



Low attainment in mathematics:

an investigation focusing on Year 9 pupils in England

TECHNICAL REPORT

Jeremy Hodgen, Robert Coe, Colin Foster and Margaret Brown
with Steve Higgins and Dietmar Küchemann



Author biographies

Jeremy Hodgen is a Professor of Mathematics Education at University College London Institute of Education. His research interests include mathematics education at all levels, assessment, international comparisons, approaches to grouping pupils and the evaluation of several large-scale educational trials. He has a particular interest in research that directly informs policy and practice.

Robert Coe currently works for Evidence-Based Education and the Education Endowment Foundation. He was Professor of Education and Director of the Centre for Evaluation and Monitoring at Durham University from 2010-2018. His research interests include evidence-based initiatives in education and the involvement of practitioners in research.

Colin Foster is a Reader in Mathematics Education in the Mathematics Education Centre at Loughborough University. His research interests focus on the learning and teaching of mathematics in ways that support pupils' conceptual understanding. He is particularly interested in the design of classroom tasks that enable pupils to develop the necessary fluency in mathematical processes that will support them in solving mathematical problems.

Margaret Brown is an Emeritus Professor of Mathematics Education at King's College London and has directed over 25 research projects. She has taught in primary and secondary schools, been a member of four government committees on the mathematics curriculum, and is a former president of the British Educational Research Association and the Mathematical Association.

Steve Higgins is Professor of Education at Durham University. He is the lead author of the Sutton Trust / Education Endowment Foundation Teaching and Learning Toolkit and has a particular interest in the use of meta-analysis to develop understanding of effective use of research evidence for both policy and practice.

Dietmar Küchemann is an associate at King's College London. He is interested in designing intriguing mathematical task for pupils. He currently works on the Increasing Competence and Confidence in Algebra and Multiplicative Structures (ICCAMS) project and has recently published *Algebradabra - Developing a better feel for school algebra*.

Acknowledgements

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

The authors are grateful to the schools, teachers and pupils who took part in this study, to the members of the project advisory group (Nicola Bretscher, Ann Dowker, Peter Gates, Matthew Inglis, Jane Jones, Anne Watson, John Westwell and Anne White), to Cheryl Lloyd of the Nuffield Foundation, and to our colleagues who previously worked on the project (Lee Copping, Karen Jones and David Martin).

This research was partly conducted when the Principal Investigator, Jeremy Hodgen, was at the University of Nottingham (until 2017).

Note

This technical report provides additional, detailed information about the methods and analysis used in the research project entitled: *Low attainment in mathematics: An investigation focusing on Year 9 students in England*. It should be read in conjunction with the main project report, which is available from www.ucl.ac.uk/ioe.

Recommended citation

Hodgen, J., Coe, R., Foster, C., Brown, M., Higgins, S., & Küchemann, D. (2020). *Low attainment in mathematics: An investigation focusing on Year 9 students in England. Technical Report*. London: Nuffield Foundation.

Table of Contents

Glossary of abbreviations	5
Introduction.....	6
Research Questions 1 & 2:.....	7
Research question 3:	24
Research question 4:	36
References.....	42
Appendix 1: IMAP Number Items, Topic and Origin	43
Appendix 2: Arithmetic fact recall items (in order as presented).....	48
Appendix 3: Comparison of scores to LNRP and ICCAMS	49
Appendix 4: Factor loadings for PCA for each item on the main IMAP test.....	51
Appendix 5: The full dataset for the literature review and second-order meta-analysis	53

Glossary of abbreviations

CAI	Computer-Aided Instruction
CEM	Centre for Evaluation and Monitoring
CSMS	Concepts in Secondary Mathematics and Science
DI	Direct Instruction
ESRC	Economic and Social Research Council
FSM	Free School Meals
GCSE	General Certificate of Secondary Education
ICCAMS	Increasing Competence and Confidence in Algebra and Multiplicative Structures
IDACI	Income Deprivation Affecting Children Index
IMAP	Investigating Mathematical Attainment and Progress
InCAS	Interactive Computer Adaptive System
ITS	Intelligent Tutoring Systems
KS1, KS2 etc	Key Stage 1, Key Stage 2 etc
LNRP	Leverhulme Numeracy Research Programme
MidYIS	Middle Years Information System
NPD	National Pupil Database
Ofqual	Office of Qualifications and Examinations Regulation
RQ1, RQ2 etc	Research question 1, research question 2 etc
SES	Socio-Economic Status
Y9, Y5 etc	Year 9, Year 5 etc

Introduction

This project investigated low attainment in mathematics by focusing on the lowest attaining 40% of pupils in Year 9 in England and addressing the following research questions:

RQ1: What mathematics do low attaining secondary pupils understand, and what are their particular strengths and weaknesses in number, multiplicative reasoning and algebra?

RQ2: Can low attainment be characterised simply as cognitive delay? If not, to what extent and in what ways do low attaining pupils understand mathematics in qualitatively different ways to high attaining pupils?

RQ3: To what extent do low attaining pupils' prior understandings of mathematics, and of particular mathematical topics, help to explain the existence of the attainment gap? What is the relative contribution of these mathematical understandings in comparison to socio-economic status and other demographic factors?

RQ4: What is currently known about the effectiveness of teaching strategies and approaches that address low attainment in secondary mathematics?

RQ5: To what extent is mathematics currently taught in appropriate ways for low attainers?

This technical report provides additional, detailed information about the methods and analysis used in this research project. See the main project report for a discussion of the background to the project.

Research Questions 1 & 2:

What mathematics do low attaining Year 9 pupils understand and what are their particular strengths and weaknesses in number, multiplicative reasoning and algebra?

Do low attaining pupils understand mathematics in qualitatively different ways to middle and/or high attaining pupils?

Our focus was on the lowest attaining 40% of the Year 9 cohort. Our overall aim was to understand the nature of low attainment in mathematics in lower secondary and to gather evidence on what mathematics these pupils know. In doing this, we aimed to investigate whether low attainers' understandings of mathematics were qualitatively different from those of middle and high attaining pupils by comparing to a group of middle and high attaining Year 5 pupils with similar overall attainment to the Year 9 low attaining group.

The test design, validation, administration and results are described in the main report. In this technical report, we provide additional information about the test items, the process of matching a Year 9 low attaining group with a Year 5 middle and high attaining group of pupils.

The test items

The final test consisted of two broad elements:

- *The main Number, or mathematics, element [henceforth, IMAP Number]:* 61 items, focused largely on key aspects of the number, calculation and 'pre-algebra' curriculum. Items were presented online in a variety of formats (including multiple choice, free entry and sliders). We reported outcomes as a total score and as scores for subsets of items grouped by pre-designed topic (see below).
- *A separate timed Arithmetic fact retrieval, or recall, element [henceforth, Arithmetic recall]:* a speeded fact retrieval test, consisting of 30 items, with as many as possible to be completed in a total of two minutes. This is reported separately from the main IMAP score.

See Appendices 1 and 2 for the full set of items in each element of the test, together with a brief justification for the inclusion of each item.

Selecting matched samples

The distribution of total scores on the 61-item IMAP Number test for the Y5 and Y9 samples can be seen in Figure 1. It is clear that the Y9 sample, as expected, is considerably higher attaining than the Y5 one, so, in order to make valid comparisons of factor structure, item characteristics, etc., we selected a matched subsample from each group, so that the two distributions were the same. All further analysis in this section is carried out on these matched samples (which are referred to as the Y9 and Y5 matched samples, respectively).

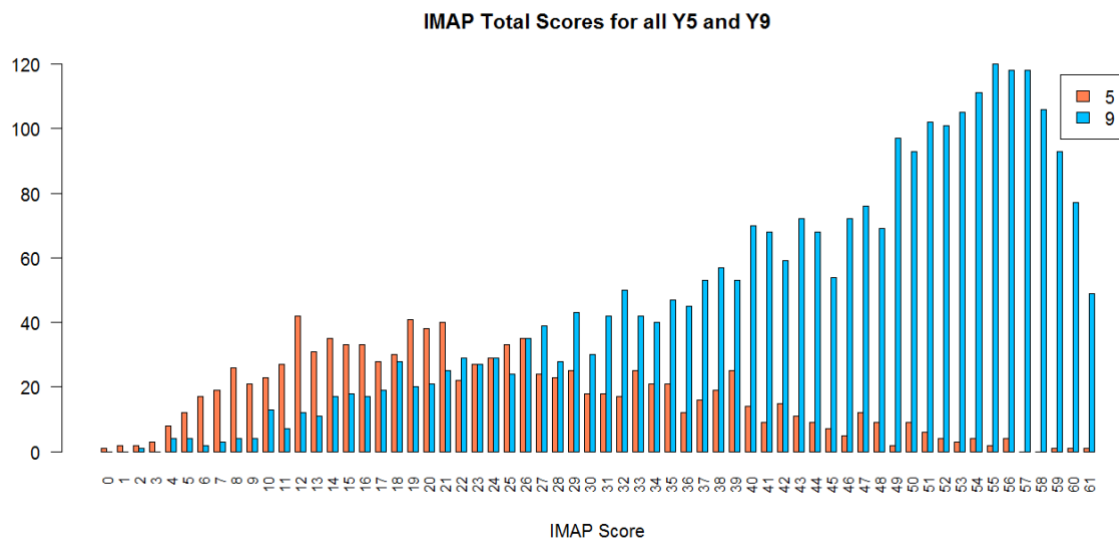


Figure 1: IMAP Number test total scores for all pupils from Y5 and Y9

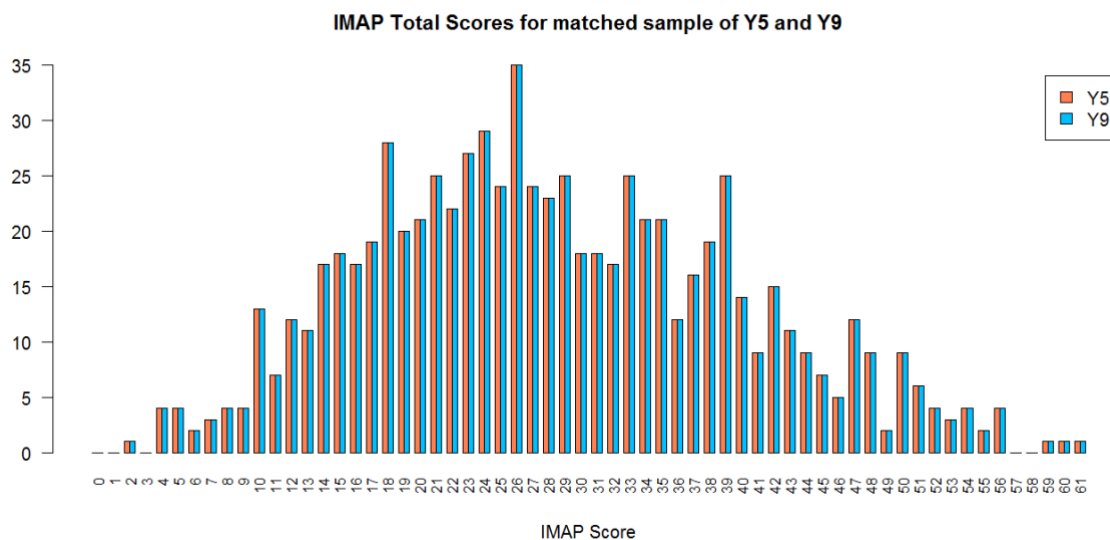


Figure 2: IMAP Number test total scores for matched pupils from Y5 and Y9

The matched samples comprised 759 pupils from each year group, and their score distributions are shown in Figure 1.2. In Table 2, we compare the distributions for the two matched samples. As can be seen from Table 2, the Y9 sample includes a small number of pupils with high scores on the IMAP Number test.

The distributions of scores on the arithmetic recall section for the matched subsamples (matched on total score from the 61 item IMAP test) are shown in Table 1 and Figure 3. We can see that the Y9s are slightly higher scoring on arithmetic recall than the Y5s, even though they are matched on their performance on the main IMAP Number test (which excludes the arithmetic recall score). The mean scores for the two groups are 12.5 and 13.5 for Y5 and Y9, respectively, corresponding to a standardised effect size difference of 0.16 (95% CI 0.06 to 0.26).

Table 1: Overview of the matched Y5 and Y9 samples

	Y5 (N=759)		Y9 (N=759)	
	Mean	SD	Mean	SD
Gender [% female]	49%	-	53%	-
Mathematics / Number [IMAP Number Score]	28.3	11.4	28.3	11.4
Arithmetic recall [Timed]	12.5	6	13.5	5.9
Maths Confidence [Beginning of test]	7.1	2.2	5.7	2.2
Maths Confidence [End of test]	7.5	2.3	5.9	2.4

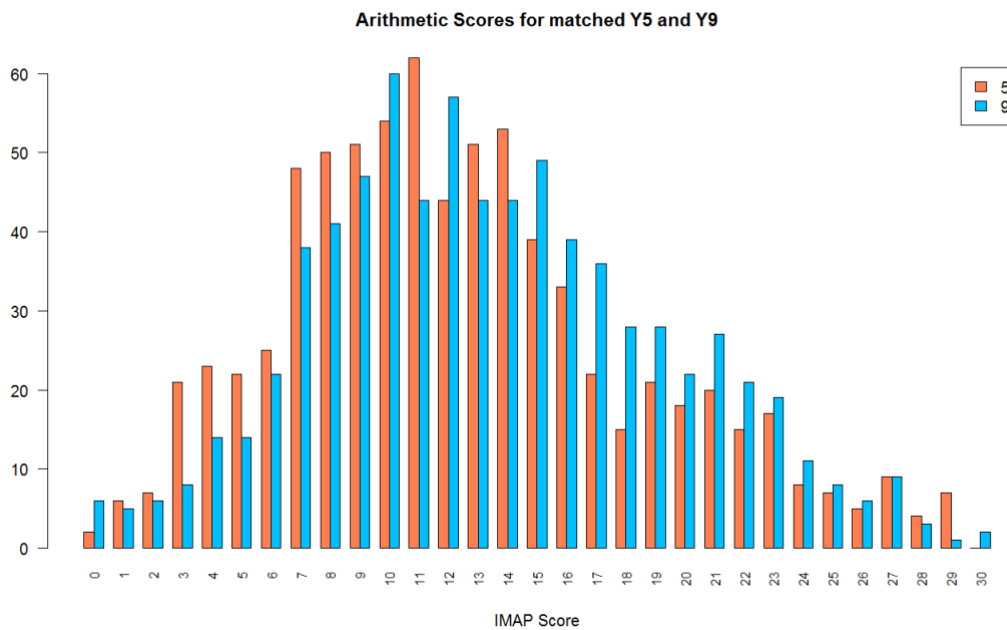


Figure 3: Distributions of total scores on the Arithmetic timed number facts test, for the matched sample

Results RQ1

What mathematics do low attaining Year 9 pupils understand and what are their particular strengths and weaknesses in number, multiplicative reasoning and algebra?

In Table 2, we present an overview of the Y9 pupils' performance overall and broken down by quintiles (within this low attaining sample). In order to enable comparison across the topics, we present the mean facility for each section. This table is also reproduced in the main report. In Figure 4, these comparisons are presented graphically using the mean facilities relative to the overall IMAP Number test mean facility. Aside from quintile 5, which is likely to be subject to ceiling effects, the profile for each quintile is broadly similar.

