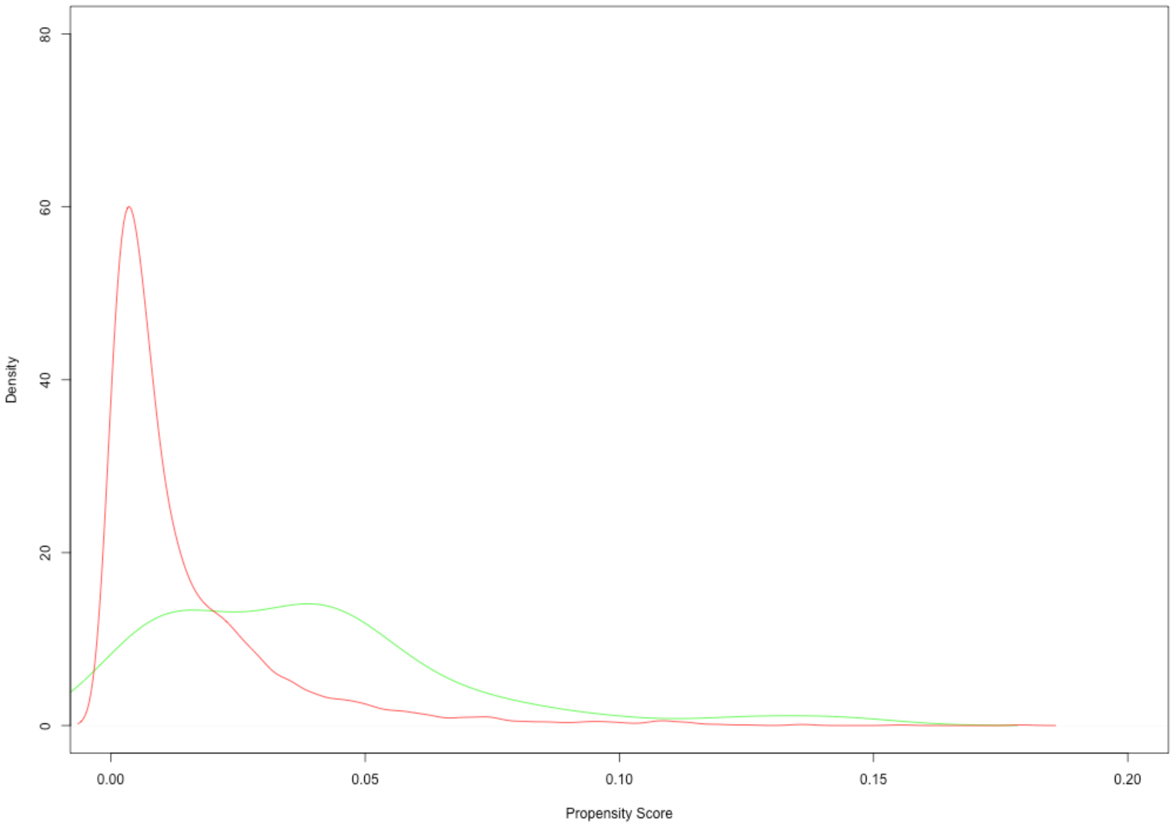
Matches are found based on a treatment propensity score estimated from the following generalised linear model with a probit link function:

$$
\begin{aligned}
probit(MixedAttain_i) \\
&= \beta_0 + \beta_1 Prop.FSM_{i2017} \\
&+ \beta_2 Prop.HighPrior_{i2017} + \beta_3 Prop.LowPrior_{i2017} + \beta_4 Academy_{i2017} + \beta_5 KS2_{i2018} \\
&+ \beta_6 KS2_{i2017} + \beta_7 KS2_{i2016} + \beta_8 No.Pupils_{i2017} + \beta' \boldsymbol{IDACI_{i2017}} + \beta' \boldsymbol{Ofsted_{i2017}} \\
&+ \beta' \boldsymbol{Region_i} + \beta' \boldsymbol{Urban_i} + \varepsilon_{it}
\end{aligned}
$$

where $MixedAttain_i$ is our 0/1 indicator of whether school i is a mixed attainment school, $Prop.FSM_{i2017}$ is proportion of the school eligible for Free School Meals in 2017, $Prop.HighPrior_{i2017}$, is proportion of the school's cohort identified as "high attainment" in 2017, $Prop.LowPrior_{i2017}$, is proportion of the school's cohort identified as "low attainment" in 2017, $KS2_{it}$ is average KS2 score of the school's intake in year t, $No.Pupils_{i2017}$ is the size of the cohort, $\boldsymbol{IDACI_{i2017}}$ is a vector of binary variables indicating the quintile group into which the school falls in terms of the average Index of Deprivation Affecting Children and Infants (IDACI) of its intake, $\boldsymbol{Ofsted_i}$ is a vector of binary variables indicating the school's most recent Ofsted overall effectiveness grade, $\boldsymbol{Region_i}$ is a vector of binary variables indicating the government office region in which the school is located, $\boldsymbol{Urban_i}$ is a vector of binary variables indicating the urban/rural setting of the school, and ε_i is an idiosyncratic error term.

This model was based on iterative testing of model fit of available matching variables. Given the small number of treatment schools (i.e. 43 mixed attainment schools), there a risk of instability from use of more complicated models. However, this model produces substantially improved balance compared to a simpler version.

Figure C.21.1. Density plot of distribution of propensity scores for mixed attainment and all other schools (unmatched)



Notes. Green line plots density for mixed attainment schools; red line plots density for all other schools.

Matching approach

For the purpose of this exercise, matches are found using an optimal matching algorithm with no replacement (in practice, allowing replacement makes no difference in this application, seemingly because there are plenty of potential matched comparators available) using the MatchIt package in R. We explore identifying both 20 and 25 potential matched comparators (from which recruitment will be attempted) for each mixed attainment school to explore the pros and cons of each.

We start with the hypothesis that 25 matches will have slightly worse match quality on average, however with 20 matches we are more likely not to successfully recruit one of these schools as part of this initial matching exercise (which would require identification of further potential matches, which would likely be of lower quality).

Match quality

In Figures C.21.2 and C.21.3 we plot an illustration of the distribution of the schools by their treatment status and whether they are identified as matched comparison schools. In Figure C.21.2 this is plotted for 1:20 matching and in Figure 3 this is plotted for 1:25 matching. These plots demonstrate a good spread of matched comparison schools across the propensity score range of the matched treatment schools, which is not appreciably different depending on the number of matched schools identified.

In Table C.21.1 we report the standardised differences between the treatment and potential matched comparison samples for 1:20 and 1:25 matching, along with these statistics for the unmatched sample.

Figure C.21.2. Plot of propensity score distribution of schools by treatment and matching status (1:20 matching)

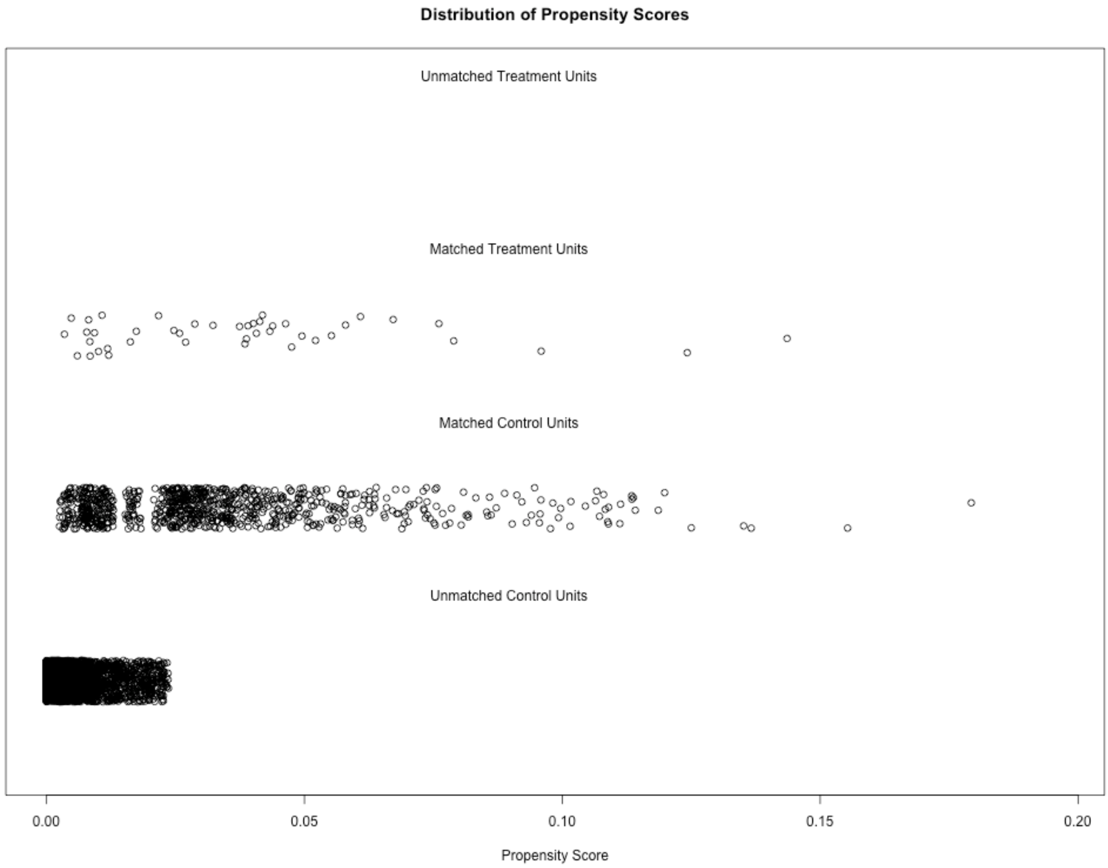


Figure C.21.3. Plot of propensity score distribution of schools by treatment and matching status (1:25 matching)