Table 2: An overview of the performance of the Y9 matched sample

Topic [# of items]	Mean total score	Overall mean facility	Mean facility by Quintile				
			Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
IMAP Number Overall [61]	28.26	0.46	0.34	0.56	0.67	0.79	0.90
<i>Elements of the overall IMAP Number score</i>							
A: Area & arrays [6]	2.65	0.44	0.34	0.53	0.62	0.71	0.86
DF: Derived facts [7]	3.66	0.52	0.42	0.61	0.66	0.80	0.92
E: Estimation [2]	0.79	0.40	0.24	0.55	0.63	0.79	0.95
F: Fractions [11]	5.76	0.52	0.39	0.63	0.76	0.86	0.91
IC: Integer calculation [4]	2.69	0.67	0.58	0.77	0.80	0.90	0.93
NL: Number lines [5]	2.01	0.40	0.27	0.49	0.62	0.74	0.92
P: Percentages [4]	0.91	0.23	0.12	0.26	0.42	0.53	0.83
PV: Place Value [11]	6.01	0.55	0.40	0.68	0.80	0.91	0.96
R: Ratio [6]	2.03	0.34	0.19	0.45	0.58	0.70	0.89
SC: Select a calculation [5]	1.76	0.35	0.23	0.45	0.57	0.69	0.78
Arithmetic fact recall [30]	12.51	0.45	0.36	0.53	0.60	0.68	0.71

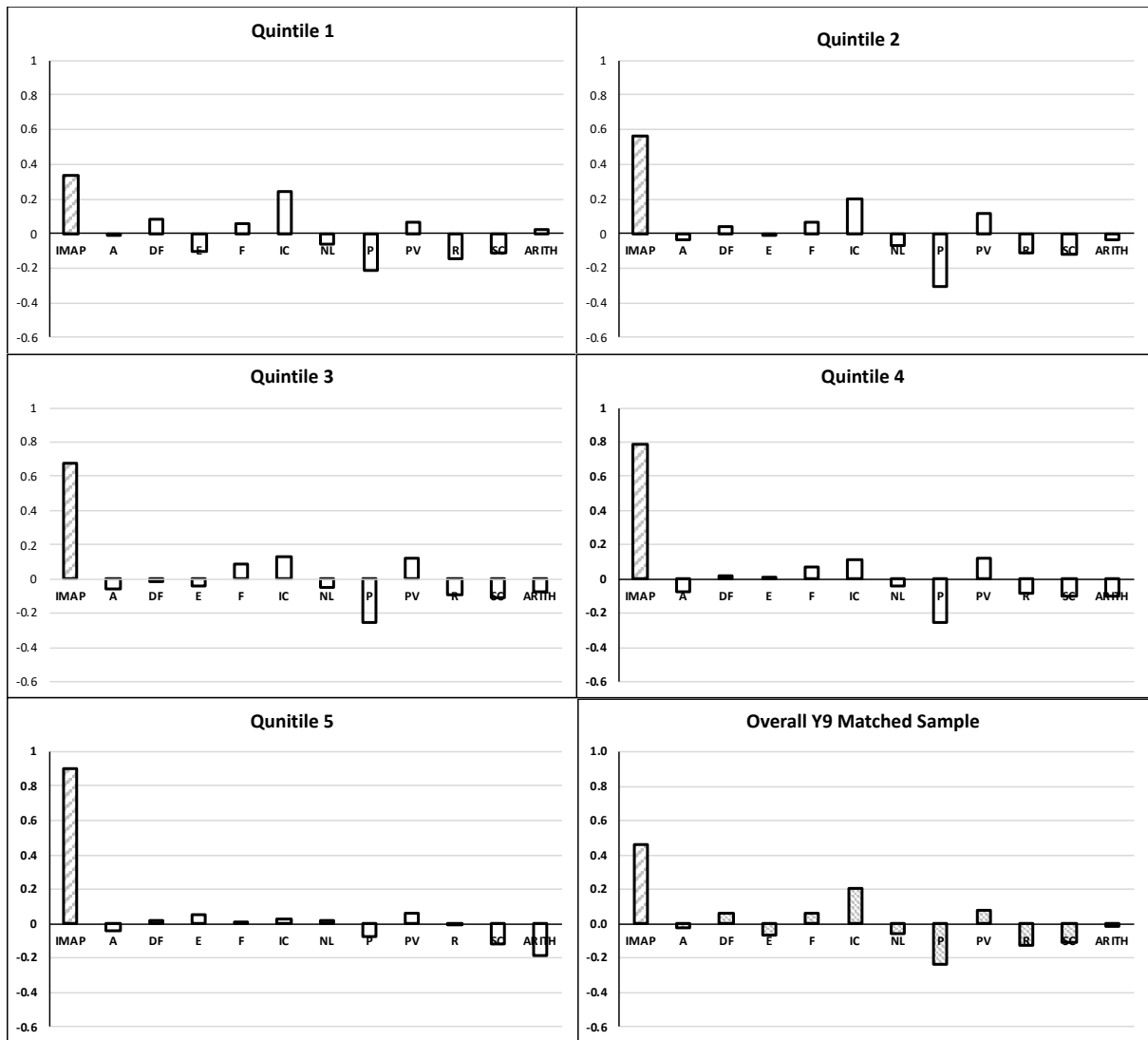


Figure 4: The performance of the Y9 matched sample by quintile, showing the mean facility for each topic relative to the overall IMAP Number score (i.e., topic mean facility less overall mean facility)

The matched samples enable a comparison of the performance of the low attaining Y9 matched sample with a group of middle and high attaining Y5 pupils, whose overall performance on the test is similar to the Y9 low attainers. We classify this group of pupils as middle and high attaining relative to the Y5 cohort of pupils as a whole as selected by their schools. This comparison is presented in Table 3, which is an extended version of the equivalent table in the main report..

This comparison indicates that the matched Y9 sample performed better on arrays and area (Cohen's $d=0.22$), percentages ($d=0.23$) and, as already noted, arithmetic recall ($d=0.16$), whereas the matched Y5 sample was stronger on derived facts ($d=0.22$) and select a calculation ($d=0.15$). Effect sizes were not significantly different from zero for estimation, fractions, integer calculation, number lines, place value or ratio. An inspection of the six array and area items indicates that the Y9 superiority resulted from the two numeric (dimensionless) area items (items 54 and 55) and, in fact, the matched Y5 sample performed better on the remaining four array items (see Appendix 3). It is important to appreciate that in our design, in which the two groups were matched on overall score on the IMAP test, there is a compensatory effect (i.e., relative strengths on a topic are compensated by relative weaknesses elsewhere within the test). This does not apply to the arithmetic recall test, since scores on this (sub)test were not matched.

Table 3: A comparison of the performance of the Y5 and Y9 matched samples.

Topic	Score				Difference Y9-Y5			Correlation with Number (Main IMAP test)		Correlation with Arithmetic (Timed test)	
	Y5 matched sample	SD	Y9 matched sample	SD	Cohen's d	95% CI	p -value	Y5	Y9	Y5	Y9
Number overall (Main IMAP score)	28.26	11.39	28.26	11.39	0	(-0.10, 0.10)	-	-	-	.69	.69
<i>Elements of overall Number score</i>											
Area & arrays [6 items]	2.33	1.44	2.65	1.46	-0.22	(-0.32, -0.12)	<.001	.66	.67	.47	.53
Derived facts [7]	4.10	1.95	3.66	2.06	0.22	(0.12, 0.32)	<.001	.68	.57	.44	.33
Estimation [2]	0.75	0.75	0.79	0.78	-0.06	(-0.16, 0.05)	.280	.56	.55	.39	.39
Fractions [11]	5.78	2.45	5.76	2.50	0.01	(-0.09, 0.11)	.880	.85	.82	.62	.58
Integer calculation [4]	2.77	0.95	2.69	0.94	0.09	(-0.01, 0.19)	.080	.64	.61	.49	.50
Number lines [5]	2.06	1.59	2.01	1.53	0.03	(-0.07, 0.13)	.590	.68	.64	.49	.43
Percentages [4]	0.69	0.88	0.91	1.03	-0.23	(-0.33, -0.13)	<.001	.59	.62	.38	.40
Place Value [11]	5.95	2.87	6.01	2.89	-0.02	(-0.12, 0.08)	.680	.86	.84	.62	.57
Ratio [6]	1.89	1.58	2.03	1.73	-0.08	(-0.18, 0.02)	.110	.64	.73	.35	.48
Select a calculation [5]	1.95	1.31	1.76	1.33	0.15	(0.04, 0.25)	<.001	.69	.68	.46	.45
Arithmetic score (Timed section)	12.51	6.01	13.47	5.88	0.16	(-0.26, -0.06)	<.001	.69	.69	-	-

Results RQ2

Do low attaining pupils understand mathematics in qualitatively different ways from middle and/or high attaining pupils?

In this analysis, we compared the Y9 and Y5 matched samples in order to investigate the extent to which the mathematical profiles of the two groups are similar or different, and whether any differences are sufficient to be classed as potential threshold concepts.

To explore whether we had evidence for different profiles of relative strengths and weaknesses for the Y9s and the Y5s, and to identify potential threshold concepts (Cousin, 2006), we addressed the following sub-questions:

- RQ2.1: In what ways do the factor and cluster structures of the IMAP test differ for the two matched groups (the Y9 low attainers and the Y5 middle and high attainers)?
- RQ2.2: How do the relative strengths in mathematical competence and arithmetic fact recall differ for the two groups?
- RQ2.3: How do the strengths of the relationships among various subscales and topics within the tests differ for the two groups?
- RQ2.4: To what extent is performance on different test items conditionally dependent on performance in key sections, specifically number lines? In other words, are there thresholds of performance on the number lines section that act as a prerequisite or enabler (i.e., a threshold concept) for expected success on other items? If so, are these dependencies different for high and low attainers?

RQ2.1 In what ways do the factor and cluster structures of the IMAP test differ for the two groups (Y5 and Y9)?

For this sub-question, we used a number of different analytic approaches: factor analysis, cluster analysis, comparison of inter-item score correlations, and comparison of inter-item Rasch residual correlations. As above, these analyses were applied to the matched dataset.

The factor analysis was conducted on the 61 items from the IMAP number test using principal components analysis (PCA) with oblimin rotation. The data were found to be suitable for PCA (Kaiser-Meyer-Olkin value = .93, Bartlett's test of sphericity: $\chi^2(1830) = 20065, p < 0.001$).

We found that extracting six factors using PCA with oblimin rotation provided the most defensible and interpretable reduction. These factors accounted for 18%, 4%, 3%, 3%, 2% and 2% of the variance, cumulatively accounting for 33% of the variance. The factors, together with their highest loading items are shown in Table 4.

Table 4: Six-factor structure of IMAP Number, from PCA with oblimin rotation. Items with highest factor loadings reported.

Factor	Items
Ratio	Q47 A soup recipe for 3 people needs 8 carrots. How many carrots are needed for 9 people? Q49 A soup recipe for 3 people needs 2 onions. How many onions are needed for 9 people? Q56 These two ticks are exactly the same shape. Find the length of the red part.
Area	Q54 What is the area of the shape below? (3 by 4 gridded rectangle) Q55 What is the area of the shape below? (6 by 10 un-gridded rectangle) Q34 Enter the missing number in the box. $1/3 = 2/?$
Number Lines	Q6 Look at the number line. What number is the arrow pointing to? Q8 Look at the number line. Click and drag the arrow so that it points to the number six thousand and twenty-five. Q9 Look at the number line. Click and drag the arrow so that it points to the number six thousand one hundred and eighty.
Derived Facts	Q20 Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $57+86$ Q21 Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $860+570$ Q19 Look at this calculation. $34+28=62$. Find a quick way to work out the answer to $34+29$?
General number	Q11 What is 1 less than 200? Q51 What is half of 16? Q1 A shirt costs £20. Alex buys 3 shirts. How much does this cost?
Whole number bias	Q43 Click on the larger fraction. $3/7, 5/7$ Q42 Click on the larger fraction. $1/4, 3/4$ Q44 Click on the larger fraction. $3/5, 3/4$ [Negative loading]

Ratio: All of the six ratio items loaded on this factor with loadings of between .36 and .66. In addition, two of the four percentage and two of the six array items had factor loadings of between .37 and .43.

Area: Just two items loaded on this factor both related to numerical (dimensionless) area and drawn from the CSMS Algebra test with factor loadings of .78 and .79. When constructing the test, it was hypothesized that the area items, both drawn from the CSMS algebra test, were elements of a broader topic, area and arrays, together with four items based on diagrams of arrays. Only one of these array items, Q58 (*Here is a 12 by 10 array. How many dots are there?*) had a weak loading (.30). Four fractions items also loaded on this factor, although the loadings (between .34 and .43) were small compared to the area loadings.

Number lines: All five number lines items loaded on this factor with factor loadings of between .41 and .69. Additionally, three place value and one percentage items loaded on this factor (between .31 and .39).

Derived facts: All six of the derived facts items loaded on this factor with loadings of between .47 and .77. No other items loaded on this factor.

General number: Twelve items loaded on this factor from a range of the test topics, place value, integer calculation, fractions and arrays). Factor loadings ranged from .31 to .52.

Whole number bias: Three items loaded on this factor, all of which are binary comparisons of two fractions (*Click on the larger fraction, 1/4 or 3/4; 3/7 or 5/7; 3/5 or 3/4*) with loadings of .80, .81 and -.39, respectively. These were the only items with significant misfit in the Rasch analysis. One plausible explanation for the positive and negative loadings is that a distinct group of pupils answered the first two correctly by simply comparing the numerators and ignoring the denominator, whereas these same pupils selected 3/5 as larger than 3/4 by incorrectly comparing the denominators. This relates to a well-known misconception that pupils believe that the numerators and denominators can be treated as separate whole numbers (Siegler et al., 2010).¹

Examination of the patterns of inter-item correlations and of residuals from the Rasch model shows them to be broadly in line with what would be expected from a test that fits the unidimensional Rasch model well, confirming a clear unidimensional latent trait. Some groups of items do show local item dependence (LID), in other words they are more highly correlated with each other than would be expected. The items that group together in this way are generally on the same topic or indicate the same misconception, and these same groups or sub-dimensions emerge from the PCA.

Crucially for our investigation, the patterns of factor structure are very similar for the Y5 (high attaining) and Y9 (low attaining) matched comparison groups. We did not find any evidence of substantive differences in the relative strengths of the two groups in their responses to a range of mathematical topic areas and skills within the IMAP Number test. All our analysis is consistent with a view of low attainment as characterised by delay rather than qualitatively

¹ Siegler, R. S., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., et al. (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide* (NCEE #2010-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

different performance. The Y9 low attainers seem to be broadly similar to matched middle and high attaining Y5 group in terms of the broad profile of things they know and can do, it is just that their general mathematical progress is some four years behind.

The factor analysis does indicate that, as predicted in the literature (e.g. Dowker, 2015), both groups, the Y9 low attainers and the Y5 middle and high attainers, are heterogeneous; in other words, individuals within both groups have a range of strengths and weaknesses. Bartelet et al., (2014) indicate that low attainers may form clusters characterised by particular strengths and weaknesses, such as a strength in number lines. However, results from a cluster analysis of these matched samples indicate at best weak clusters that appear to be quite unstable, so are not reported here.

RQ2.2 How do the relative strengths in mathematical competence and arithmetic fact recall differ for the two groups?

As noted above, we found that the Y9 low attaining group performed better than the Y5 group on arithmetic fact retrieval. This was somewhat surprising, since Geary's (2011) review suggests that a proportion of low attainers have particular difficulties with factual recall and that this may differ across the attainment range. We investigated this by comparing the performance on the main IMAP Number test (on which the Y5 and Y9 samples were matched) and the arithmetic fact recall test. A scatter plot showing scores on both sections for both groups (Y5 in red, Y9 in blue), together with the trend line for each (from LOESS) is shown in Figure 5.

The relationship between scores on these two tests is similar for both year groups, and, aside from the highest scores, the Y9 fit line on the graph is below the Y5. So, for a given score on the IMAP Number test, the Y9 low attaining pupils performed better on arithmetic fact recall than Y5 middle and high attainers. The gap opens up to become a statistically significant difference of around two points on the arithmetic recall scores for IMAP Number scores between about 30 and 40.

Our initial hypothesis was that the Y5s would be relatively stronger on fact retrieval (i.e., the timed Arithmetic test) than the Y9s, as we hypothesised that a lack of facility in recalling basic number facts (times tables and number bonds) could be a factor in explaining why the Y9s were low attaining. On the basis of this evidence, that hypothesis is not true. Indeed, there is some, rather weak, evidence to suggest that the highest attainers within the low attaining Y9 group are relatively strong on arithmetic recall. It may be that pupils categorised as low attaining on entry to secondary school are provided with a mathematical diet strongly weighted towards arithmetic, leading to some general improvement on arithmetic at the expense of other mathematical areas.

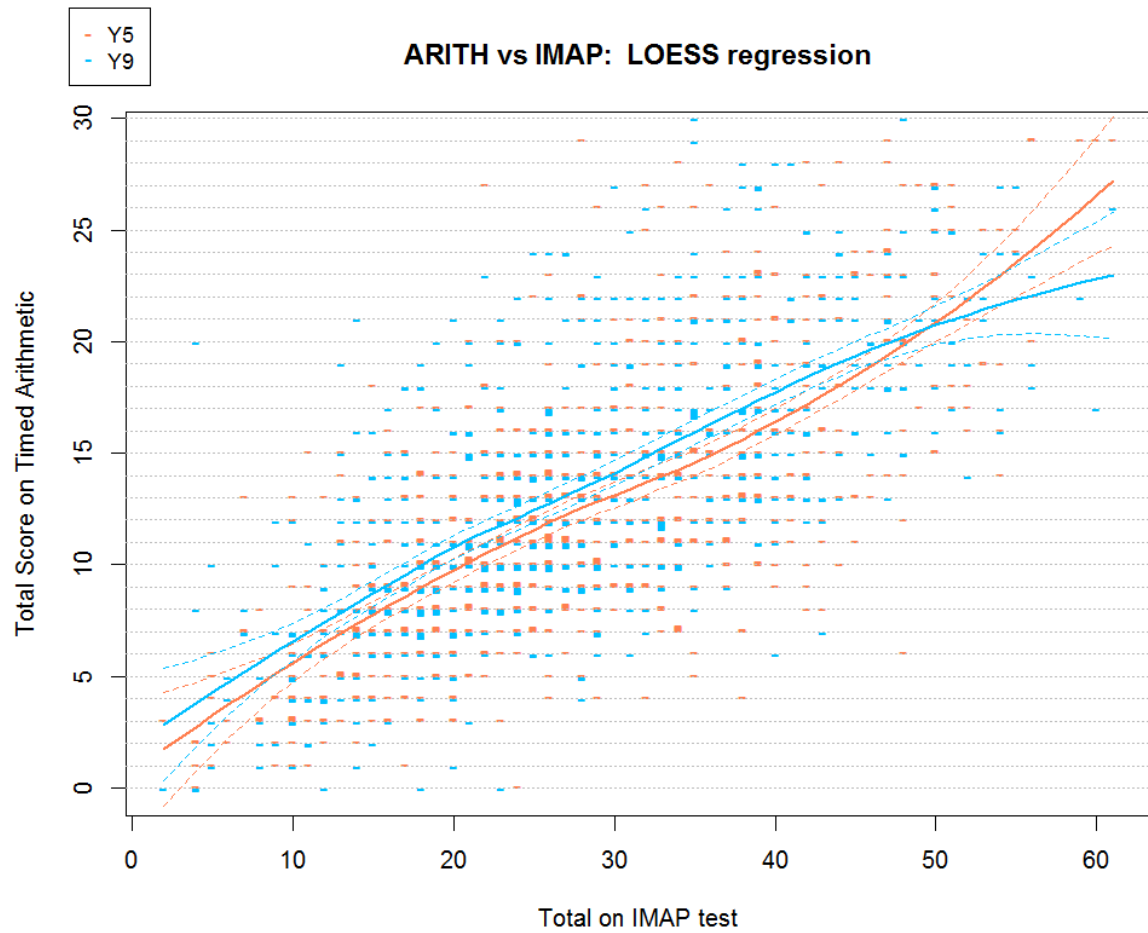


Figure 5: Scatter plot of timed Arithmetic against total on the IMAP Number test for Y5 (red) and Y9 (blue), with LOESS regression lines and 95% CIs

RQ2.3 How do the strengths of the relationships among various subscales and topics of the tests differ for the two groups?

In this analysis, we investigated the relationship between specific subsections of the IMAP Number test, such as the Derived Facts items (Q19-25), and either the total on the rest of the IMAP Number items or the score on the arithmetic recall test, in order to examine whether there were any differences within the two groups. We adopted a similar approach to that used for the last sub-question.

For most of the comparisons we have made, we found that the relationships between different sections and subsections were very similar for Y5 (high attaining, relative to their cohort) and Y9 (low attaining, relative to their cohort) pupils. One exception was the relationship between the Derived Facts section and Arithmetic scores, as shown in Figure 6.

On closer examination, we see that the divergence of the two fit lines for medium-high Arithmetic scores is a result of a relatively small number of mis-fitting pupils (about 40 pupils) who score at most 1 out of the 7 Derived Facts items (hence are in the bottom 16% of the sample on that measure) but at least 15 on the Arithmetic test (hence are in the top 30% on that). The vast majority of these misfits (33 of the 40) are in Y9. Removing those 40 cases makes the two lines coincide. From this, it is hard to say whether this would be a robust, replicable finding, but it certainly warrants further research.

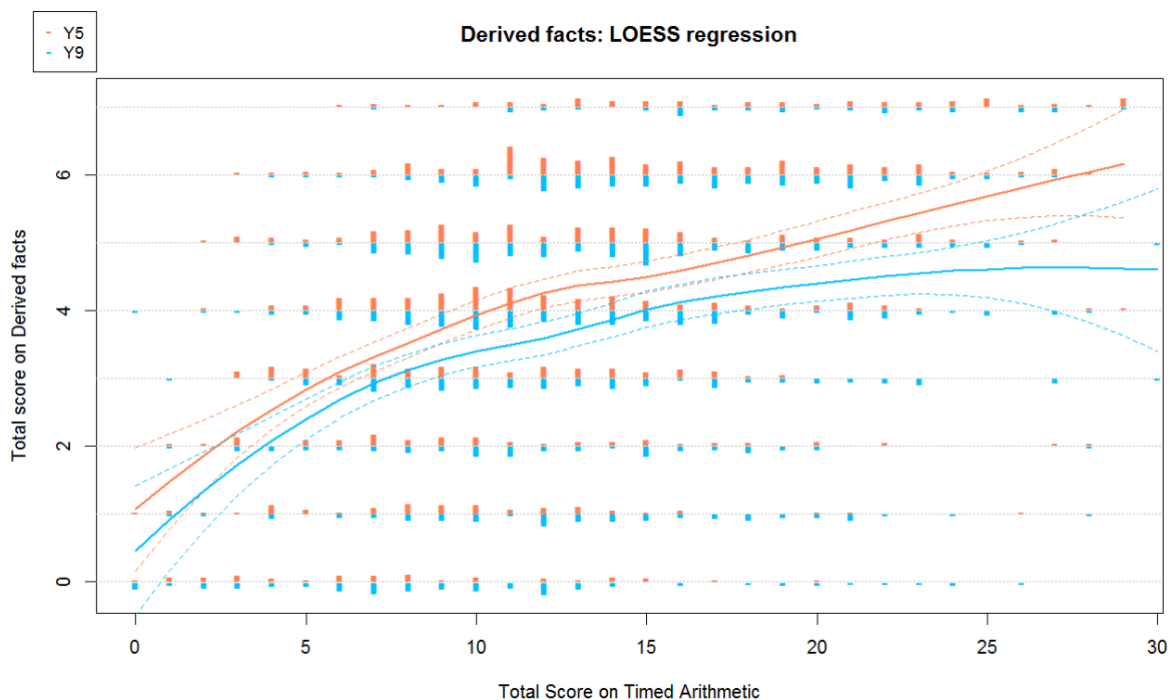


Figure 6: Scatter plot of Total on Derived Facts section against total on the Timed Arithmetic section for Y5 (red) and Y9 (blue), with LOESS regression lines and 95% CIs

RQ2.4: To what extent is performance on different test items conditionally dependent on performance in key sections, specifically number lines? Are there differences between high and low attainers?

The focus of analysis for the final sub-question is on the Number Lines section and whether it can be seen as a ‘threshold concept’ that unlocks performance in other areas of mathematics. In other words, are there thresholds of performance on the Number Lines section that act as a prerequisite or enabler for expected success on other items? Although we did not find statistically significant differences overall between the two groups on Number Lines, we judged it necessary to investigate this further, because the literature indicates that increased number line knowledge is associated with increased mathematical performance (e.g., Schneider et al., 2018) and that number lines may be key to pupil learning (e.g., Siegler et al., 2010).

We investigated this by examining the item characteristic curves for the number line items, and additionally using logistic regression.

Item characteristic curve against section score

One simple approach is to look at item characteristic curves (ICCs) for each item in the IMAP Number test against Number line (NL) score. These graphs show the proportion of pupils getting the item correct among those with each possible score on the NL section (from 0 to 5), split by year group (Y5/Y9). One manifestation of a threshold effect would be to see a discontinuity in the curve, where the probability of a correct response to an item increases dramatically at a particular NL score. This would suggest that a person’s probability of getting the item right would be noticeably different for those below a certain score on NL than for those above that threshold.

A selection of these graphs is shown in Figure 7. Overall, it is not clear from visual inspection of all these graphs that any such discontinuities exist.

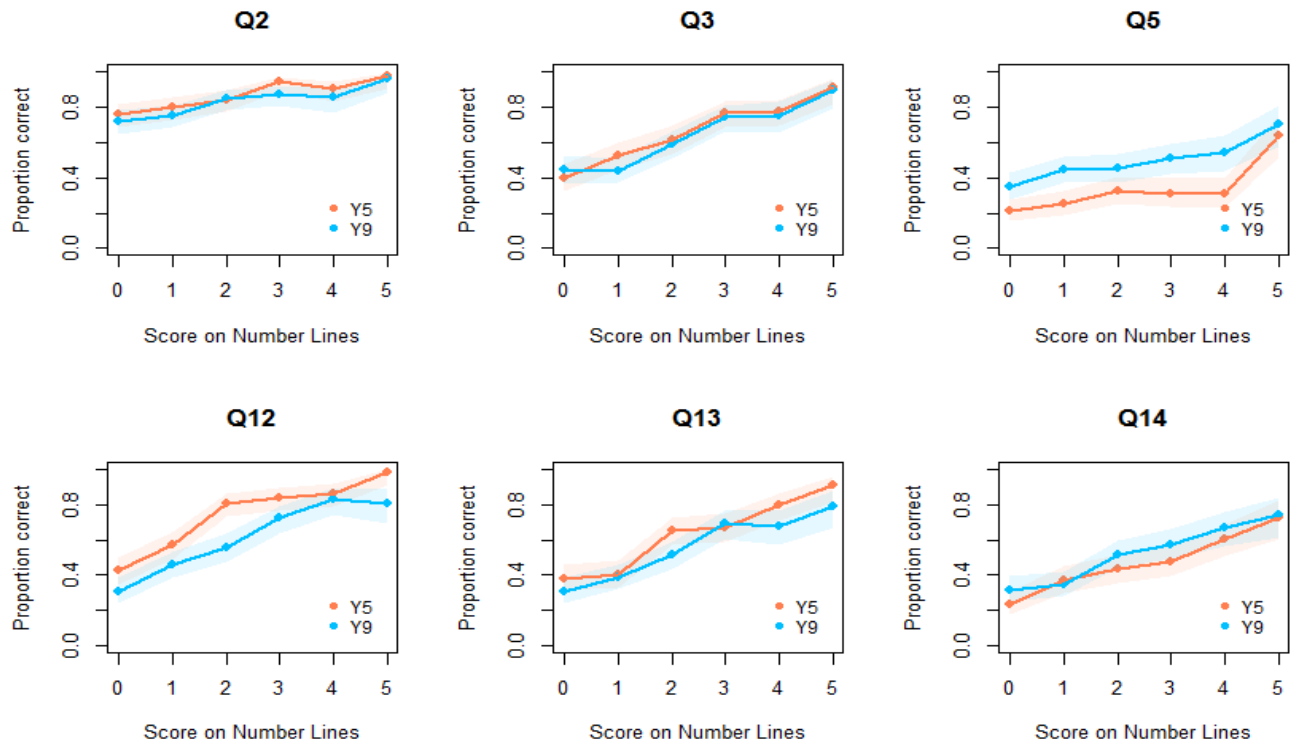


Figure 7: A selection of item characteristic curves (ICCs) for each item in the IMAP Number test against NL score

A further question concerns whether the dependency on Number Lines is different for Y5 and Y9. Here we are looking for differences in the slopes of the two lines. Given the number of comparisons made here, there do not seem to be any substantive differences between Y5 and Y9 in the relationship between performance on any item and performance in Number Lines. It is true that some items seem to favour Y5 (for the same NL score, Y5 are more likely to get the item correct, for example Q12), while others favour Y9 (eg Q5). By design, with the matched dataset, these differences balance out overall.

Logistic regression

A second way we can approach the question of conditional dependence on Number Lines is to compare the predictive power of logistic regression models with different predictors (a mediation analysis).

A logistic regression model attempts to predict the outcome on a particular item (either a score of 0 or 1) from explanatory variables entered into the model. In fact, the model estimates the probability of each pupil giving a correct answer as a function of those variables. One way to assess the power of the model is to ask how often it predicts accurately. If the predicted probability of a correct response is greater than 0.5 and the person gets it right, or if the probability is below 0.5 and they get it wrong, then we can class that as an accurate prediction. Across the sample, the proportion of accurate predictions gives a measure of the power of the prediction.

If we compare the accuracy of the predictions from a model that uses NL score as an explanatory variable with those from a ‘null model’ prediction, we will have an indicator of the additional explanatory power offered by performance in Number Lines. The null model here is the best prediction we can make without knowing anything about each individual pupil. We base this purely on the item facility: if 90% of the sample get an item right (or any percentage above 50%), then we predict everyone will get it right and the prediction will be accurate 90% of the time. Clearly, for items with facilities close to either 0 or 100%, the scope for improving on the null model prediction is somewhat limited. However, when the facility is close to 50% we should hope to see some improvement from including a predictor such as NL score.

A further comparison we can make is between the accuracy of predictions from models that include different kinds of explanatory variables. For example, we might expect NL score to improve accuracy above the null model, but does adding NL increase explanatory power in a model that already includes another variable, such as the total score on the rest of IMAP Number (IMAP-NL), the total from the timed Arithmetic fact recall section (ARITH) or a prior attainment measure such as KS1?

The results of these analyses may be hard to interpret for a number of reasons. For example, in the case of IMAP-NL, the amount of information in the IMAP-NL score is rather more than in the NL score: the former contains 56 items, the latter only 5. Even if, in theory, an understanding of Number Lines is a crucial prerequisite for performance on certain items, a measure of Number Lines understanding that contains only 5 items may not produce a strong predictor. We would expect a measure based on 56 items to dominate and would probably not be surprised if the additional explanatory power of NL were quite small. To some extent, the same argument applies to ARITH, though with 30 items in that measure, the balance of weight of information is slightly less one-sided.

One way to contextualise the explanatory power of the NL score is to compare it with the predictive power of other sections of comparable length within the IMAP test. We can do this either with sections based on genuine topic groups (such as Ratio, Arrays, Derived Facts or Percentage) that have similar lengths, or with an artificial section consisting of a random selection of 5 other items from the test.

We have run all these comparisons. For the other topic sections – Ratio (6 items), Arrays (6 items), Derived Facts (7 items) and Percentage (4 items) – we find that the predictive power of the remaining items is comparable to that of Number Lines. For the random pseudo-section (5 items chosen at random from the rest of the IMAP Number test) we find that the pseudo-section predicts slightly *better* than NLs. For illustration, we show the results of the comparison between Number Lines and Derived Facts in Figure 8.

Overall, therefore, we find no evidence that performance in Number Lines has any special place in explaining performance in other areas of mathematics. Moreover, this holds when each of the two year groups is analysed separately.

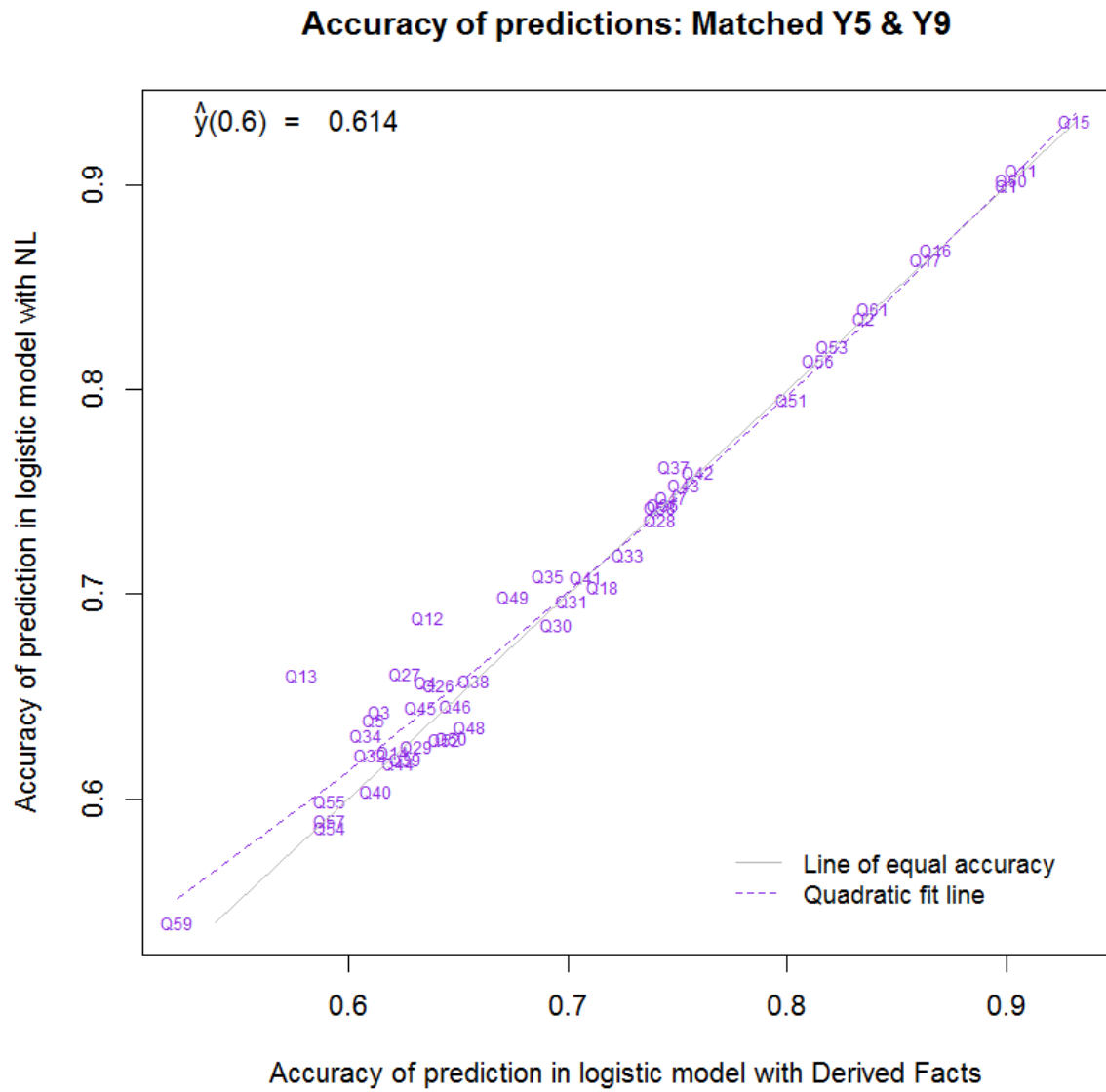


Figure 8: Comparison between the accuracy of prediction of response on each item of the IMAP Number test using Number Lines score (vertical axis) and Derived Facts score (horizontal axis). Prediction is from a logistic regression model with each item (excluding those in either section) as the DV and the section score as the IV

Research question 3:

To what extent do low attaining pupils' prior understandings of mathematics, and of particular mathematical topics, help to explain the existence of the attainment gap?