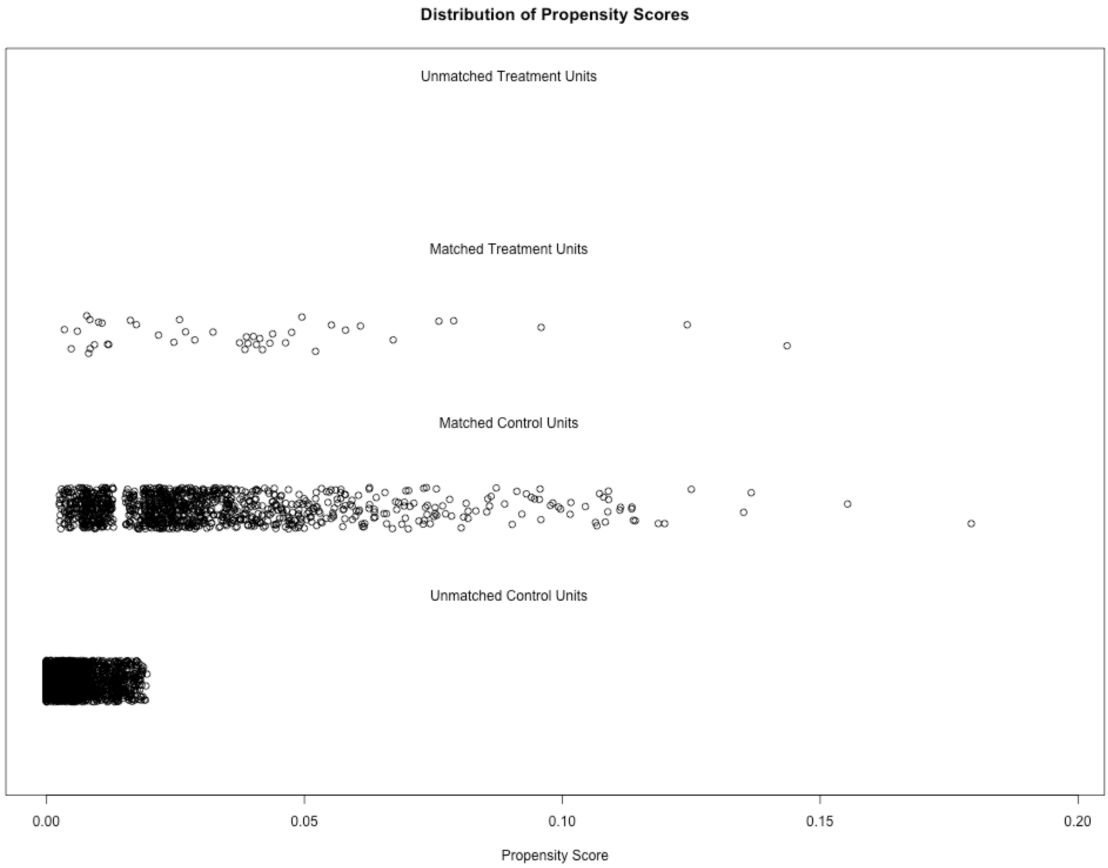


Table C.21.1. Standardised differences in characteristics by matching methods

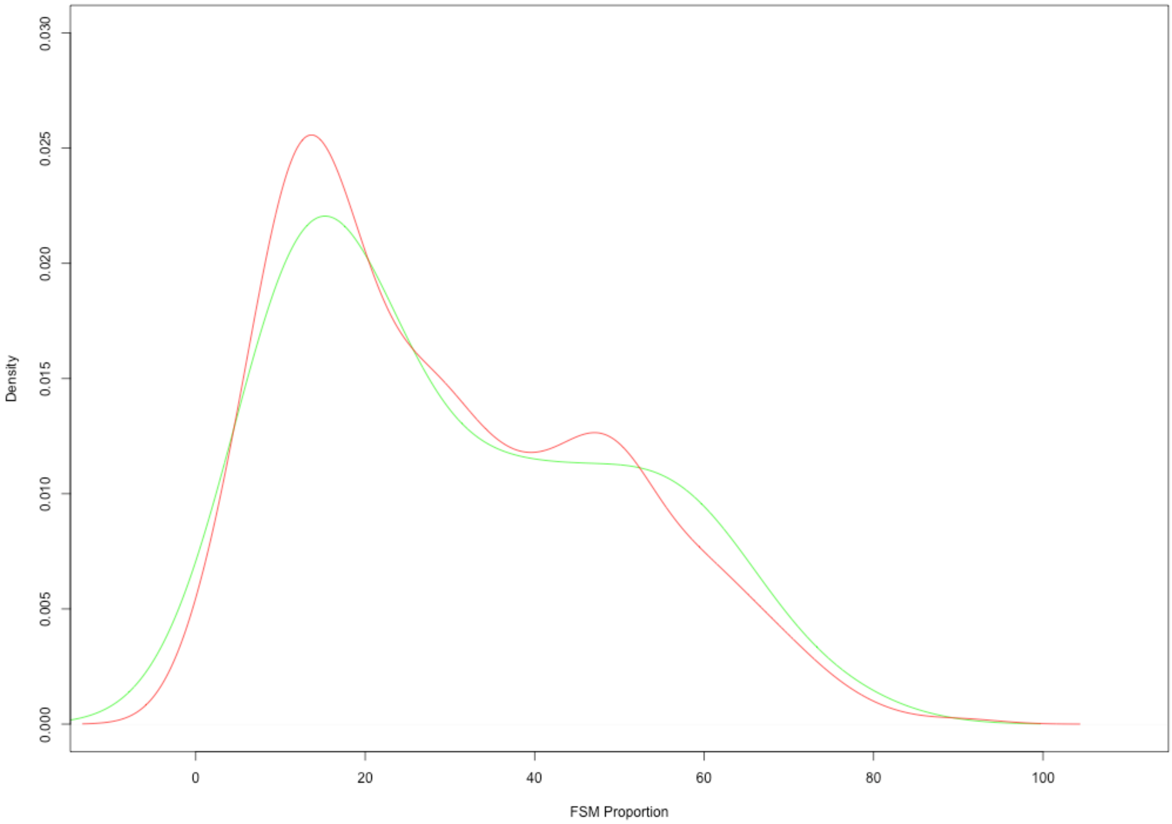
Characteristics	20 Matches	25 Matches	Unmatched
No. Pupils	0.10	0.08	0.28
Academy Proportion	-0.26	-0.28	-0.10
FSM Proportion	0.05	0.12	0.21
KS2 APS 2018	-0.01	-0.03	0.19
KS2 APS 2017	-0.02	-0.06	0.08
KS2 APS 2016	-0.02	-0.05	0.11
Low Attainers Prop.	-0.02	0.00	-0.21
High Attainers Prop.	-0.02	-0.04	0.18
IDACI Q1 Prop.	0.04	-0.01	0.32
IDACI Q2 Prop.	0.00	-0.01	-0.16
IDACI Q4 Prop.	-0.01	-0.01	-0.21
IDACI Q5 Prop.	0.03	0.08	0.27
Ofsted Outstanding Prop.	0.03	0.00	0.21
Ofsted Good Prop.	0.04	0.05	-0.06
Urban Setting Prop.	-0.05	-0.04	0.14
East Mids Prop.	0.04	0.03	0.12
East of England Prop.	0.09	0.02	-0.21
London Prop.	0.05	0.13	0.71
North West Prop.	-0.02	-0.05	-0.15
South East Prop.	0.00	-0.06	-0.11
South West Prop.	-0.04	-0.03	-0.11
West Mids Prop.	-0.10	-0.09	-0.07
Yorks/Humb Prop.	-0.01	-0.01	-0.08
Average	*0.05*	*0.06*	*0.19*

Notes. Reporting "Std. Diff" between treated and comparison schools identified by each matching method described. Standard differences calculated by dividing means by overall sample standard deviation. "Mean Abs. Std. Diff" = Mean absolute standard difference calculated across characteristics in table. IDACI Quintile 5 and Ofsted: Inadequate categories excluded since these are determined by the remainder of the other categories of this variable.

We consider the distribution of selected continuous characteristics and how this differs between treatment and matched comparison groups in the following plots. In Figures C.21.4 and C.21.5, we plot the density of the proportion of FSM pupils in the school for 1:20 and 1:25 matching, respectively. In Figures C.21.6 and C.21.7, we do the same for average KS2 points score on intake. In Figures C.21.8 and C.21.9, this is repeated for the proportion of the school's intake identified as low attainment by the DfE, while Figure C.21.10 and C.21.11 do the same for the proportion of the school's intake identified as high attainment.