What is the relative contribution of these mathematical understandings in comparison to socio-economic status and other demographic factors?

In this analysis we addressed the following sub-questions:

- *RQ3.1 What factors determine success in GCSE maths?*
What factors explain attainment on our various measures of mathematical attainment (ICCAMS test scores in Y7, 8,9; KS3 teacher assessments)
- *RQ3.2 Are some areas of maths crucial?*
Are there specific items or sections of the ICCAMS test that are particularly predictive of later outcomes (eg GCSE maths grade or gaining A*-C)
- *RQ3.3 Can low attainment be overcome?*
What factors are associated with later success of pupils who were low attaining earlier in their schooling?

The results of these analyses are outlined and discussed in the main report. In this technical report, we present the technical details of these analyses in more depth.

Sample and variables

The sample analysed here is the 10,913 pupils who took one or more ICCAMS tests in Y7, 8 or 9 in 2008 or 2009 as part of the ICCAMS study. Pupils were randomly allocated to take two of the three ICCAMS tests (Algebra, Number, Ratio) and an attitude questionnaire. We constructed a variable 'iccams789' which is the latent trait estimate (from the Rasch model) for each pupil from all the assessments they took.

These pupils were matched with data held in the National Pupil Database (NPD), including their KS1 levels, KS2 levels, KS3 teacher assessments and KS4 (GCSE) results. Also included are a range of variables captured in the annual school census, with data for each term of their school career: Free School Meals (FSM) entitlement, Income Deprivation Affecting Children Index (IDACI), recorded absences, and number of days excluded. Other variables held in the NPD were not available to us for this analysis, including English as an Additional Language (EAL), Special Educational Needs (SEN) status and ethnicity.

Items from the ICCAMS attitudes questionnaire were grouped based on factor analysis to produce the following scales:

- *Self-efficacy*: a person's confidence and belief in their own ability in mathematics, not finding maths too hard or feeling too anxious about it and rating themselves high in relation to others (5 items, $\alpha=0.74$; labelled as 'seffic')
- *Intrinsic enjoyment*: enjoyment of mathematics for its own sake, particularly enjoying being challenged and having to work (3 items, $\alpha=0.55$; labelled as 'intrins')
- *Performance goals*: the extent to which a person is motivated by manifest achievement, particularly in comparison to others (2 items, $\alpha=0.77$; labelled as 'perfgo')

Results RQ3

RQ3.1: What factors determine success?

One way to explain and understand the attainment gap is to find the best predictors of attainment. If we can predict an attainment outcome, then we can say what characteristics are associated with doing well – or less well – on it. Factors that strongly predict who does well can also help to explain why some do better than others.

Relationships with individual variables

Figure 9 to 11 show density scatter plots for the relationship between each of our variables and GCSE maths outcome, along with the trend line (from LOESS) and overall correlation.

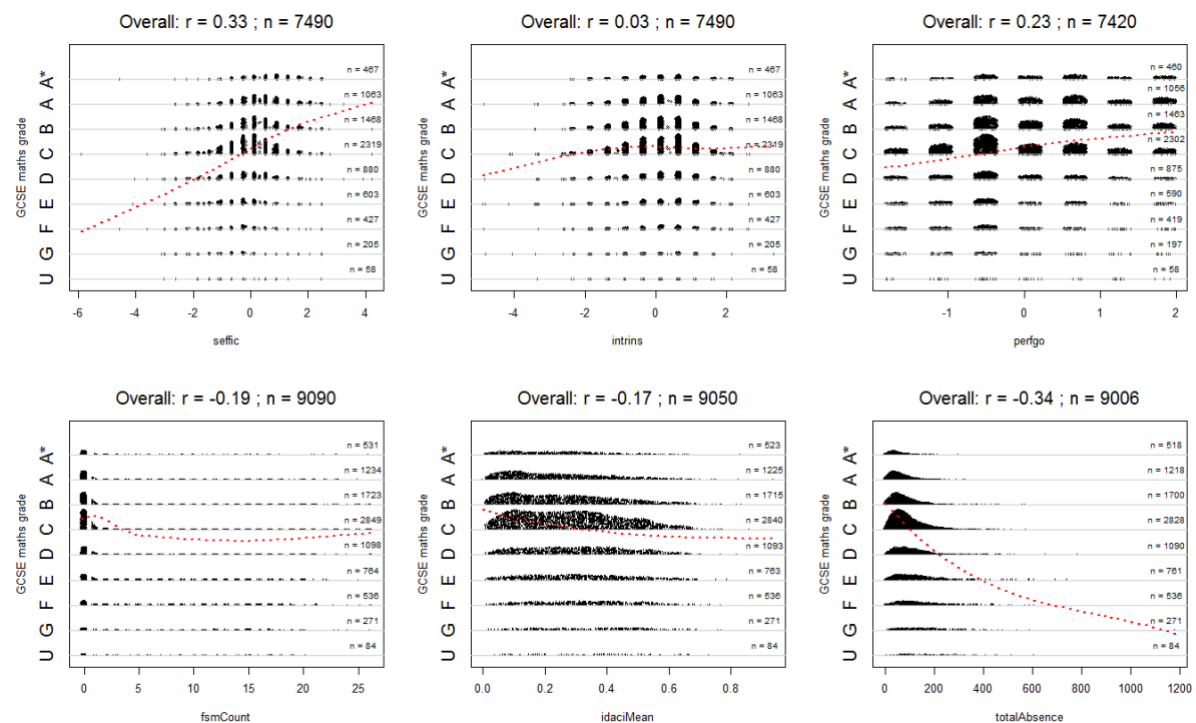


Figure 9: Density scatter plots for the relationship between GCSE maths grade and: self-efficacy; intrinsic motivation; performance goal orientation; no. of school terms eligible for free school meals; mean IDACI score over school career; total number of half days absent from school.

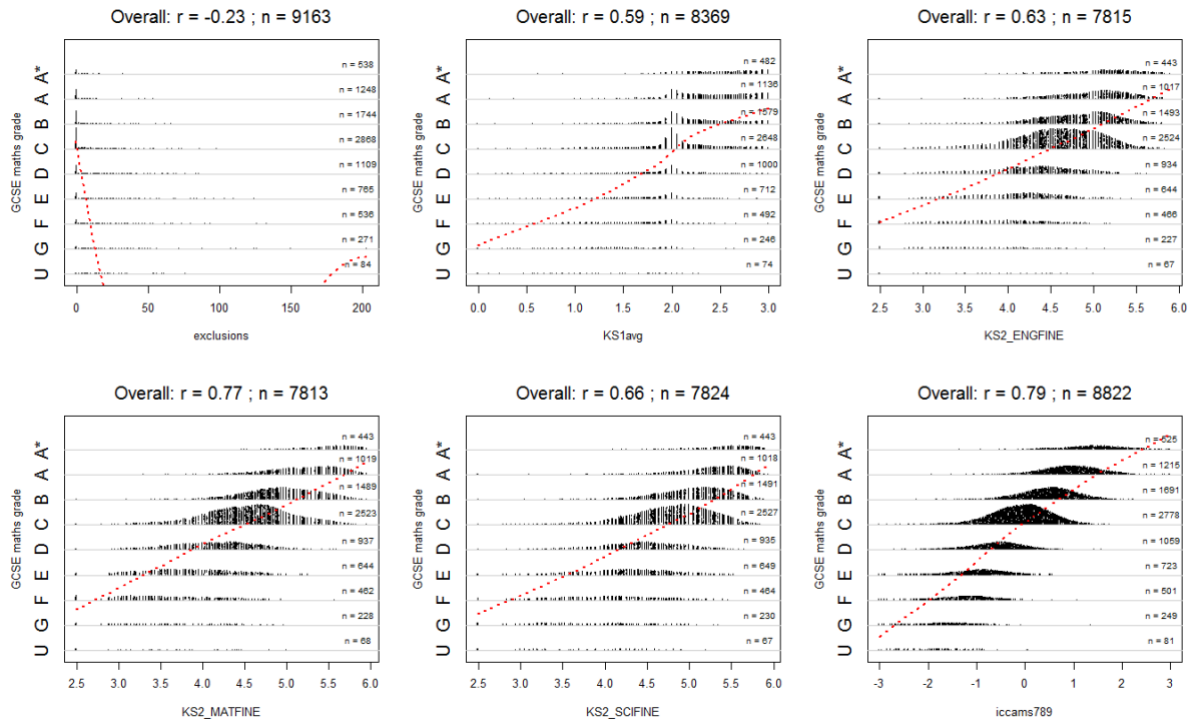


Figure 10: Density scatter plots for the relationship between GCSE maths grade and: no. of half days excluded from school; Key Stage 1 average fine level; KS2 English fine level; KS2 maths fine level; KS2 science fine level; ICCAMS score (Rasch estimate from all tests taken in either Y7, 8 or 9)

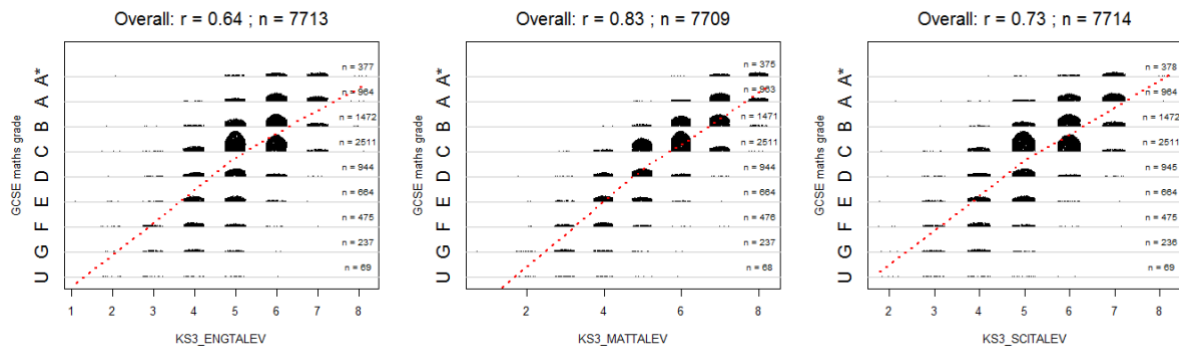


Figure 11: Density scatter plots for the relationship between GCSE maths grade and: KS3 English teacher assessment level; KS3 maths teacher assessment level; KS3 science teacher assessment level.

The strongest single predictors are KS3 maths teacher assessment level (KS3_MATTAEV), ICCAMS test scores (iccams789), and KS2 maths fine level (KS2_MATFINE), with correlations of 0.83, 0.79 and 0.77, respectively. Other prior attainment measures (KS3 and KS2 levels in science and English, KS1 average fine level) have correlations around the 0.6-0.7 range, so are also strong predictors.

Moderate predictions (correlations around 0.2-0.3) are found for the number of recorded absences (totalAbsence), self-efficacy (seffci), number of half-days excluded from school, (exclusions), number of terms eligible for free school meals (fsmCount), performance goal orientation (perfgo) and IDACI (idaciMean: an index of deprivation based on pupil's home postcode, recorded each term and averaged over all recorded values).

In order to simplify the analysis, we combined fsmCount and idaciMean into a single variable to represent socio-economic status: ses.² The correlation between this measure and GCSE maths grade is -0.22.

Other variables, including intrinsic motivation, the number of schools attended, the amount of variation in recorded IDACI scores and gender, had no correlation with GCSE maths grade.

Regression analyses

Multiple linear regression models allow us to see the combined effects of multiple predictors. We fitted the models progressively, adding in more predictors to investigate their cumulative explanatory power. We also fitted linear mixed effects (multilevel) models, with pupils nested within schools. Results were in line with those from ordinary least squares (OLS) regression. The predictors included in each model are as follows:

- Model 1: SES only;
- Model 2: SES, gender, no. of absences, no. of half-days excluded, no. of schools attended;
- Model 3: As Model 2, plus KS1 average fine level;
- Model 4: As Model 3, plus KS2 fine levels in maths, science and English;
- Model 5: As Model 4, plus KS3 teacher assessment levels in maths, science and English;
- Model 6: As Model 5, plus ICCAMS overall score (Rasch estimate from all tests taken);
- Model 7: As Model 6, plus self-efficacy, intrinsic motivation and performance goals.

Regression coefficients from these seven models, with progressive addition of independent variables, are shown in Table 5.

Many previous studies find socio-economic status (SES) variables to be a predictor of academic outcomes, and Model 1 confirms this relationship, albeit with a lower correlation than sometimes found. However, once a prior achievement measure was entered into the model (ie Models 3-7), the additional predictive power of SES drops to zero or close to it. This suggests that the impact of SES is already captured in those prior attainment measures, but also that the progress made from KS2 to GCSE is typically not related to socio-economic status (Model 4).

Similarly, we found gender to be only very weakly related to either outcomes or progress to GCSE.

In the full model (Model 7) we found that the overall predictive power of our variables was very strong (adjusted *R*-squared of 0.78, which corresponds to a multiple correlation of 0.88). As previously, the variables that capture prior attainment in maths are the strongest predictors. We note that there is some potential for collinearity among these predictors, so the exact coefficients and standard errors may not be wholly dependable; the coefficients of KS2/3 in English and science seem a little unstable, for example. Nevertheless, the finding that a combination of KS3 maths teacher assessment, ICCAMS test score and KS2 maths fine level provides a very strong prediction of GCSE grade in maths is robust. Interestingly, KS1 level is reduced to having almost no predictive power once KS2 is included.

² The single SES measure was calculated as the mean of the two standardised scores, then re-standardised and winsorised at ± 3 .

Among the non-cognitive variables that retain substantive predictive power once the attainment variables are included are exclusions and absences; both have standardised coefficients around 0.12. Translated into the scales on which each of the variables is measured, these are equivalent to saying that a reduction of one grade in GCSE maths would typically be associated with one additional day of exclusion from school, or to nine additional days of absence over a school career. These relationships are found after taking into account other factors such as prior attainment, SES and gender. However, it does not of course follow that these are causal relationships: it could easily be that some factor we have not observed causes both poorer GCSE performance and a propensity for exclusion or absence. Nevertheless, the strength of the relationship is such that there is a case for further research to investigate whether there are things schools could do to reduce exclusions and absences, and, if so, whether increased GCSE results ensue.

Somewhat smaller relationships, but still above the $p < 0.05$ threshold for statistical significance, are found with self-efficacy, intrinsic motivation and number of schools attended.

Table 5: Regression coefficients from seven regression models with different independent variables.
GCSE maths grade (the dependent variable) is standardised (sd=1)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
IVs	SES	Model 1 + Absence. Exclusions	Model 2 + KS1	Model 3 + KS2	Model 4 + KS3	Model 5 + ICCAMS	Model 6 + Self-Eff, Intrinsic Mot, Perf Goals
SES	-0.222 (0.010)*	-0.150 (0.010)*	-0.046 (0.009)*	-0.013 (0.007)	-0.008 (0.006)	-0.009 (0.006)	-0.007 (0.006)
Gender		0.033 (0.020)	0.067 (0.017)*	-0.041 (0.014)*	0.004 (0.012)	-0.005 (0.012)	-0.020 (0.013)
Absences		-0.347 (0.012)*	-0.265 (0.010)*	-0.201 (0.008)*	-0.124 (0.007)*	-0.118 (0.007)*	-0.121 (0.008)*
Exclusions		-0.317 (0.018)*	-0.268 (0.015)*	-0.211 (0.012)*	-0.133 (0.011)*	-0.128 (0.011)*	-0.128 (0.012)*
Num_Schools		0.068 (0.011)*	0.056 (0.009)*	0.043 (0.007)*	0.039 (0.006)*	0.034 (0.006)*	0.030 (0.007)*
KS1			0.537 (0.009)*	0.038 (0.011)*	0.001 (0.009)	-0.010 (0.009)	-0.010 (0.010)
KS2 Ma				0.581 (0.013)*	0.238 (0.013)*	0.181 (0.013)*	0.181 (0.014)*
KS2 Sc				0.083 (0.013)*	-0.005 (0.011)	0.001 (0.011)	-0.001 (0.012)
KS2 En				0.079 (0.013)*	-0.037 (0.011)*	-0.033 (0.011)*	-0.025 (0.012)*
KS3 Ma					0.436 (0.012)*	0.354 (0.012)*	0.335 (0.014)*
TA					0.180 (0.010)*	0.157 (0.010)*	0.156 (0.011)*
KS3 Sc					0.083 (0.010)*	0.062 (0.010)*	0.059 (0.011)*
TA						0.198 (0.011)*	0.197 (0.011)*
ICCAMS							0.034 (0.012)*
Self-eff							0.020 (0.007)*
Intrinsic mot							0.013 (0.007)*
Perf Goals							0.013 (0.006)
(Intercept)	-0.004 (0.010)	-0.057 (0.014)*	-0.071 (0.012)*	-0.025 (0.010)*	-0.062 (0.009)*	-0.048 (0.009)*	-0.044 (0.009)*
R-squared	0.049	0.189	0.440	0.666	0.764	0.773	0.780
df	9088	8315	8301	7313	7004	6813	5735

* p<0.05

RQ3.2: Are some areas of mathematics crucial?

To answer our second sub-question, we assessed the strength of relationship between each individual question in the ICCAMS test and pupils' subsequent GCSE maths grade. We used two different analytical approaches to address this question.

Relationship between individual ICCAMS items and GCSE grade

In the first approach, we calculate the point-biserial correlation between scores on that item (0,1) and the grade achieved in GCSE maths (coded as A*=8, A=7, B=6, etc). We expect these correlations to be attenuated by the item facility (items with facilities close to 0 or 1 will tend to have lower correlations) and by the distribution of GCSE grades. For comparison, therefore, we created a simulated set of responses to each question, to provide an indication of the expected correlation.

Some items were found to be correlated with GCSE maths grade a little higher than we would expect if they followed the model, while others were a little lower than expected. Comparing these correlations with the within-test discrimination of each item shows a pretty close match: the correlation between an item and GCSE grade reflects that item's correlation with the other items in its test. Hence items that are highly discriminating within their own test are also good predictors of GCSE, and conversely.

From this, it is not clear that there are particular mathematical topics, skills or items that punch above their weight in predicting subsequent GCSE performance.

Individual items to supplement the prediction from logistic regression

The second approach is designed to address the question of whether there is any evidence for threshold concepts: crucial parts of mathematical knowledge or understanding that are an enabler (or barrier) for future success.

We applied logistic regression models with a range of predictors for GCSE grade (dichotomised as A*-C or not), and then used this predictor with two subsets for each item in the ICCAMS test, for pupils who got that item right and wrong, respectively. This allowed us to compare the modelled chance of getting A*-C for pupils with the same background characteristics, depending whether they got each item right or wrong.

We compared these differences with the results of a simulation of random responses. Broadly speaking, the observed differences were very similar to those that arose from simulation, and this was true across all three ICCAMS tests. This suggests that there is no evidence of any specific items having particular predictive power for GCSE outcomes, over and above what is already captured in their prior attainment, attitudes, gender, absence and exclusions.

This is probably not surprising, given the weight of information in the combined predictors compared with that in a single item. However, our hypothesis was that there might be certain crucial elements of mathematical knowledge or competence that were a barrier to or an enabler of further learning. If this was the case, we might expect to see an additional effect in the form of different outcomes for those pupils who had mastered that learning in KS3 from those who had not. Controlling for all the other predictors is necessary here in order to rule out the competing explanation that those differences are just a reflection of general mathematical

attainment, since higher attaining pupils will both be more likely to get an item correct and to be successful at GCSE. However, it does make this a tough test for the existence of threshold concepts, particularly when, as here, those concepts are represented by a single item in the test.

RQ 3.3: Can low attainment be overcome?

Our third sub-question focusses specifically on low attainers. We looked at pupils who were low attaining in maths at different points in their school career and estimated their chances of achieving a grade C or above at GCSE. We then investigated how these chances may be different for pupils who differ on any of our additional variables.

Chances of low attainers with different characteristics achieving C+

Pupils in the first decile (lowest 10%) of attainment, according to their KS3 teacher assessment in maths, have only a 7% chance of achieving C+ at GCSE. For those in the top three deciles at KS3 (top 30%), over 99% achieve C+. The prospects for pupils who are low attaining at KS1 (i.e. aged 7) are not quite as firmly determined, though the relationship is still strong: of those in the bottom decile 23% go on to achieve C+, rising to 98% for those in the top decile. (See Table 6.)

Table 6: Percentage of pupils in each decile of performance on each of KS1 average level, KS2 maths fine level, ICCAMS Y7, 8 or 9 test score or KS3 teacher assessment who go on to achieve grade A-C in maths GCSE*

Percentage who achieve A*-C in GCSE maths, from each decile of prior attainment				
Decile	KS1avg	KS2_MATFINE	iccams789	KS3_MATTALEV
1	22.9	10.3	11.2	6.8
2	39.1	25.6	25.4	17.9
3	56.3	52.5	44.2	51.5
4	69.2	64.5	61.7	49.8
5	68.3	76.6	79.0	86.3
6	73.8	84.8	89.4	91.0
7	85.4	93.6	95.6	91.7
8	89.9	96.9	98.5	99.2
9	95.1	99.0	99.9	99.1
10	97.7	99.3	99.7	99.5

Figure 12 addresses the question: ‘If you are a low attainer on some measure (eg KS1), what are your chances of going on to achieve A*-C in GCSE mathematics?’ A further question we might ask is: ‘If we then also know some other piece of information about you (for example, your score on another assessment, or demographic information), how much difference does that make to our expectation?’ This is the question we address now.

In this section we look at the prospects of attaining A*-C in GCSE mathematics for those who are in the bottom 40% (approximately³) of attainment on some prior attainment measure. For these initially low attainers, we calculated the proportion who went on to achieve C or above at GCSE, depending on whether they also scored above or below average on each of our other measures. Figure 13 shows these proportions when KS1 is used to define ‘low attainers’ and the other variables we consider are dichotomised at the median. In other words, we split the

³ Some of the variables used to identify these low attainers are lumpy in their distributions, and hence it is impossible to separate the bottom 40% precisely. For example at KS1, over a quarter of pupils achieve an average level of exactly 2. Hence, we identify low attainers as those with KS1 average level strictly below 2, and end up with only 26% of the population of those for whom we have KS1 levels.

population into those in the top half (above the median) or bottom half (below the median) on that variable, and calculate the proportion who achieve A*-C for each half. Similar graphs were produced for low attainers defined by bottom 40% performance on KS2, ICCAMS and KS3, with similar patterns emerging.

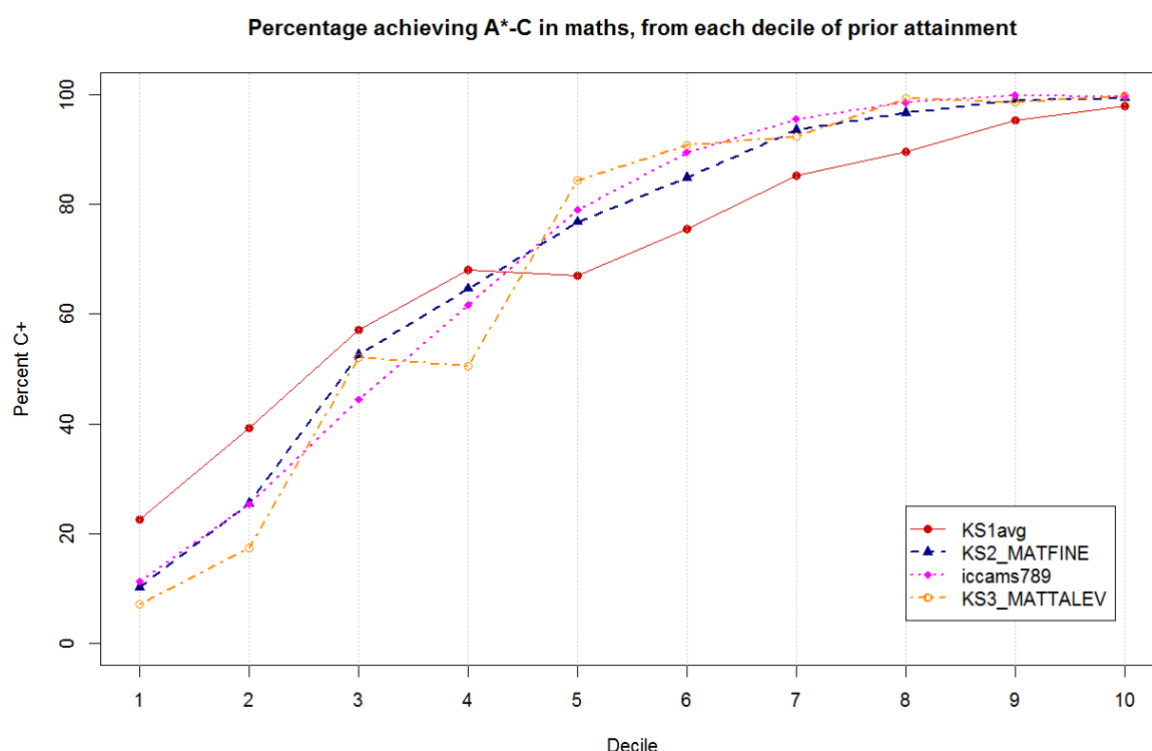


Figure 12: Graph showing results of Table 6. Percentage achieving A*-C in maths GCSE, from each decile of prior attainment on four measures

Figure 13 shows that being in the top half on some measures offers a very different prospect from being in the bottom half. For example, those whose scores on any of the ICCAMS tests put them in the top half have a better than 80% chance of going on to achieve C+ at GCSE, even though their KS1 levels were in the bottom 26% of the cohort. This compares with chances around 30% of gaining C+ for those whose ICCAMS scores were in the bottom half. A similar difference can be seen for performance in maths at KS2 and KS3, and a slightly smaller effect for performance in science and English: again those in the top half on these measures have around a 70-80% chance of success, while for those in the bottom half it remains around 30%.

Prospects for low attainers at KS1

To put this in context, the overall proportion of these low attainers (ie bottom 26% of the cohort on KS1) who achieve C+ at GCSE is 36%. Most of these low attainers at KS1 are also in the bottom half of attainment on the other measures: in the case of ICCAMS, KS2 and KS3, over 80% of low attainers on KS1 are in the bottom half on these measures too. However, for the 15-20% of pupils who are in the top half, their prospects are very different.

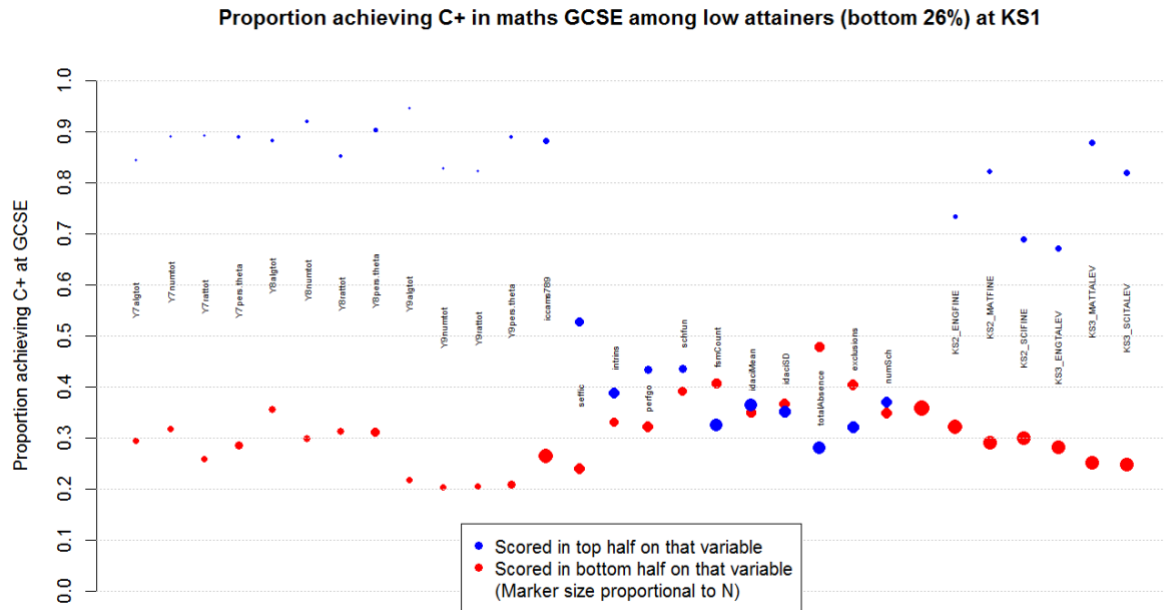


Figure 13: Proportion achieving C+ in GCSE maths among low attainers at KS1 with different characteristics on other variables

For the demographic and attitude variables, the picture is more mixed. First of all, none of these variables makes as much difference to the chances of getting C+ as the attainment variables. Nevertheless, for some of these variables, there is a significant difference. In particular, a pupil's self-efficacy ('sefffic') and number of absences ('totalAbsences') are both associated with differences of 25 and 20 percentage points, respectively, in the chances of getting C+ between those in top and bottom halves on these measures. The number of terms eligible for FSM ('fsmCount'), number of half-days excluded ('exclusions') and performance goal orientation ('perfgo') are each associated with differences of around 10 percentage points.

Research question 4:

What is currently known about the effectiveness of teaching strategies and approaches that address low attainment in secondary mathematics?

In this section, we provide additional technical detail about the literature review.

Methods of synthesising the literature

Our approach was to carry out second-order meta-analysis – i.e., meta-analyses of existing meta-analyses – and occasionally third-order meta-analyses, where we summarised the findings of existing second-order meta-analyses. Where possible, we conducted a quantitative meta-analysis of the first-order meta-analyses. All aggregation of effect sizes was carried out using the random-effects model, which is the more conservative option and does not assume that all of the effect sizes being combined are homogeneous. We performed our calculations using the *metafor* package in *R* (Viechtbauer, 2010; see <http://www.metafor-project.org/doku.php>).

The dataset

Our data set consists of a total of 107 items: 76 meta-analyses and 31 other relevant papers (mainly systematic reviews), written in English, relevant to the learning of mathematics of low attaining pupils aged 11-14, and published (or publicly available) between 1970 and August 2018. These were identified using searches of electronic databases, the reference lists of the literature obtained and our own and colleagues' knowledge of the research literature. The original search was conducted in 2016, then updated in February 2017 and again in September 2018. See Appendix 5 for the full dataset together with examples of excluded meta-analyses.

Coding and data extraction

Each paper was coded as a meta-analysis, systematic review or 'other literature', and details were recorded, including year of publication, author key words, abstract, content area, main focus, secondary focus, key definitions, research questions, ranges of effect sizes, any pooled effect sizes and standard errors, number of studies and number of pupils, age range, countries in which studies were conducted, study inclusion dates, any pedagogic or methodological moderators or other analyses, inclusion/exclusion criteria and quality judgments.