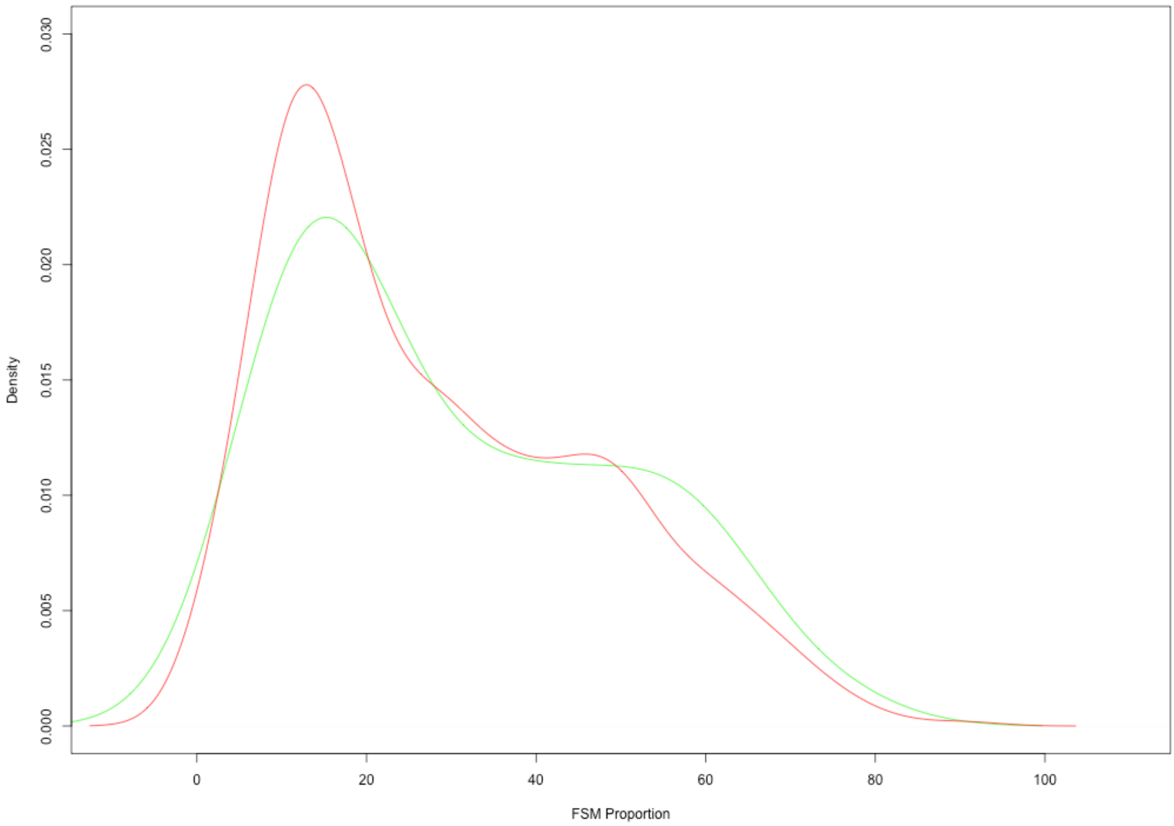
Overall, it is unclear that balance is substantially worse among the potential comparators in the case of 1:25 matching. Later in this document, we check that our simulated responses patterns among these potential comparators does not alter this picture.

Figure C.21.4. Distribution of proportion of pupils identified as FSM in treatment and potential comparison groups (1:20 matching)



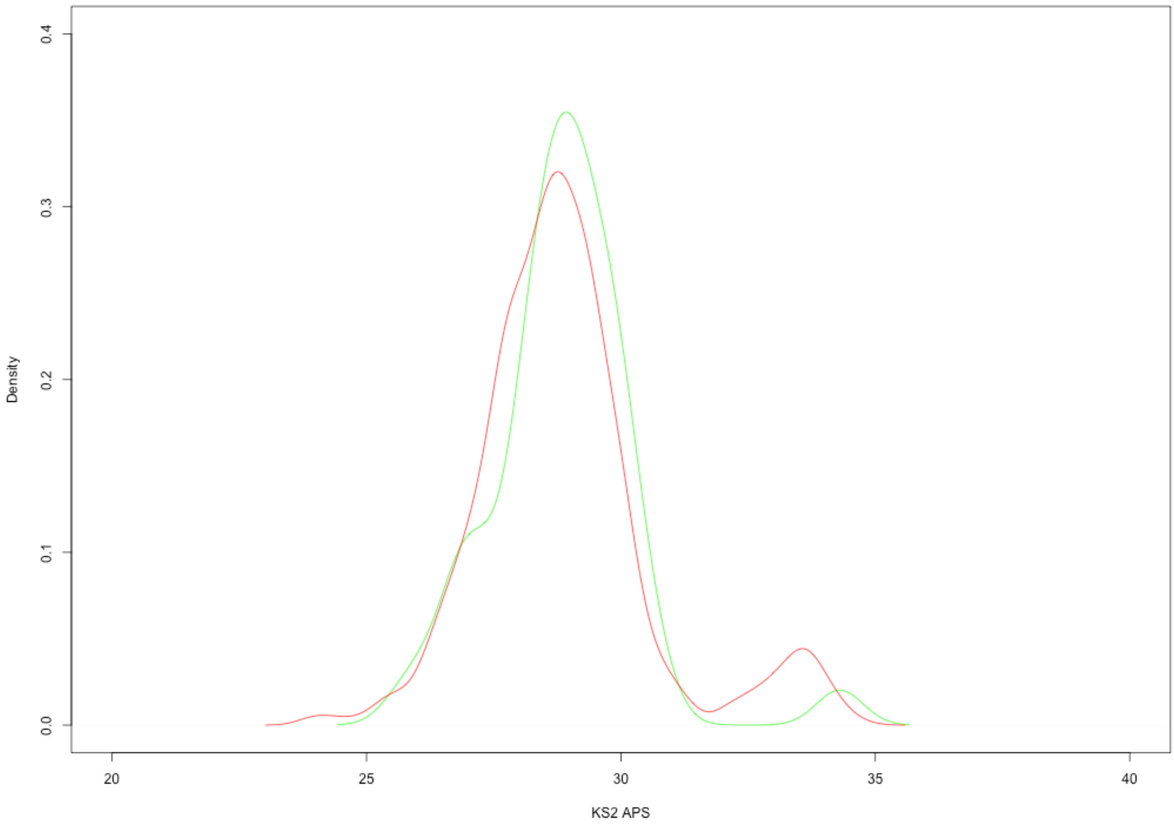
Notes. Kernel density plot of school FSM proportion for treated (green) and comparison (red) schools.

Figure C.21.5. Distribution of proportion of pupils identified as FSM in treatment and potential comparison groups (1:25 matching)



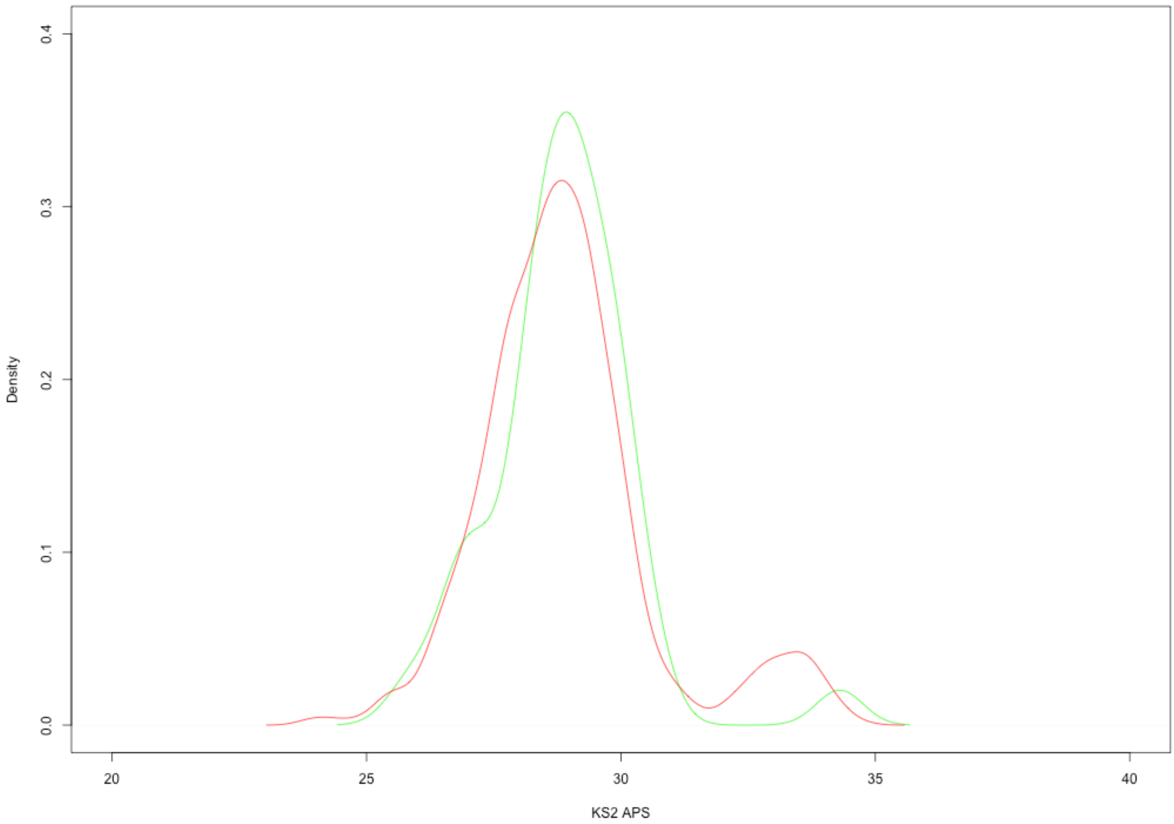
Notes. Kernel density plot of school FSM proportion for treated (green) and comparison (red) schools.