There is little consistency in the definition of low attainment in the literature. Indeed, we found only one meta-analysis which specifically targeted low attaining pupils in its title (Baker et al., 2002). Hence, our study includes meta-analyses (and systematic reviews) relating to "strugglers", pupils with mathematical learning difficulties or learning difficulties generally, "persistent low attainers", pupils "at risk" of underachievement and disadvantaged pupils. In order to reflect this, we adopted an "onion" approach (Coffield et al., 2004) whereby literature was coded as Level 1, 2 or 3, with Level 1 being the most relevant. Only Level 1 and 2 meta-analyses were considered suitable for inclusion in any quantitative meta-analysis. See Table 7 for an overview.

Table 7: An overview of the data set

Category	Number of key meta-analyses	Number of other meta-analyses	Number of additional studies	Total	Meta-analyses considered for aggregation	Description
Level 1	35	-	-	35	Yes	Meta-analyses focusing on or highly relevant to the teaching and learning of mathematics to low attaining pupils
Level 2	13	-	-	13	Yes	Meta-analyses focusing on the teaching and learning of mathematics and relevant to low attaining pupils
Level 3	-	28	31	59	No	Systematic reviews (and some meta-analyses) relevant to the teaching and learning of mathematics to low attaining pupils (including some generic pedagogical and some broad brush reviews). Includes some narrative or theoretical reviews providing insights from research or theory on key themes and strategies highlighted in analysis of Level 1 and 2 meta-analyses
Total	48	28	31	107	-	-

We assessed the methodological quality of the meta-analyses using six criteria, which we developed ourselves, informed by the PRISMA framework for rating the methodological quality of meta-analyses (<http://www.prisma-statement.org/>) and the AMSTAR criteria (Shea et al., 2009). For each meta-analysis, we graded each of our six criteria on a 1-3 (1 low, 3 high) scale.

For the strategies identified in the review, we assessed the strength of evidence using an approach based on the GRADE system in medicine (Guyatt et al., 2008). This is an expert-judgment-based approach that is informed, but not driven, by quantitative metrics (such as number of studies included). The strength of evidence judgements took account of the number of original studies, the methodological quality of the meta-analysis (including limitations in the approach or corpus of studies considered), consistency (and homogeneity) of results, the directness of results, any imprecision, and any reporting bias.

Many of the original studies that are synthesised in the meta-analyses and systematic reviews were conducted in the US, which has a very different (mathematics) education system. In some cases, implementation would require some recontextualisation and translation for the UK classroom and school contexts and to align with the curriculum. To address this, we assessed the relevance, or the directness, of the evidence for UK classrooms using an approach also based on the GRADE system. Two members of the research team independently gave a high/medium/low rating for each section. Disagreements were resolved through discussion.

Results

As discussed in the main report, for 12 strategies, the evidence was judged sufficient to aggregate through a secondary meta-analysis of the original primary meta-analyses: computer-aided instruction (CAI), cooperative learning, explicit and direct instruction, feedback to pupils, feedback to teachers, heuristics, manipulatives, peer tutoring, representations and visualisation, self-instruction, pupil centred approaches and tutoring by adults. The results of this aggregation are presented in Table 8 together with our judgements about the strength of the evidence and its relevance to UK mathematics classrooms. These effects are also compared in Figure 14.

We consider the effect sizes should be regarded as indicative of the relative impact of different strategies rather than a precise estimate of the actual impact. Mindful of this lack of precision, in this review, we will treat effects of up to $d=0.25$ as small, of $0.25 < d < 0.75$ as moderate and of $d=0.75$ or greater as large. However, our interpretation will be cautious and highly dependent on the context.

The evidence relating to the remaining seven strategies is summarised in Table 9: attitudinal and behavioural interventions, cross-age tutoring, individualised instruction, instructional components, providing information to parents, technology tools and textbooks.

Table 8: Second-order meta-analysis results for the twelve strategies with a sufficient quantity and quality of evidence to aggregate across primary meta-analyses. Results ordered highest to lowest by strength of evidence, then size of effect. Results highlighted under heterogeneity indicate strategies for which no evidence of heterogeneity was found.

Strategy	Meta-analyses	Metas excluded from aggregation	Meta-analyses aggregated	Number of effects (k)	High quality meta-analyses	Number of medium quality meta-analyses	Number of low quality meta-analyses	Effect Size			Heterogeneity			Strength of Evidence	Relevance
								d	CI-	CI+	p	I ²	p(I ²)		
Explicit and direct instruction	13	4	9	283	7	6	0	0.59	0.53	0.65	<.0001	1.1	0.077	Very Strong	Strong
Computer-Aided Instruction (CAI)	11	2	9	207	7	1	1	0.18	0.04	0.31	0.1	91.7	<.0001	Strong	Moderate
Manipulatives	10	5	5	119	2	3	0	0.39	0.29	0.48	<.0001	60	0.019	Moderate	Strong
Tutoring by adults	4	1	3	77	2	0	1	0.36	0.21	0.51	<.0001	81.8	0.012	Moderate	Strong
Peer tutoring	7	2	5	94	2	2	1	0.66	0.52	0.79	<.0001	0	0.871	Moderate	Moderate
Heuristics	10	4	6	55	2	3	0	0.62	0.51	0.74	<.0001	3.1	0.103	Moderate	Moderate
Feedback to pupils	9	4	5	42	3	2	0	0.51	0.27	0.76	<.0001	74.6	0.006	Moderate/Weak	Strong
Representations	8	2	6	39	3	3	0	0.45	0.27	0.63	<.0001	69.2	0.012	Moderate/Weak	Strong
Feedback to teachers	4	0	4	19	2	2	0	0.39	0.13	0.66	0.003	41.3	0.226	Moderate/Weak	Strong
Self-instruction	5	1	4	32	2	2	0	1.02	0.63	1.42	<.0001	0	0.979	Moderate/Weak	Moderate
Cooperative Learning	7	1	6	68	2	4	0	0.29	0.15	0.43	<.0001	75.7	0.003	Moderate/Weak	Moderate
Student-centred approaches	6	1	5	47	0	5	0	0.73	0.46	0.99	<.0001	0	0.647	Weak	Weak

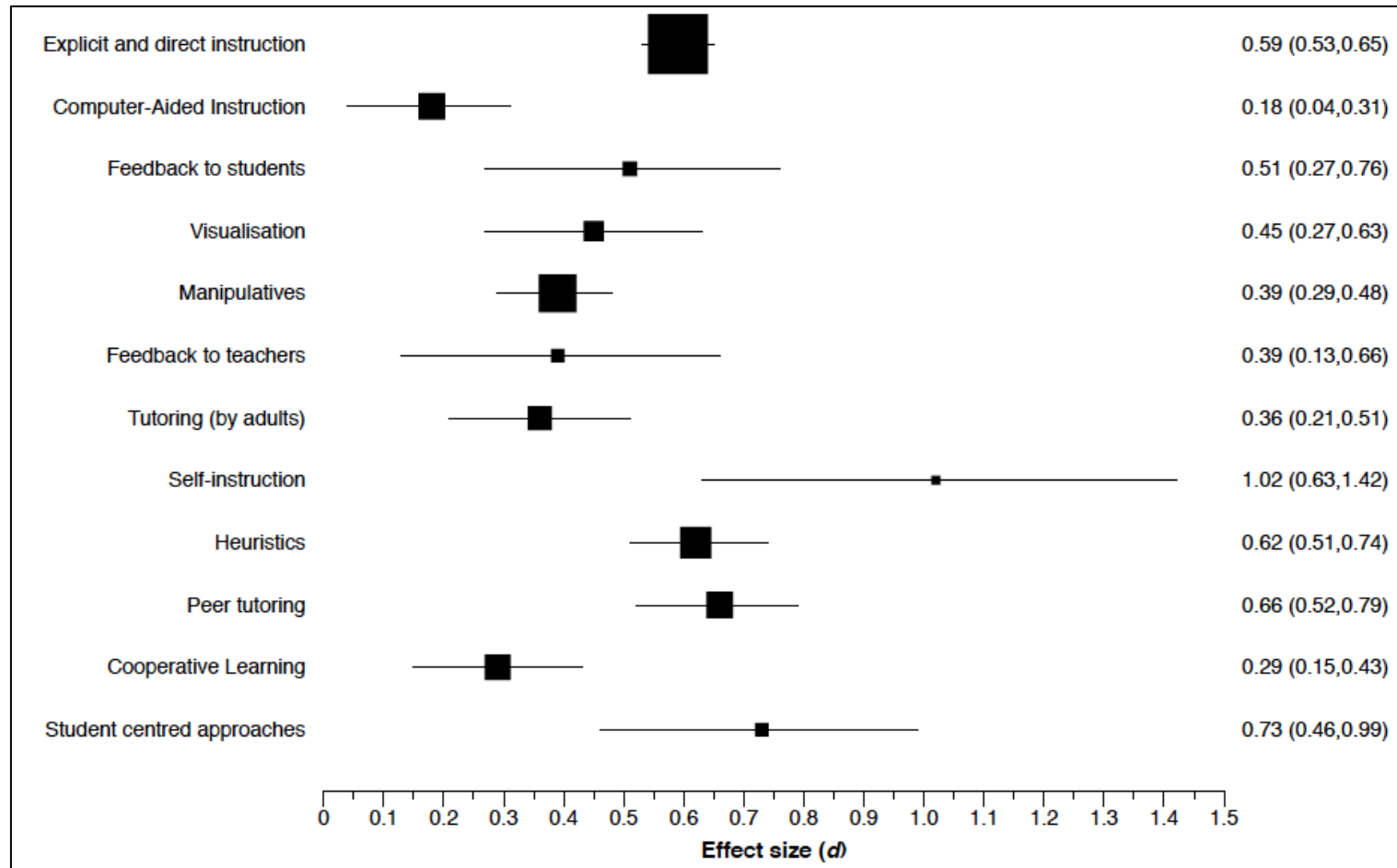


Figure 14: Comparing the effects of the twelve strategies and approaches with sufficient evidence to aggregate with point estimates and 95% confidence intervals. Strategies ordered as for Table 8 highest to lowest by strength of evidence, then relevance, then size of effect. Box size proportional to number of effects (k) aggregated.

Table 9: Evidence relating to seven strategies not aggregated

Strategy	Meta-analyses	Number of effects	Reason / Comment
Motivation, behavioural and attitudes	3	36	Wide variation of interventions, many of which are poorly defined. Results of meta-analyses inconsistent.
Textbooks	3	81	Studies conducted in the US where textbooks are conflated with curricular programmes.
Individualised instruction	2	357	Variation between results. Both metas are very dated and refer to individualised programmes that are no longer available. Only one has sufficient information for aggregation.
Instructional components	12	-	Components too varied for meaningful aggregation. Includes several meta-analyses of only single-case designs.
Technology tools	8	73	Different technology tools too disparate to aggregate. Calculators reported under number and calculation topic.
Providing information to parents	1	2	Too few original studies.
Cross-age tutoring	1	2	Too few original studies.

References

- Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities*, 35(3), 657-670. doi:http://dx.doi.org/10.1016/j.ridd.2013.12.010
- Cousin, G. (2006). An introduction to threshold concepts. In Meyer, J., & Land, R. (2006). *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge*. Routledge.
- Dowker, A. (2015). Individual differences in arithmetical abilities: the componential nature of arithmetic. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford Handbook of Numerical Cognition* (pp. 878-894). Oxford: Oxford University Press.
- Geary, D. C. (2011). Consequences, Characteristics, and Causes of Mathematical Learning Disabilities and Persistent Low Achievement in Mathematics. *Journal of developmental and behavioral pediatrics: JDBP*, 32(3), 250-263. doi:10.1097/DBP.0b013e318209edef
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924-926. doi:10.1136/bmj.39489.470347.AD
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis. *Child Development*, 0(0). doi:10.1111/cdev.13068
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., et al. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), 1013-1020. doi:10.1016/j.jclinepi.2008.10.009
- Siegler, R. S., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., . . . Wray, J. (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide (NCEE #2010-4039)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Appendix 1: IMAP Number Items, Topic and Origin

Question text		Topic	LNRP	CSMS / ICCAM S	New item	Justification
1	A shirt costs £20. Alex buys 3 shirts. How much does this cost?	IC			Y	Multiplicative operators: multiplicand / multiplier in opposite order to item 18.
2	What is 12 more than 26?	IC	Y5			Computation: two-digit addition.
3	Here are some marbles. Peter wins a quarter of these marbles. How many marbles does Peter win? [8 marbles]	F	Y5			Fraction as part-of-a-whole.
4	A stick of rock is shared equally among 5 children. What fraction of the stick does each child get?	F	Y5			Sharing a continuous quantity.
5	There are 12 eggs in a box. 5 of the eggs are cracked. What fraction of the eggs are cracked?	F	Y5			Fraction as part-of-a-whole.
6	Look at the number line. What number is the arrow pointing to? [6230]	NL	Y5 & Y6			Measures: Identifying four-digit number on a number line.
7	Look at the number line. What number is the arrow pointing to? [387]	NL	Adapted NL			Measures: Identifying three-digit number on a number line.
8	Look at the number line. Click and drag the arrow so that it points to the number six thousand and twenty-five.	NL	Y5 & Y6			Measures: Placing four-digit number on a number line.
9	Look at the number line. Click and drag the arrow so that it points to the number six thousand one hundred and eighty.	NL	Y5			Measures: Placing four-digit number on a number line.
10	Look at the number line. What number is the arrow pointing to? Use the decimal point on the keypad to enter your answer. [5.8: Decimals]	NL	Y5 & Y6			Measures: Identifying decimal number on a number line.
11	What is 1 less than 200?	PV	Y5 & Y6			Borrowing aspect of subtraction
12	What is 1 less than one thousand one hundred?	PV	Y5 & Y6			Borrowing aspect of subtraction
13	What is 10 more than three thousand five hundred and ninety-seven?	PV		Decimal s 5a		Carrying aspect of addition
14	4 children out of the 100 children on a school trip forgot to bring their lunch. What percentage is this?	P	Y5 & Y6	Ratio 8a		Percentage item from CSMS / ICCAMS ratio test
15	The newspaper says 24 out of 400 cars have faulty brakes. What percentage is this?	P	Y5 & Y6	Ratio 8c		Percentage item from CSMS / ICCAMS ratio test

Question text		Topic	LNRP	CSMS / ICCAM S	New item	Justification
16	6% of children travel to school by train. There are 200 children in the school. How many children travel to school by train?	P		Adapted Ratio 8b		Percentage item from CSMS / ICCAMS ratio test
17	Some jeans cost £20. In a sale the price is reduced by 10%. How much do the jeans cost now?	P		Adapted Ratio 8d		Percentage item from CSMS / ICCAMS ratio test
18	Erica buys 20 magazines. Each magazine costs £3. What is the total cost?	IC			Y	Multiplicative operators: multiplicand / multiplier in opposite order to item 1.
19	Look at this calculation. $34+28=62$. Find a quick way to work out the answer to $34+29$.	DF			Y	Derived fact item: Ans+1
20	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $57+86$	DF	Y5 & Y6			Derived fact item: Commutativity of addition; (this, and 21-23, originally from Van Den Heuvel-Panhuizen, 1996). ⁴
21	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $860+570$	DF	Y5 & Y6			Derived fact item: Effect of multiplication by a power of 10.
22	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $85+57$	DF	Y5			Derived fact item: Ans-1
23	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $86+86+57+57$	DF	Y5 & Y6			Derived fact item: Doubling.
24	Look at this calculation. $16.25891+23.74109=40$. Find a quick way to work out the answer to $17.25891+23.74109$ [Decimals]	DF	Y5 & Y6			Derived fact item: Decimals, Ans+1.
25	Look at this calculation. 15 multiplied by 24=360. Find a quick way to work out the answer to 16 multiplied by 24.	DF	Y5 & Y6			Derived fact item: Multiplication, Ans+1 lot (ie $16 \times 24 = 15 \times 24 + 24$)
26	Click and drag the three numbers below into the boxes so that they are in order from highest to lowest: 0.07, 0.23, 0.1 [Decimals]	PV	Y5 & Y6			Comparison of decimals.

⁴ Van Den Heuvel-Panhuizen, M. (1996). *Assessment and realistic mathematics education*. Utrecht: CD-β Press / Freudenthal Institute.