Figure C.21.6. Distribution of average KS2 prior attainment in treatment and potential comparison groups (1:20 matching)



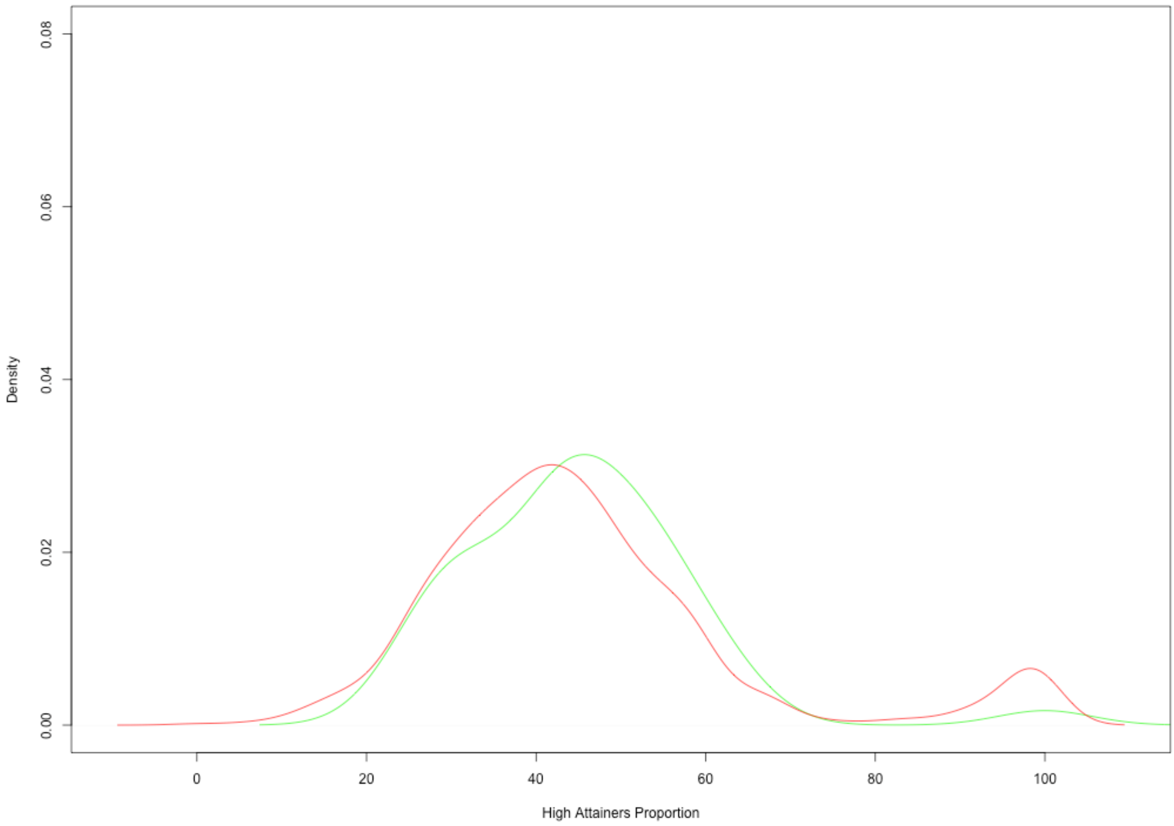
Notes. Kernel density plot of school average KS2 prior attainment for treated (green) and comparison (red) schools.

Figure C.21.7. Distribution of average KS2 prior attainment in treatment and potential comparison groups (1:25 matching)



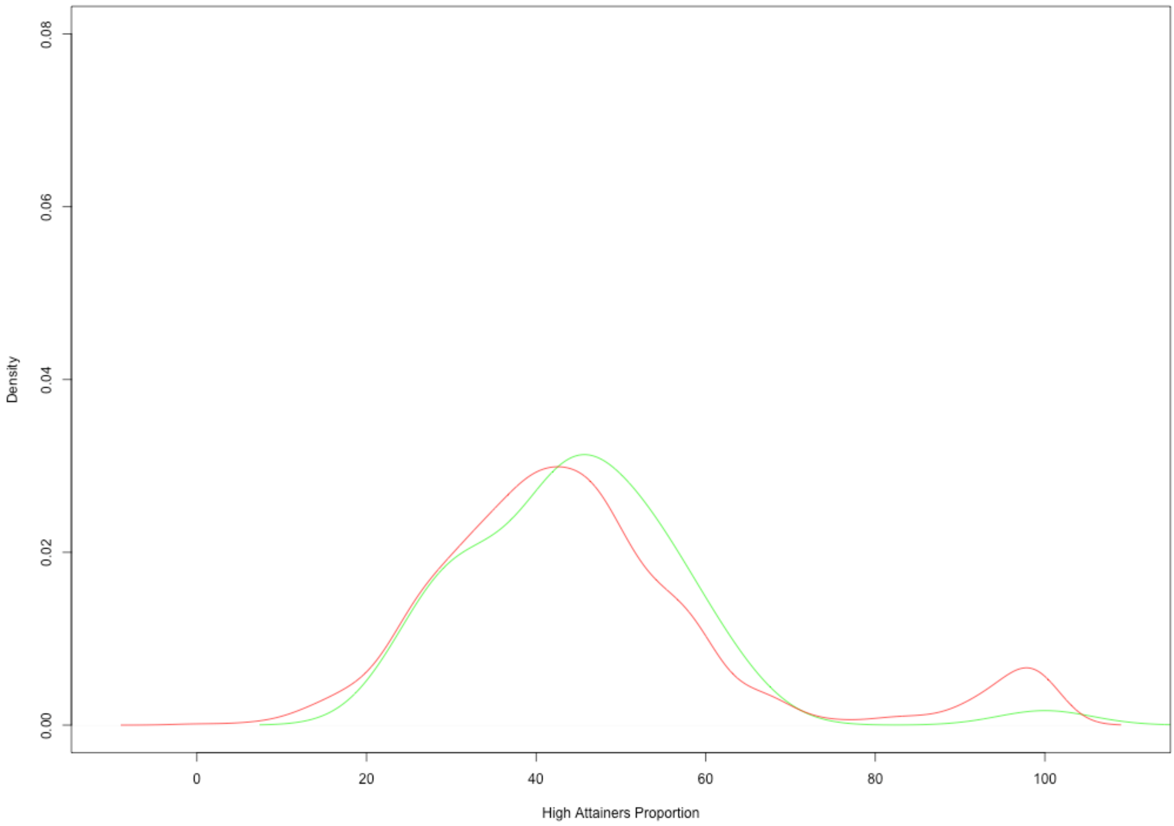
Notes. Kernel density plot of school average KS2 prior attainment for treated (green) and comparison (red) schools.

Figure C.21.8. Distribution of proportion of pupils identified as high attainment on intake in treatment and potential comparison groups (1:20 matching)



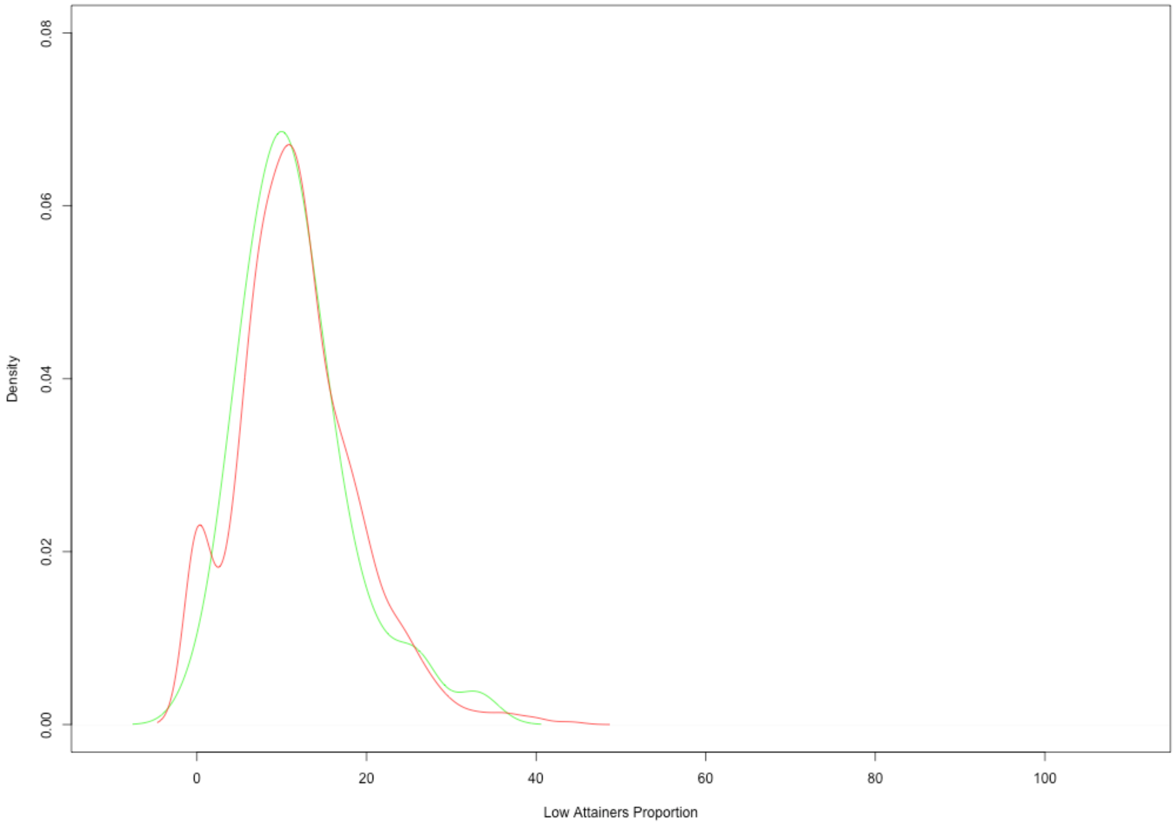
Notes. Kernel density plot of proportion of pupils identified as high attainment on intake by DfE for treated (green) and comparison (red) schools.

Figure C.21.9. Distribution of proportion of pupils identified as high attainment on intake in treatment and potential comparison groups (1:25 matching)



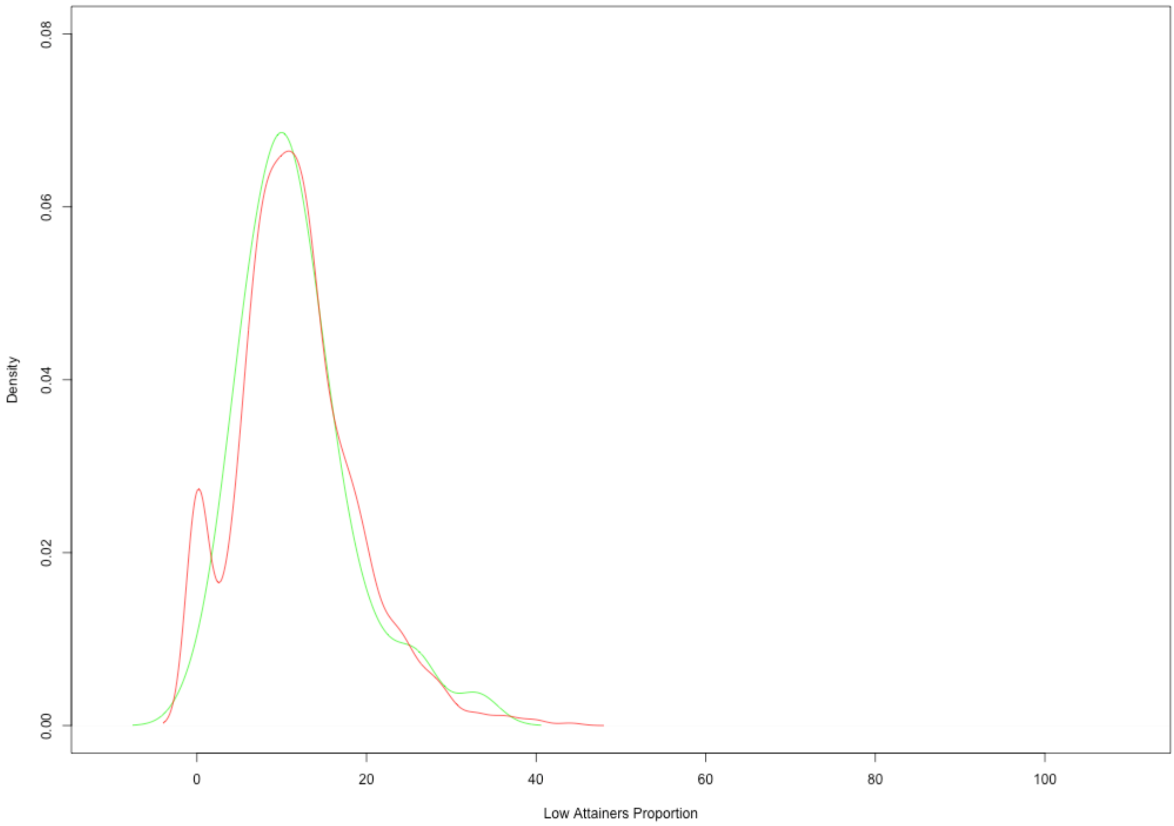
Notes. Kernel density plot of proportion of pupils identified as high attainment on intake by DfE for treated (green) and comparison (red) schools.

Figure C.21.10. Distribution of proportion of pupils identified as low attainment on intake in treatment and potential comparison groups (1:20 matching)



Notes. Kernel density plot of proportion of pupils identified as low attainment on intake by DfE for treated (green) and comparison (red) schools.

Figure C.21.11. Distribution of proportion of pupils identified as low attainment on intake in treatment and potential comparison groups (1:25 matching)



Notes. Kernel density plot of proportion of pupils identified as low attainment on intake by DfE for treated (green) and comparison (red) schools.

Simulated response: failure to recruit

For this study, it is necessary to actively recruit schools, since the data needed for this evaluation cannot be extracted entirely from administrative datasets. As such, we carry out a basic simulation of this recruitment process, as follows.

In 1000 simulations, we assign all schools identified as matched comparators a response probability drawn randomly from a uniform distribution between 0 and 1. We assume that those schools with response probabilities above 0.8 will join the study if contacted to do so. We then treat as recruited the two schools with a response probability above 0.8 with the smallest difference in propensity score from its respective treated school for each mixed attainment school. In doing so, we mimic the recruitment process in which we will work systematically through a matched recruitment list for each mixed attainment school sorted in the same way, continuing until two schools have been recruited or the list has been exhausted.

In the same way, in some simulations it is the case that there is only one school, or even no schools, with a response probability about 0.8 in the potential matched comparator list for each mixed attainment school. This is more likely to be the case when only 20 potential matched comparators are identified, rather than 25, which we demonstrate with the following analysis.

Table C.21.2. Proportion of simulations in which the column title number of schools achieves the row title number of responses – 1:20 matching

	0	1	2	3	4					
No responses	0.612	0.315	0.065	0.007	0.001					
	0	1	2	3	4	5	6	7	8	
One responder	0.082	0.208	0.27	0.212	0.135	0.055	0.026	0.01	0.002	
	34	35	36	37	38	39	40	41	42	43
Two responders	0.002	0.004	0.024	0.045	0.079	0.178	0.241	0.215	0.163	0.049

Table C.21.3. Proportion of simulations in which the column title number of schools achieves the row title number of responses – 1:25 matching

	0	1	2	3		
No responses	0.841	0.147	0.01	0.002		
	0	1	2	3	4	5
One responder	0.328	0.394	0.183	0.073	0.02	0.002
	38	39	40	41	42	43
Two responders	0.006	0.036	0.094	0.208	0.37	0.286

We note that there are limitations to this approach. The 0.8 probability cut off is an assumption (based on an estimated recruitment probability of 0.2) and this simple process makes the assumption of no correlation between school characteristics and response probability. Note, however, that because the list is worked through systematically from the potential matched comparator for each mixed attainment school with the smallest difference in propensity scores from the treated school to the one with the largest.

Simulated response: effects on imbalance

Response patterns will also have an effect on imbalance, relative to the matched sample. We can use our simulations to explore these. Given the recruitment strategy we intend to follow, i.e. prioritising those with the most similar propensity scores to the mixed attainment schools, our simulations suggest that, if anything this process is likely to reduce imbalance relative to the initial matched sample. This is perhaps unsurprising, given the large size of the matched sample that we generate.

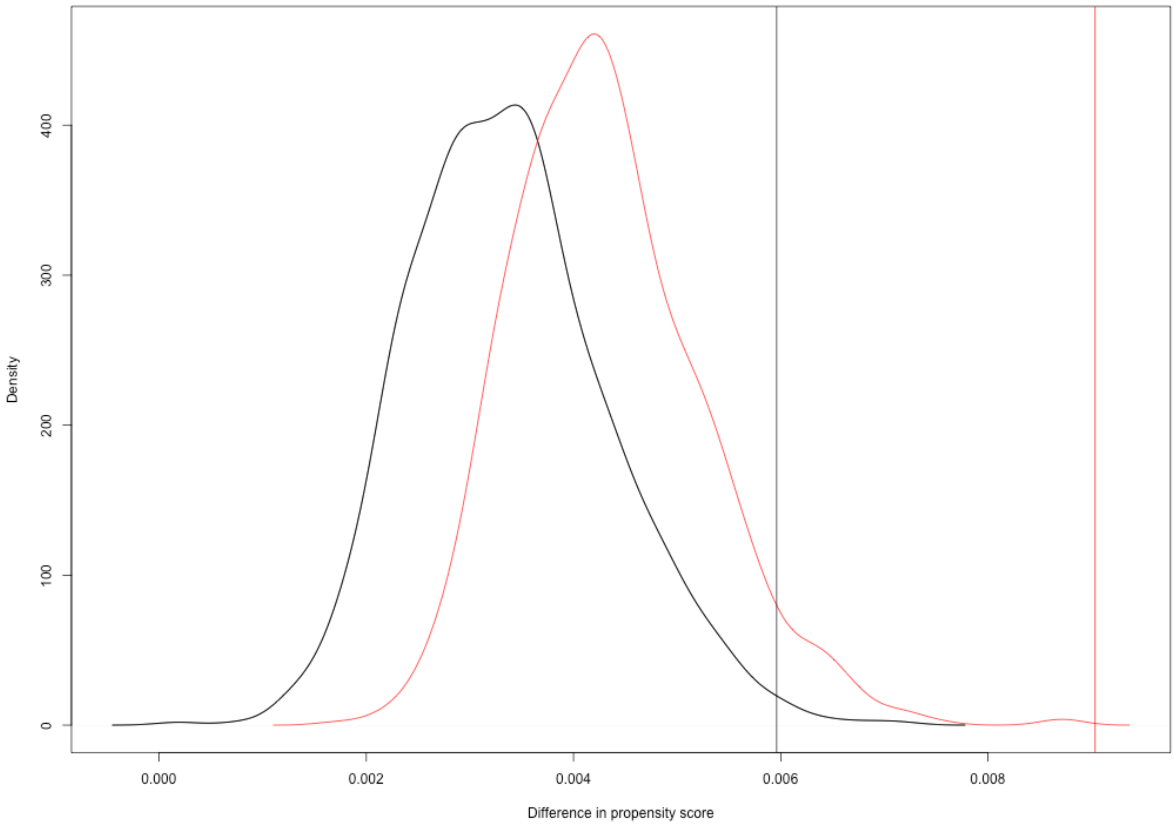
Figure C.21.12 plots the distribution of the difference in propensity score between treatment and simulated recruited schools by whether 1:20 or 1:25 matching was carried out in the initial matching process. Figures C.21.13 to C.21.18 repeat this but for standardised difference measures of imbalance in key characteristics. Overall, we judge that it is not particularly the case that there is systematically better balance between the simulated recruited samples and the mixed attainment schools in the case of 1:20 matching compared to 1:25 matching.

Conclusions

Based on the above analysis, we are minded to adopt a 1:25 approach to matching given the increased probability that this will lead to successful recruitment within the initial matched sample, without evidence of this compromising the match quality of the finally recruited sample.