Question text		Topic	LNRP	CSMS / ICCAM S	New item	Justification
27	Enter a number that is larger than nought point six but smaller than nought point seven. {Decimals]	PV	Y5 & Y6			Infinite nature of the set of rational numbers.
28	A box contains 70 pencils. How many pencils are there in 10 boxes?	PV			Y	Effect of multiplication by a power of 10; grouping.
29	What are 10 lots of 324?	PV	Y5 & Y6			Effect of multiplication by a power of 10.
30	What are 10 lots of 5.3?	PV	Y5 & Y6			Effect of multiplication by a power of 10.
31	How many 10s are there in 200?	PV	Y5			Effect of division by a power of 10.
32	How many 10s are there in 4500?	PV	Y5 & Y6			Effect of division by a power of 10.
33	How many 10s are there in 489? Enter your answer as a decimal. [48.9; Decimals]	PV	Y5 & Y6			Effect of division by a power of 10; decimals as a result of a division operation.
34	Enter the missing number in the box. $\frac{1}{3} = \frac{2}{?}$	F	Y5 & Y6	Fraction s 9a		Equivalence of fractions.
35	Enter the missing number in the box. $\frac{4}{12} = \frac{1}{?}$	F	Y5 & Y6	Fraction s 9e		Equivalence of fractions.
36	John buys 8 cakes. They cost two pounds forty in total. Each cake costs the same amount. What calculation would you do to work out the cost of one cake?	SC	Y5 & Y6			Recognition of the correct operation to apply in a problem involving decimals: multiplication with grouping.
37	Pencils cost 18 pence each. What calculation would you do to work out how many you could buy for 90 pence?	SC	Y5 & Y6			Recognition of the correct operation to apply in a problem: division with allocation/rate.
38	Meg wears an outfit each day made up of a T-shirt and a skirt. She has 3 different T-shirts and 6 different skirts. Click on the calculation that would tell you how many different possible outfits ...	SC	Y5 & Y6			Recognition of the correct operation to apply in a problem involving decimals: Cartesian Product.
39	The Green family have to drive 261 miles to get from London to Leeds. After driving for 87 miles, they stop for lunch. How do you work out how far they still have to drive?	SC	Y5 & Y6			Recognition of the correct operation to apply in a problem involving decimals: subtraction.
40	A motorbike can travel 40 miles on each litre of petrol. Which calculation would tell you how many miles it would be able to travel on 8 litres?	SC		Adapted Decimal s		Recognition of the correct operation to apply in a problem involving decimals: multiplication with rate.

Question text		Topic	LNRP	CSMS / ICCAM S	New item	Justification
41	12 friends are making a total of 300 hats for a party. Each friend will make the same number of hats. How many hats will each friend make?	IC			Y	Computation: division with allocation/rate.
42	Click on the larger fraction. $\frac{1}{4}$, $\frac{3}{4}$	F	LNRP Example			Comparison of two fractions, same denominator (familiar).
43	Click on the larger fraction. $\frac{3}{7}$, $\frac{5}{7}$	F	Y5 & Y6			Comparison of two fractions, same denominator (less familiar).
44	Click on the larger fraction. $\frac{3}{5}$, $\frac{3}{4}$	F	Y5 & Y6			Comparison of two fractions, same numerator.
45	Click on the number that is nearest in size to nought point one eight. [0.18: Decimals]	E	Y4	Decimals 18c		Significant figures and approximation: decimals.
46	Click on the number that is nearest in size to 2.9 multiplied by 7. [Decimals]	E	Y5 & Y6	Decimals 18d		Significant figures and approximation: multiplication involving decimals.
47	A soup recipe for 3 people needs 8 carrots. How many carrots are needed for 9 people?	R			Y	Ratio scalar (integer) multiplier: $\times 3$.
48	A soup recipe for 3 people needs 6 potatoes. How many potatoes are needed for 2 people?	R	Y5 & Y6			Ratio scalar (rational) multiplier: $\times \frac{2}{3}$.
49	A soup recipe for 3 people needs 2 onions. How many onions are needed for 9 people?	R	Y5 & Y6			Ratio scalar (integer) multiplier: $\times 3$.
50	A soup recipe for 3 people needs 8 carrots and 2 onions. Lee uses 4 onions. How many carrots will he need?	R			Y	Ratio scalar (integer) multiplier: doubling with redundant distractor (3 people).
51	What is half of 16?	F			Y	Computation: halving.
52	What is one third of 12?	F			Y	Computation: simple division expressed as fraction (or fraction as multiplier).
53	What is two fifths of 20?	F	Y5 & Y6			Computation: multiplication with fraction multiplier.
54	What is the area of the shape below? (3 by 4 'grid-ed' rectangle)	A		Algebra 7a		Area model of multiplication; unexpectedly high non-correct responses on ICCAMS algebra test.
55	What is the area of the shape below? (6 by 10 'non-grid-ed' rectangle)	A		Algebra 7b		Area model of multiplication; unexpectedly high non-correct responses on ICCAMS algebra test.
56	These two ticks are exactly the same shape. Find the length of the red part.	R		Adapted Ratio 7		
57	These two Vs are exactly the same shape. Find the length of the red part.	R		Adapted Ratio 7		
58	Here is a 12 by 10 array. How many dots are there?	A			Y	Array / area model of multiplication; effect of multiplication by a power of 10.

Question text		Topic	CSMS / LNRP ICCAMS		New item	Justification
59	This is a 14 by 36 array. How many more dots would there be in a 15 by 36 array?	A			Y	Array / area model of multiplication: Similar to derived fact item 25, but with array representation: $15 \times 36 = 14 \times 36 + 36$.
60	This is part of a sheet of paper which had two arrays drawn on it. The arrays have the same number of dots. The size of the top array is 4 by 23. What was the size of the bottom array?	A			Y	Array / area model of multiplication: Associativity, or $4 \times 23 = 2 \times 46$.
61	The chairs in a room are arranged in rows. There are 14 chairs in each row. How many rows are needed for 84 people?	A			Y	Array / area model of multiplication: Division & rate.
		Totals	40	14	13	

Appendix 2: Arithmetic fact recall items (in order as presented)

A1	$2 + ? = 10$
A2	$4 \times 5 = ?$
A3	$25 + ? = 50$
A4	$? + 14 = 20$
A5	$18 \div 9 = ?$
A6	$30 \div 5 = ?$
A7	$9 \times 9 = ?$
A8	$99 \div 3 = ?$
A9	$5 \times 11 = ?$
A10	$50 = 18 + ?$
A11	$6 \times 8 = ?$
A12	$48 + ? = 90$
A13	$27 \div 3 = ?$
A14	$8 \times 3 = ?$
A15	$54 \div 6 = ?$
A16	$7 \times 8 = ?$
A17	$36 = 4 \times ?$
A18	$5 \times 25 = ?$
A19	$12 + ? = 60$
A20	$63 \div 7 = ?$
A21	$7 \times 11 = ?$
A22	$88 \div 22 = ?$
A23	$11 \times 12 = ?$
A24	$4 \times 13 = ?$
A25	$15 \times 5 = ?$
A26	$80 = 16 + ?$
A27	$96 \div 12 = ?$
A28	$70 = 21 + ?$
A29	$121 \div 11 = ?$
A30	$90 \div 15 = ?$

Appendix 3: Comparison of scores to LNRP and ICCAMS

In this appendix, we present the topic area scores and facilities (percentage correct) of the matched samples of Year 5 and Year 9 pupils.

Question		Topic	IMAP Matched Sample	
			Y5	Y9
Q02	What is 12 more than 26?	IC	0.85	0.82
Q03	Here are some marbles. Peter wins a quarter of these marbles. How many marbles does Peter win? [8 marbles]	F	0.62	0.60
Q04	A stick of rock is shared equally among 5 children. What fraction of the stick does each child get?	F	0.60	0.68
Q05	There are 12 eggs in a box. 5 of the eggs are cracked. What fraction of the eggs are cracked?	F	0.31	0.47
Q06	Look at the number line. What number is the arrow pointing to? [6230]	NL	0.29	0.29
Q08	Look at the number line. Click and drag the arrow so that it points to the number six thousand and twenty-five.	NL	0.33	0.27
Q09	Look at the number line. Click and drag the arrow so that it points to the number six thousand one hundred and eighty.	NL	0.59	0.57
Q10	Look at the number line. What number is the arrow pointing to? Use the decimal point on the keypad to enter your answer. [5.8: Decimals]	NL	0.42	0.47
Q11	What is 1 less than 200?	PV	0.92	0.90
Q12	What is 1 less than one thousand one hundred?	PV	0.70	0.57
Q14	4 children out of the 100 children on a school trip forgot to bring their lunch. What percentage is this?	P	0.43	0.48
Q15	The newspaper says 24 out of 400 cars have faulty brakes. What percentage is this?	P	0.05	0.09
Q20	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $57+86$	DF	0.73	0.70
Q21	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $860+570$	DF	0.67	0.65
Q22	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $85+57$	DF	0.68	0.63
Q23	Look at this calculation. $86+57=143$. Find a quick way to work out the answer to $86+86+57+57$	DF	0.48	0.44
Q26	Click and drag the three numbers below into the boxes so that they are in order from highest to lowest: 0.07, 0.23, 0.1 [Decimals]	PV	0.30	0.46
Q27	Enter a number that is larger than nought point six but smaller than nought point seven. {Decimals}	PV	0.43	0.54
Q28	A box contains 70 pencils. How many pencils are there in 10 boxes?	PV	0.72	0.76
Q29	What are 10 lots of 324?	PV	0.48	0.51
Q30	What are 10 lots of 5.3?	PV	0.28	0.38
Q31	How many 10s are there in 200?	PV	0.72	0.67
Q32	How many 10s are there in 4500?	PV	0.51	0.42
Q33	How many 10s are there in 489? Enter your answer as a decimal. [48.9: Decimals]	PV	0.31	0.30
Q34	Enter the missing number in the box. $1/3 = 2/?$	F	0.44	0.51

Question		Topic	IMAP Matched Sample	
			Y5	Y9
Q35	Enter the missing number in the box. $4/12 = 1/?$	F	0.27	0.34
Q36	John buys 8 cakes. They cost two pounds forty in total. Each cake costs the same amount. What calculation would you do to work out the cost of one cake?	SC	0.23	0.28
Q37	Pencils cost 18 pence each. What calculation would you do to work out how many you could buy for 90 pence?	SC	0.21	0.29
Q38	Meg wears an outfit each day made up of a T-shirt and a skirt. She has 3 different T-shirts and 6 different skirts. ... how many different possible outfits ...	SC	0.39	0.30
Q39	The Green family have to drive 261 miles to get from London to Leeds. After driving for 87 miles, they stop for lunch. How do you work out how far they still have to drive?	SC	0.45	0.39
Q43	Click on the larger fraction. $3/7$, $5/7$	F	0.79	0.72
Q44	Click on the larger fraction. $3/5$, $3/4$	F	0.45	0.43
Q46	Click on the number that is nearest in size to 2.9 multiplied by 7. [Decimals]	E	0.38	0.36
Q48	A soup recipe for 3 people needs 6 potatoes. How many potatoes are needed for 2 people?	R	0.40	0.34
Q49	A soup recipe for 3 people needs 2 onions. How many onions are needed for 9 people?	R	0.30	0.35
Q53	What is two fifths of 20?	F	0.19	0.17
		<i>Mean facility:</i>	0.47	0.48

Appendix 4: Factor loadings for PCA for each item on the main IMAP test

Shading indicates factor loadings between 0.3 and 0.5 (light blue); between 0.5 and 0.7 (dark blue); greater than 0.7 (black); and negative loading of magnitude greater than 0.3 (orange).

	Topic	Ratio	Area	Number lines	Derived facts	General number	Whole number bias
Q1	IC	0.07	-0.04	-0.07	0.07	0.45	-0.08
Q2	IC	0.07	0	0.08	-0.06	0.45	0
Q3	F	0.22	0.09	0.2	-0.06	0.34	-0.08
Q4	F	0.06	0.28	0.26	-0.01	0.12	0.01
Q5	F	0.05	0.3	0.19	0.01	-0.09	-0.07
Q6	NL	-0.03	0	0.69	-0.01	-0.01	0.07
Q7	NL	-0.15	0.04	0.49	0.09	0.04	0.04
Q8	NL	-0.04	-0.06	0.65	0.01	0.1	0.04
Q9	NL	-0.12	-0.02	0.54	0.12	0.17	0.02
Q10	NL	0.06	0.08	0.41	0.05	0.1	-0.05
Q11	PV	0.01	-0.04	0.11	-0.01	0.52	0.03
Q12	PV	0.09	-0.14	0.39	0	0.39	-0.06
Q13	PV	0.06	0	0.32	-0.02	0.36	0.03
Q14	P	0.21	0.04	0.23	0.16	0.03	-0.08
Q15	P	0.41	0.08	0.27	0	-0.26	0.03
Q16	P	0.26	0.04	0.31	0.05	-0.2	-0.04
Q17	P	0.43	0.06	0.28	-0.02	-0.19	-0.01
Q18	IC	0.08	0.22	0.14	0.01	0.33	-0.09
Q19	DF	-0.01	-0.03	-0.07	0.67	0.01	0.04
Q20	DF	-0.03	-0.01	-0.04	0.77	-0.05	0
Q21	DF	-0.06	0.06	0	0.69	0.04	-0.01
Q22	DF	-0.04	0.06	0	0.62	-0.05	-0.03
Q23	DF	0.13	0.03	0.13	0.47	0.05	0.02
Q24	DF	0.12	-0.11	0.07	0.59	0	-0.03
Q25	DF	0.33	-0.08	0.24	0.24	-0.07	0.06
Q26	PV	0.19	0.27	0.24	0.02	-0.07	-0.07
Q27	PV	0.14	0.31	0.27	0.03	0.08	-0.01
Q28	PV	-0.01	0.19	0.15	0.12	0.39	-0.13
Q29	PV	0.1	0.26	0.23	0.05	0.23	-0.08
Q30	PV	0.14	0.36	0.33	0	0.05	-0.05

	Topic	Ratio	Area	Number lines	Derived facts	General number	Whole number bias
Q31	PV	0.17	-0.05	0.13	0.05	0.39	0.06
Q32	PV	0.14	0.07	0.16	0.07	0.32	0.04
Q33	PV	0.17	0.17	0.25	0.06	0.14	0.07
Q34	F	0.12	0.43	0.12	0.06	0.04	0.03
Q35	F	0.24	0.38	0.14	-0.02	-0.02	0.02
Q36	SC	0.1	0.31	0.11	0.09	0.06	0.07
Q37	SC	0.26	0.3	0.21	-0.01	-0.04	0.03
Q38	SC	0.07	0.09	0.03	0.11	0.01	-0.02
Q39	SC	0.21	0.13	0.06	0.11	0.17	0.12
Q40	SC	0.1	0.06	0.01	0.17	0.26	0.08
Q41	IC	0.19	0.19	0.15	0.1	-0.05	0.07
Q42	F	0.01	0.03	0.02	0.01	0.01	0.8
Q43	F	0.01	-0.03	0.02	0	0	0.81
Q44	F	0.21	0.16	0.16	0.11	0.06	-0.39
Q45	E	0.32	0.26	0.14	-0.08	-0.01	0.03
Q46	E	0.28	0.13	0.07	0.03	0.08	0.09
Q47	R	0.66	-0.07	-0.13	0.03	0.12	-0.03
Q48	R	0.46	-0.04	-0.06	0.07	0.19	0.03
Q49	R	0.59	-0.08	-0.07	0.06	0.12	-0.05
Q50	R	0.41	0.11	-0.09	0.09	0.21	0.07
Q51	F	0.01	0.26	-0.04	0.16	0.47	0.06
Q52	F	0.12	0.37	0.16	0.09	0.03	0.11
Q53	F	0.24	0.34	0.11	0.09	-0.13	0.08
Q54	A	-0.07	0.79	-0.1	0.05	0.02	0.01
Q55	A	-0.04	0.78	-0.04	-0.01	0.01	-0.04
Q56	R	0.5	0.28	-0.08	-0.09	-0.03	0.09
Q57	R	0.36	0.07	-0.16	0.01	0.18	0.09
Q58	A	0	0.3	0.05	0.05	0.31	0.14
Q59	A	0.12	-0.06	0.11	-0.02	-0.12	0.26
Q60	A	0.37	0.09	0.13	0.1	-0.08	0.02
Q61	A	0.37	0	0.18	0.05	-0.06	0.07

Appendix 5: The full dataset for the literature review and second-order meta-analysis

Level 1: Meta-analyses judged to be core with mathematics education as central to the review, relevant to Key Stage 3 and a specific focus on low attaining pupils

Note: This category includes some meta-analyses focusing exclusively on aggregating single-case designs.

- Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, 103(1), 51-73. doi:10.1086/499715 <http://www.jstor.org/stable/1002308>
- Browder, D. M., Spooner, F., Ahlgrim-Delzell, L., Harris, A. A., & Wakemanxya, S. (2008). A meta-analysis on teaching mathematics to students with significant cognitive disabilities. *Exceptional Children*, 74(4), 407-432. <http://dx.doi.org/10.1177/001440290807400401>
- Burns, M. K., Coddling, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, 39(1), 69.
- Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, 105(2), 380. doi: 10.1037/a0031084
- Chen, H. (2004). The efficacy of mathematics interventions for students with learning disabilities: A meta-analysis. (Order No. 3157959, The University of Iowa). ProQuest Dissertations and Theses, 136 pages. Retrieved from <http://search.proquest.com/docview/305196240?accountid=14533>. (305196240).
- Chodura, S., Kuhn, J.-T., & Holling, H. (2015). Interventions for Children With Mathematical Difficulties. *Zeitschrift für Psychologie*, 223(2), 129-144. doi:10.1027/2151-2604/a000211
- Coddling, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice*, 26(1), 36-47.
- Dennis, M. S., Sharp, E., Chovanes, J., Thomas, A., Burns, R. M., Custer, B., & Park, J. (2016). A Meta-Analysis of Empirical Research on Teaching Students with Mathematics Learning Difficulties. *Learning Disabilities Research & Practice*, 31(3), 156-168.
- Domino, J. (2010). The Effects of Physical Manipulatives on Achievement in Mathematics in Grades K-6: A Meta-Analysis. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.
- Donker, A. S., De Boer, H., Kostons, D., van Ewijk, C. D., & Van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1-26. DOI: 10.1016/j.edurev.2013.11.002
- Fyfe, E. R., & Brown, S. A. (2018). Feedback influences children's reasoning about math equivalence: A meta-analytic review. *Thinking & Reasoning*, 24(2), 157-178. doi:10.1080/13546783.2017.1359208
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of

- instructional components. *Review of Educational Research*, 79(3), 1202-1242. <http://dx.doi.org/10.3102/0034654309334431>
- Guskey, T. R., & Pigott, T. D. (1988). Research on Group-Based Mastery Learning Programs: A Meta-Analysis. *The Journal of Educational Research*, 81(4), 197-216. doi:10.1080/00220671.1988.10885824
- Haas, M. (2005). Teaching methods for secondary algebra: A meta-analysis of findings. *Nassp Bulletin*, 89(642), 24-46. <http://dx.doi.org/10.1177/019263650508964204>
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, Vol. 21, No. 1, pp. 33-46 [Meta-analysis of 151 studies]
- Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education*, 242-273. <http://www.jstor.org/stable/749120>
- Hembree, R., & Dessart, D. J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for research in mathematics education*, 17(2), 83-99.
- Hughes, E. M., Witzel, B. S., Riccomini, P. J., Fries, K. M., & Kanyongo, G. Y. (2014). A Meta-Analysis of Algebra Interventions for Learners with Disabilities and Struggling Learners. *Journal of the International Association of Special Education*, 15(1).
- Jacobse, A. E., & Harskamp, E. (2011). A meta-Analysis of the Effects of instructional interventions on students' mathematics achievement.
- Jitendra, A. K., Lein, A. E., Im, S.-h., Alghamdi, A. A., Hefte, S. B., & Mouanoutoua, J. (2017). Mathematical Interventions for Secondary Students With Learning Disabilities and Mathematics Difficulties: A Meta-Analysis. *Exceptional Children*, 84(2), 177-196. doi:10.1177/0014402917737467
- Jitendra, A. K., Nelson, G., Pulles, S. M., Kiss, A. J., & Houseworth, J. (2016). Is Mathematical Representation of Problems an Evidence-Based Strategy for Students With Mathematics Difficulties?. *Exceptional Children*, 83(1), 8-25.
- Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs a meta-analysis. *Remedial and special education*, 24(2), 97-114.
- Kuchler, J.M. (1998) The effectiveness of using computers to teach secondary school (grades 6-12) mathematics: A meta-analysis. Ph.D. thesis, University of Massachusetts Lowell.
- Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of Mastery Learning Programs: A Meta-Analysis. *Review of Educational Research*, 60(2), 265-299. doi:doi:10.3102/00346543060002265
- Lee, D. S. (2000). A meta-analysis of mathematics interventions reported for 1971-1998 on the mathematics achievement of students identified with learning disabilities and students identified as low achieving. Doctoral Thesis, University of Oregon ProQuest UMI 9963449
- Melby-Lervåg, M. m., & Hulme, C. c. (2013). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*, 49(2), 270-291.
- Methe, S. A., Kilgus, S. P., Neiman, C., & Riley-Tillman, T. C. (2012). Meta-analysis of interventions for basic mathematics computation in single-case research. *Journal of Behavioral Education*, 21(3), 230-253.
- Pellegrini, M., Lake, C., Inns, A., & Slavin, R. E. (2018 - DRAFT). Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis.
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. *Educational Psychologist*, 50(2), 138-166. doi: 10.1080/00461520.2015.1036274

- Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education*, 498-505. <http://www.jstor.org/stable/749423>
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970-987. doi:10.1037/a0032447
- Stevens, E. A., Rodgers, M. A., & Powell, S. R. (2017). Mathematics Interventions for Upper Elementary and Secondary Students: A Meta-Analysis of Research. *Remedial and Special Education*, 0741932517731887. doi:10.1177/0741932517731887
- Stockard, J., Wood, T. W., Coughlin, C., & Rasplika Khoury, C. (2018). The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. *Review of Educational Research*, 88(4), 479-507. doi:10.3102/0034654317751919
- Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education*, 32(4), 207-225.
- Zhang, D., & Xin, Y. P. (2012). A follow-up meta-analysis for word-problem-solving interventions for students with mathematics difficulties. *The Journal of educational research*, 105(5), 303-318. <http://dx.doi.org/10.1080/00220671.2011.627397>

Level 2: Meta-Analyses judged to be relevant to the teaching of mathematics to Key Stage 3 low attaining pupils but does not fully meet the level 1 criteria

- Athappilly, K., Smidchens, U., & Kofel, J. W. (1983). A computer-based meta-analysis of the effects of modern mathematics in comparison with traditional mathematics. *Educational Evaluation and Policy Analysis*, 485-493.
- Capar, G., & Tarim, K. (2015). Efficacy of the cooperative learning method on mathematics achievement and attitude: A meta-analysis research. *Educational Sciences: Theory & Practice*, 2, 553-559. http://www.estp.com.tr/wp-content/uploads/2015/04/ESTPApril2015_553_559.pdf
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88-113.
- Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research*, 87(2), 243-282. doi:10.3102/0034654316687036
- Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education*, 433-463.
- Hartley, S.S. (1977) Meta-Analysis of the Effects of Individually Paced Instruction In Mathematics. Doctoral dissertation University of Colorado at Boulder.
- Holmes, A. B. (2013). Effects of Manipulative Use on PK-12 Mathematics Achievement: A Meta-Analysis. Society for Research on Educational Effectiveness.
- Horak, V. M. (1981). A meta-analysis of research findings on individualized instruction in mathematics. *The Journal of Educational Research*, 74(4), 249-253. <http://dx.doi.org/10.1080/00220671.1981.10885318>
- LeNoir, P. (1989). The effects of manipulatives in mathematics instruction in grades K-college: A meta-analysis of thirty years of research. . Doctoral Thesis, North Carolina State University at Raleigh, ProQuest UMI 8918109.

- Nunnery, J.A., Chappell, S. & Arnold, P. (2013). A meta-analysis of a cooperative learning model's effects on student achievement in mathematics. *Cypriot Journal of Educational Sciences*, 8(1), 34-48.
- Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of Instructional Improvement in Algebra A Systematic Review and Meta-Analysis. *Review of Educational Research*, 80(3), 372-400. <http://dx.doi.org/10.3102/0034654310374880>
- Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*. 79(2), 839-911. <http://dx.doi.org/10.3102/0034654308330968>

Level 3: Systematic reviews focused on mathematics teaching for low attaining pupils and meta analyses judged to be relevant to low attainers but either with insufficient information to extract data or a generic (non- mathematics) focus

- Anthony, G., & Walshaw, M. (2007). *Effective pedagogy in mathematics/Pāngarau: Best evidence synthesis iteration*. Wellington, New Zealand: Ministry of Education.
- Askew, M., Hodgen, J., Hossain, S., & Bretscher, N. (2010). *Values and variables: Mathematics education in high-performing countries*. London: Nuffield Foundation.
- Athappilly, K. K. (1978). *A meta-analysis of the effects of modern mathematics in comparison with traditional mathematics in the American educational system*. Western Michigan University, ED.D.,
- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of learning disabilities*, 38(4), 333-339.
- Cheung, A. C., & Slavin, R. E. (2011). *The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis*. (Baltimore: The Best Evidence Encyclopedia, Center for Data-Driven Reform in Education, Johns Hopkins University, 2011).
- Cowan, R., Saxton, M., & Tolmie, A. (Eds.). (2010). *Understanding diversity in number development*. BJEP Monograph Series II: Part 7 Understanding Number Development and Difficulties. London: British Psychological Society.
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231-264.
- Dowker, A. (2004). *What works for children with mathematical difficulties?* DfES Research Report RR554. London: Department for Education and Skills (DfES).
- Dowker, A. (2009). *What works for children with mathematical difficulties?* Nottingham: DCSF Publications.
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics Anxiety: What Have We Learned in 60 Years? *Frontiers in Psychology*, 7(508). doi:10.3389/fpsyg.2016.00508
- Dugan, J.J. (2007). *A systematic review of interventions in secondary mathematics with at-risk students: Mapping the Literature* Doctoral dissertation Colorado State University, Fort Collins Colorado. ProQuest UMI 3266389
- Durkin, K. (2011). *The self-explanation effect when learning mathematics: A meta-analysis*. Presented at the Society for Research on Educational Effectiveness 2011, Available online [accessed 13th March 2017]: <http://eric.ed.gov/?id=ED518041>.
- Ellis, A. K., Denton, D. W., & Bond, J. B. (2014). An analysis of research on metacognitive teaching strategies. *Procedia-Social and Behavioral Sciences*, 116, 4015-4024.

- Friso-van den Bos, I. i., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, 1029-44.
- Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide (NCEE 2014-4005)*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities*, 45(3), 257-269.
- Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, 26(1), 9-25.
- Gascoine, L., Higgins, S., & Wall, K. (2017). The assessment of metacognition in children aged 4–16 years: a systematic review. *Review of Education*, 5(1), 3-57.
- Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting Students Struggling with Mathematics: Response to Intervention (RtI) for Elementary and Middle Schools. NCEE 2009-4060. IES Practice Guide: What Works Clearinghouse.*
- Gersten, R., Woodward, J., & Darch, C. (1986). Direct instruction: A research-based approach to curriculum design and teaching. *Exceptional Children*, 53(1), 17-31.
- Gilmore, C. K., & Papadatou-Pastou, M. (2009). Patterns of individual differences in conceptual understanding and arithmetical skill: A meta-analysis. *Mathematical Thinking and Learning*, 11(1-2), 25-40.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Higgins, S., Hall, E., Baumfield, V., & Moseley, D. (2005). *A meta-analysis of the impact of the implementation of thinking skills approaches on pupils*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Hord, C., & Xin, Y. P. (2013). Intervention Research for Helping Elementary School Students with Math Learning Difficulties Understand and Solve Word Problems: 1996-2010. *Learning Disabilities: A Multidisciplinary Journal*, 19(1), 3-17. [Narrative review of multiple studies]
- Jacob, R., & Parkinson, J. (2015). The Potential for School-Based Interventions That Target Executive Function to Improve Academic Achievement. *Review of Educational Research*, 85(4), 512-552. doi:doi:10.3102/0034654314561338
- Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P. G., & Egan, A. M. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of learning disabilities*, 48(1), 51-72.
- Kidron, Y., & Lindsay, J. (2014). The Effects of Increased Learning Time on Student Academic and Nonacademic Outcomes: Findings from a Meta-Analytic Review. REL 2014-015. Regional Educational Laboratory Appalachia.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding It Up: Helping Children Learn Mathematics* (Prepared by the Mathematics Learning Study Committee, National Research Council). Washington DC: The National Academies Press.
- Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28-37. doi:10.1111/j.1745-3992.2011.00220.x

- Kluger, A. N., & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2), 254-284.
- Kyriacou, C., & Issitt, J. (2008). What Characterises Effective Teacher-initiated Teacher-pupil Dialogue to Promote Conceptual Understanding in Mathematics Lessons in England in Key Stages 2 and 3: A Systematic Review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Lai, E. R. (2011). Metacognition: A literature review. Always learning: Pearson research report.
- Lazowski, R. A., & Hulleman, C. S. (2016). Motivation Interventions in Education. *Review of Educational Research*, 86(2), 602-640. doi:10.3102/0034654315617832
- Li, Q. and Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, 22(3), 215-243.
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30(5), 520-540.
- Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: a meta-analysis. *Journal for Research In Mathematics Education*, 28(1), 26-47.
- McMullen, F., & Madelaine, A. (2014). Why is there so much resistance to Direct Instruction? *Australian Journal of Learning Difficulties*, 19(2), 137-151.
- Misquitta, R. (2011). A review of the literature: Fraction instruction for struggling learners in mathematics. *Learning Disabilities Research & Practice*, 26(2), 109-119. <http://dx.doi.org/10.1111/j.1540-5826.2011.00330.x>
- Muenks, K., & Miele, D. B. (Forthcoming, online first). Students' Thinking About Effort and Ability. *Review of Educational Research*, 0034654316689328. doi: 10.3102/0034654316689328
- Nunes, T., Bryant, P., & Watson, A. (2009). Key understandings in mathematics learning. London: Nuffield Foundation.
- Othman, N. (1996). The effects of cooperative learning and traditional mathematics instruction in grade K-12: A meta-analysis of findings. Doctoral Thesis, West Virginia University. ProQuest UMI 9716375
- Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*.
- Preston, J. A. (2007). Student-centered versus teacher-centered mathematics instruction: A meta-analysis. Doctoral Thesis, Indiana University of Pennsylvania. ProQuest UMI 3289778.
- Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, 1-13.
- Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., & Bakker, A. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review*, 19, 158-172.
- Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). Review and meta-analyses of school and teaching effectiveness. Enschede: Department of Educational Organisation and Management, University of Twente.
- Shin, M., & Bryant, D. P. (2015). Fraction Interventions for Students Struggling to Learn Mathematics: A Research Synthesis. *Remedial And Special Education*, 36(6-), 374-387.

- Siegler, R. S., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., . . . Wray, J. (2010). Developing effective fractions instruction for kindergarten through 8th grade: A practice guide (NCEE #2010-4039). . Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Slavin, R. E. (1987). Mastery Learning Reconsidered. *Review of Educational Research*, 57(2), 175-213. doi:doi:10.3102/00346543057002175
- Slavin, R. E. and Lake, C. (2007). *Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis* (Baltimore: The Best Evidence Encyclopedia, Center for Data-Driven Reform in Education, Johns Hopkins University, 2007).
- Slavin, R. E., Lake, C., & Groff, C. (2007). *Effective Programs in Middle and High School Mathematics: A Best-Evidence Synthesis* (Baltimore: The Best Evidence Encyclopedia, Center for Data-Driven Reform in Education, Johns Hopkins University, 2007).
- Sokolowski, A., Li, Y., & Willson, V. (2015). The effects of using exploratory computerized environments in grades 1 to 8 mathematics: a meta-analysis of research. *International Journal of STEM Education*, 2(1), 1-17.
- Star, J. R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M. R., . . . Zbiek, R. M. (2015). Teaching strategies for improving algebra knowledge in middle and high school students (NCEE 2014-4333). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.
- Stoner, D. A. (2004). The effects of cooperative learning strategies on mathematics achievement among middlegrades students: a meta-analysis. University of Georgia, Athens.
- Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, 76(2), 249-274. <http://dx.doi.org/10.3102/00346543076002249>
- Templeton, T. N., Neel, R. S., & Blood, E. (2008). Meta-analysis of math interventions for students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 16(4), 226-239. <http://dx.doi.org/10.1177/1063426608321691>
- Verschaffel, L., Lehtinen, E., & Van Dooren, W. (2016). Neuroscientific studies of mathematical thinking and learning: a critical look from a mathematics education viewpoint. *ZDM*, 48(3), 385-391.
- Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, 78(3), 516-551.
- Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, 22(4), 393-409. [May be relevant here if explores worked problem-solving examples and transfer]
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M. L., Herzig, P., Jitendra, A., . . . Ogbuehi, P. (2012). Improving mathematical problem solving in grades 4 through 8: A practice guide (NCEE 2012-4055). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Additional references cited in the RQ4 review

- Ashman, G. (2018). *The Truth about Teaching: An evidence-informed guide for new teachers*. SAGE.

- Askew, M., Brown, M., Rhodes, V., Johnson, D. C., & Wiliam, D. (1997). *Effective teachers of numeracy*. London: King's College.
- Baines, E., Blatchford, P., Kutnick, P., Chowne, A., Ota, C., & Berdondini, L. (2