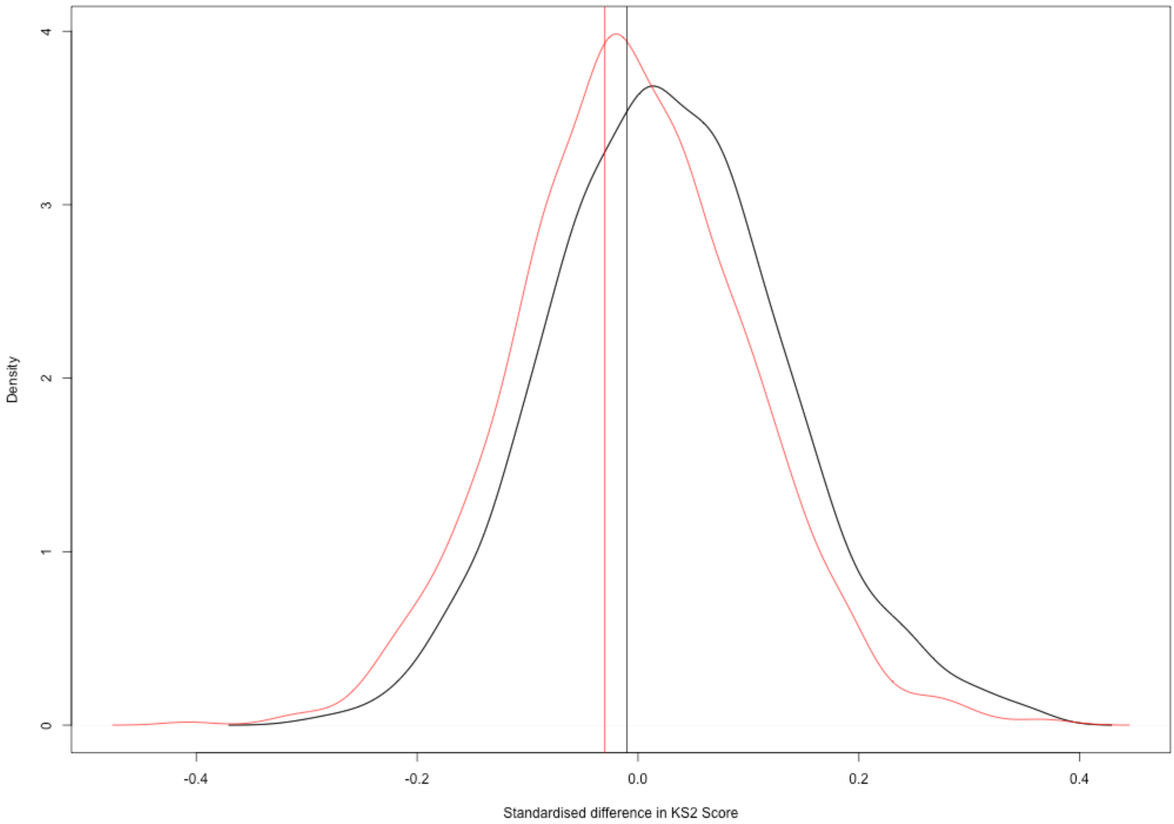
The exercise has also emphasised the particular importance of estimating propensity scores using a parsimonious model as part of this project, given the small number of mixed attainment schools available. This makes prioritisation of characteristics on which we need to achieve a good match to have confidence in the estimates from this project a particularly key issue.

Figure C.21.12. Simulated density of imbalance in propensity score measures after response: comparing 1:20 (black) to 1:25 (red) matching



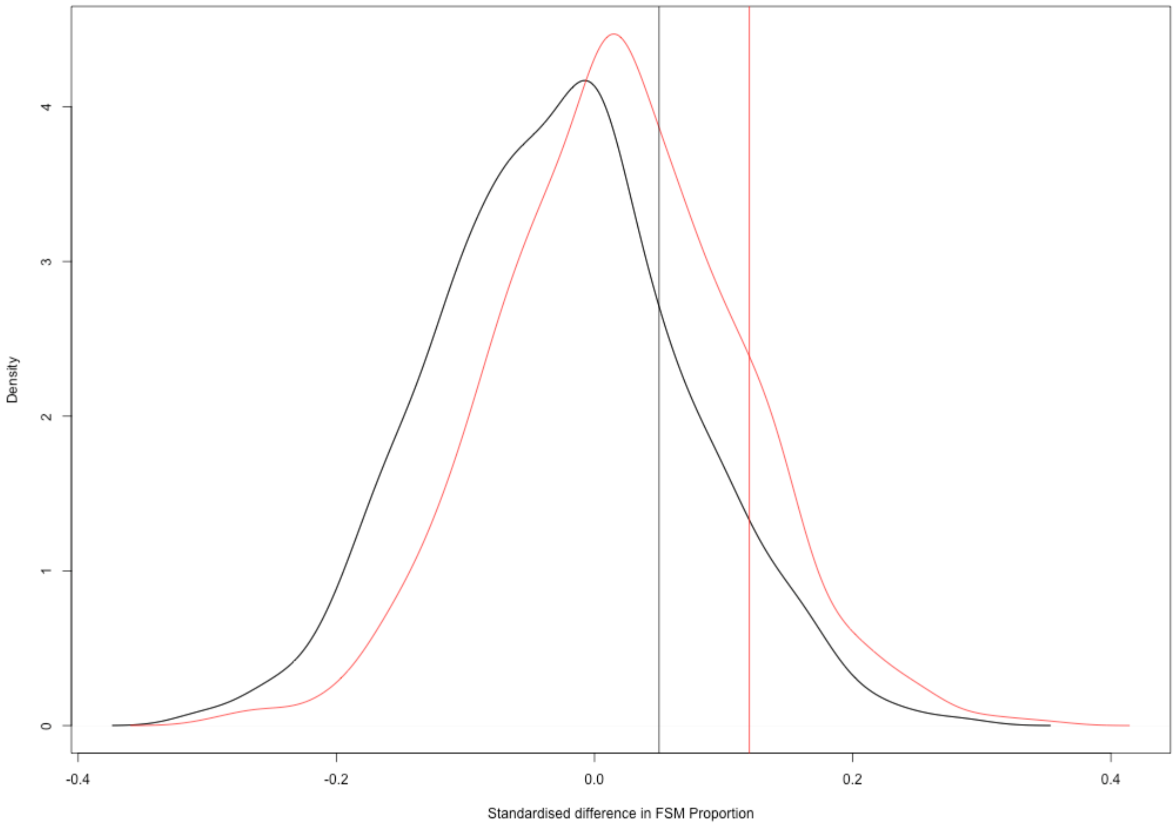
Notes. Density plots showing difference in propensity score between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Figure C.21.13. Simulated density of standardised imbalance in KS2 score after response: comparing 1:20 (black) to 1:25 (red) matching



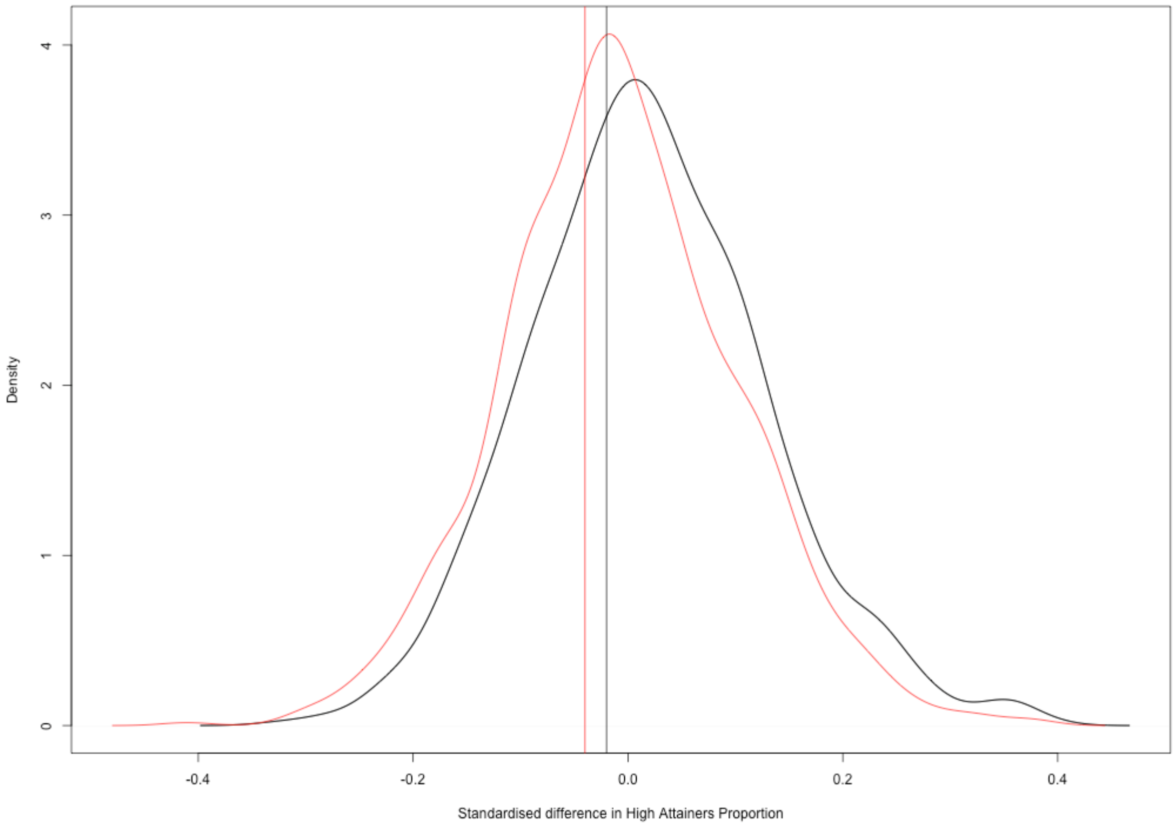
Notes. Density plots showing standardised difference in KS2 score between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Figure C.21.14. Simulated density of standardised imbalance in FSM proportion after response: comparing 1:20 (black) to 1:25 (red) matching



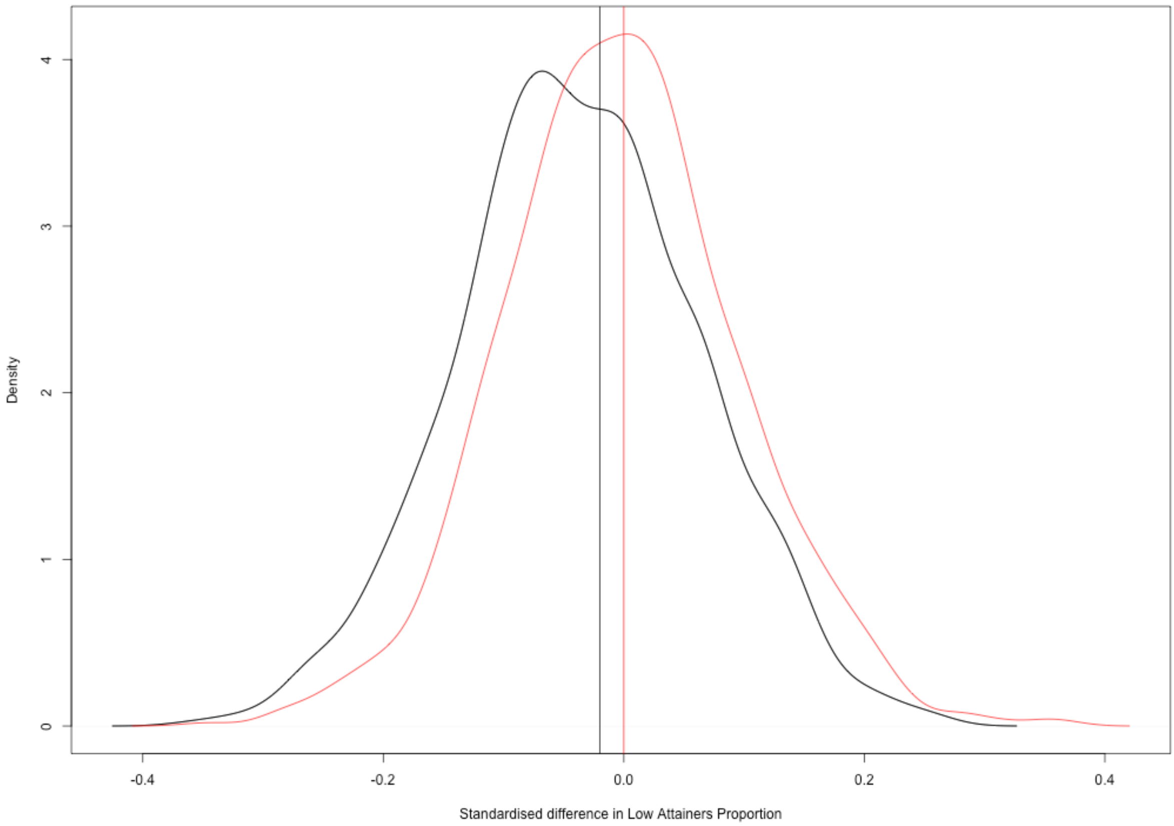
Notes. Density plots showing standardised difference in FSM proportion between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Figure C.21.15. Simulated density of standardised imbalance in proportion of high attainers after response: comparing 1:20 (black) to 1:25 (red) matching



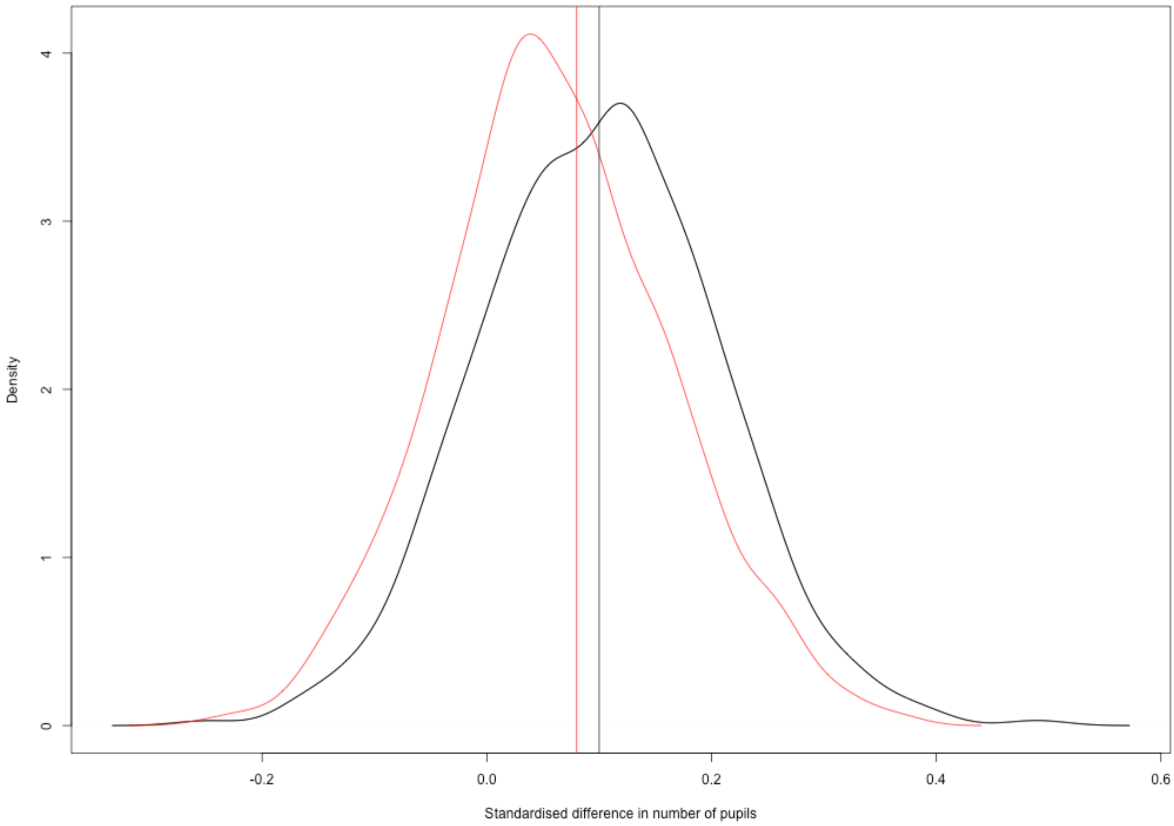
Notes. Density plots showing standardised difference in proportion of high attainers between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Figure C.21.16. Simulated density of standardised imbalance in proportion of low attainers after response: comparing 1:20 (black) to 1:25 (red) matching



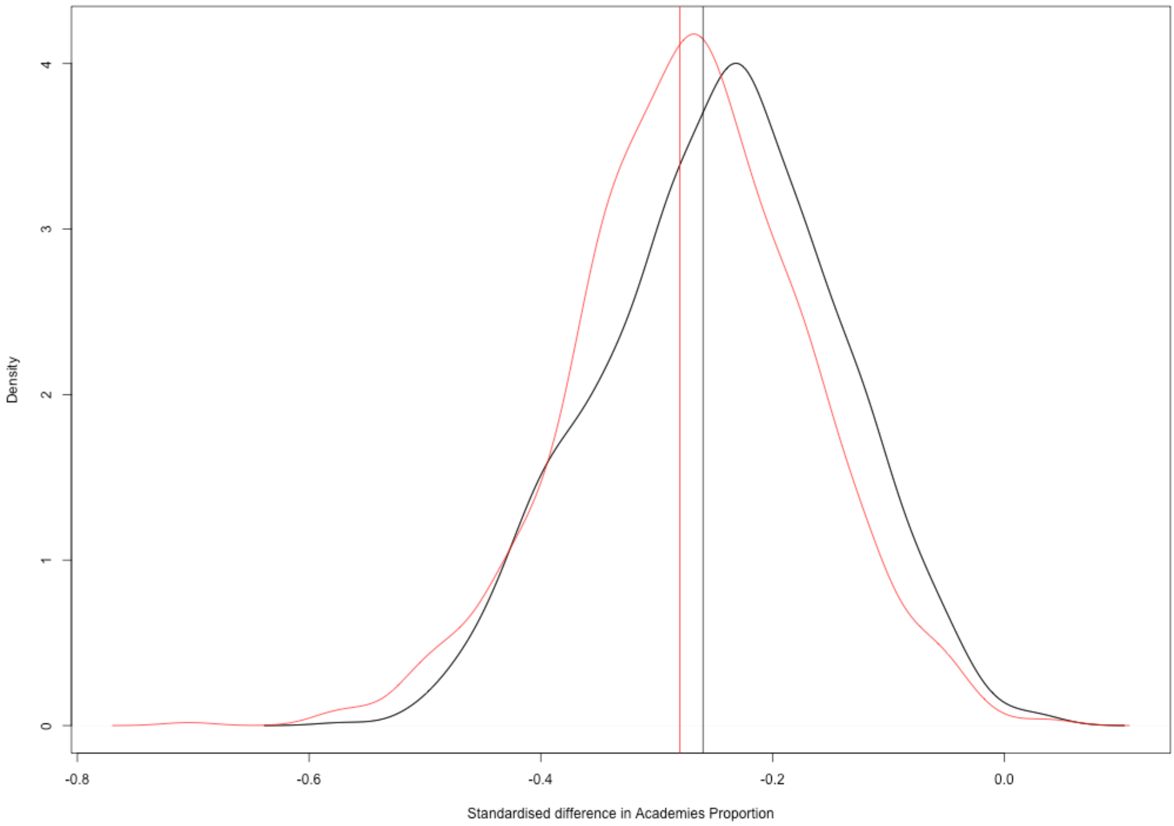
Notes. Density plots showing standardised difference in proportion of low attainers between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Figure C.21.17. Simulated density of standardised imbalance in number of pupils after response: comparing 1:20 (black) to 1:25 (red) matching



Notes. Density plots showing standardised difference in number of pupils between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Figure C.21.18. Simulated density of standardised imbalance in proportion of academies after response: comparing 1:20 (black) to 1:25 (red) matching



Notes. Density plots showing standardised difference in proportion of academies between treated and simulated responders among the matched comparison sample. Simulations based on 1:20 matched sample plotted in black; simulations based on 1:25 matched sample plotted in red. Simple imbalance from full sample of potential matches plotted as vertical line (1:20 matched sample plotted in black; 1:25 matched sample plotted in red).

Appendix C.22: Assessing threats to the validity of the findings

This study used a natural experiment design based on two matched samples of schools. In order to enable an assessment of the security of the findings, the EEF (2019) guidance was considered in the design and analysis as follows:

MAT.1. Explain how different variables are expected/hypothesised to be correlated with the treatment status and outcomes (i.e. confounders that will be considered). A key component of these evaluations requires exploring the validity of these hypothesised relationships.

The choice of variables for matching is explained in the Matching section of the main report (p.19) and in Appendix C.3. In the event, a few of these variables were not used in order to achieve a sufficient set of potential matches for each mixed attainment school. Given the focus of the study, the attainment variables were considered to be the most important: average school prior attainment and proportions of low and high attainers at entry.

This study sits within a larger body of work by the evaluation team in which these hypothesised relationships have been explored, as discussed in the TIDieR framework section of the main report. The IPE further explores the relationships through identifying how school characteristics and context influence grouping decisions, and by exploring mediators and moderators in the logic model.

MAT.2. Explore the sensitivity of results including appropriate sensitivity analyses which may include alternative specifications of the Matching/Weighting, additional variables and interaction effects. As there is no consensus on the primacy of one approach or a specific matching algorithm irrespective of the characteristics of the sample, it is necessary to discuss why the chosen approach is suitable to analyse the sample under study.

In contrast to many matched designs, matching was used to recruit fixed treatment *and control* samples for which implementation and outcome data would be collected. This limited the options available for matching. In the Matching section of the main report (p.19), we discuss the choice of PSM.

Sensitivity and robustness analyses were carried out: a replication of the main analysis using matched pupil-level data and a quantile analysis.

We also considered the effect of missing data. See Missing Data sections in the main report.

MAT.3. Assess the balance in the distribution of relevant covariates included in the matching/weighting between treatment and comparison groups, before and after the matching is done. Express differences in terms of standardised differences, as those are not dependant on sample sizes. Assess differences in mean values and higher order moments between the groups. When some differences remain even after matching/weighting, consider the use of alternative methods.

Matching was conducted at school-level before pupil-level data was available. However, we consider the balance at pupil-level to be the most important consideration for this study. The imbalance of key variables at pupil-level for the actual achieved sample is reported in the main body of the report. These are expressed using standardised differences where appropriate.

Imbalance for the key independent variable (KS2 Mathematics Attainment) is good; the standardised difference of 0.01 is within EEF's low risk category. All the continuous variables are within the threshold of 0.1 established for the study at design stage. Differences in means and higher moments are reported in Appendix C.19 together with standardised differences of school-level matching variables at design, allocation and analysis stages. Given the variable number and size of recruited matched comparison schools for each mixed attainment school, we used weighting to reduce the importance for estimation of pupils in setting schools where we recruited multiple (or different sizes of) matched setting schools for each mixed attainment school. (See Recruitment of setting schools, p.19, main report.) In addition, we included a large number of covariates in our models, thus increasing the precision and power. (See Appendix C.16.)

Differential attrition between the two groups is 6.7 percentage points, which is within the US What Work Clearinghouse's (2022) optimistic threshold for differential attrition.

MAT.4. Explore the area of common support and the characteristics of those included.

Compare the characteristics of those included in the common support and those for whom no match was found. Explain whether common support is imposed, why, as well as its implications. Consider using methods that employ information from all individuals (for example, inverse probability weighting on the propensity score). When using Inverse Probability Weighting, consider exploring the distribution of weights and including robustness excluding large weights.

Common support was imposed; two mixed attainment schools were dropped at allocation stage because no matches were identified. Because of the small number of dropped schools, it was not considered appropriate to compare their characteristics with those of the included schools. The loss of just two schools is not considered a serious threat.

For this study, matching was used to recruit fixed treatment *and control* samples. Hence, the main analysis used information from all the recruited schools. As noted above, weighting was used to account for the variable number and size of setting schools matched to each mixed attainment school. Specifically, we scaled the weights such that each student in a mixed attainment school is worth '1', but those in the comparator setting schools are down-weighted such that the collective comparators in each stratum add up to the number of students in the mixed attainment school.

Additionally, as already noted, a sensitivity analysis was conducted using matched samples of pupils (matched using pupil-level variables).

MAT.5. As Matching/Weighting cannot account for unobservable heterogeneity, consider including additional robustness checks of the sensitivity to hidden bias, e.g. using Rosenbaum Bounds.

As outlined in the methods, we planned to do include additional robustness checks, but it was not considered appropriate due the inconclusive nature of the findings. Intuitively, these methods quantify the extent of hidden bias that would be needed to overturn a finding: with no significant finding to overturn this is not a meaningful analysis.

MAT.6. Select the approach to used based on its ability to reduce imbalance. It is strongly preferred that this choice is made before outcomes are observable to the research team.

As discussed above, the PSM matching process was combined with weighting in order to reduce the importance of pupils in setting schools where multiple matches were recruited. Both approaches were chosen before the matching and recruitment process was carried out and before the outcome data was collected.

Appendix C.23: Random selection of case study schools

Following Maxwell et al.'s (2021) assessment of high-quality sampling for qualitative work in IPEs, we used a random sampling approach. Our random case study school sample was stratified by prior attainment in line with evidence that grouping practices are associated with prior attainment of pupils in secondary schools in England (see Taylor et al., 2020). A school level 'prior attainment variable' was produced, based on the average Key Stage 2 point score for students from each school who sat GCSE examinations in summer 2019 and for whom data was therefore available. Schools were then ranked by prior attainment and divided into two equally-sized strata. The characteristics of the strata for the achieved sample are described in Table C.23.1.

Table C.23.1. Number of schools by grouping practices, stratified by high and lower prior attainment

	Total schools	Prior attainment variable range **Higher** stratum	No of schools	Prior attainment variable range **Lower** stratum	No of schools
Mixed	30	30.4-29.0	15	28.8-27.0	15
Setting	80	31.5 - 29.3	40	29.2 - 25.7	40

Randomisation was conducted using the Data Analysis tool pack in MS Excel. To enable replication, random seed values were generated using the RAND function (see Table C.23.2). Using these seed values, four sets of random numbers were generated from a uniform distribution. These sets of randomly-generated numbers were then assigned to the schools in each of the four strata. Schools were then re-ranked within each stratum by random number, from lowest to highest.

Table C.23.2. Random seed values used to generate random number sequences

	Seed (set)
Mixed Attainment Higher	52
Mixed Attainment Lower	66
Setting Higher	22
Setting Lower	18

Appendix C.24: Deductive coding framework using the Teaching for Robust Understanding framework for different data sources

TRU Dimension code	(A) Lesson narratives	(B) Focal student interviews	(C) Teacher interviews
Mathematical content	Examples from the lesson of mathematical content, discussed by teacher and/or students	Segments of the interview where students reflect on the mathematical content in the lesson observed. (Interview schedule q1-q4)	Segments of the interview where the teacher reflects on the intended and lived mathematical content of the lesson observed.
Cognitive demand	Examples from the lesson where students are engaged in different forms of mathematical thinking (e.g. single word responses, explaining own reasoning, explaining interpretation of others' thinking, asking the teacher/ class mathematical questions)	Segments of the interview where students reflect on the level of mathematical challenge they experienced in the lesson observed. (Interview schedule q3-q10)	Segments of the interview where the teacher reflects on the intended and lived levels of mathematical challenge in the lesson observed.
Access	Examples from the lesson where teaching practices engage different students' participation. (e.g. whiteboard sharing, group work)	Segments of the interview where students reflect on their own and others' participation in the lesson observed. (Interview schedule q11-q14)	Segments of the interview where the teacher describes the practices they used to ensure wide participation and reflects on the different participation patterns of different students in the lesson observed.
Agency, ownership and identity	Examples from the lesson of teaching practices involving students' mathematical contributions i.e. where these are built on or not. (e.g. the use of a student's mathematical error, bringing a student to the board to be 'the teacher')	Segments of the interview where the students reflect on the usefulness of their own and others' contributions in the lesson observed. (Interview schedule q15-q18)	Segments of the interview where the teacher reflects on moments in the lesson where students' contributions were useful or not in the observed lesson.
Formative assessment	Examples from the lesson where the teacher solicits student responses and does or does not make use of them.	Segments of the interview where the students reflect on how well they think they did in the lesson and how they think their teacher knows how well they did. (Interview schedule q19-q22)	Segments of the interview where the teacher reflects on what they learned about their students' during the lesson and moments where the teacher did or did not choose to develop on students' ideas.

Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn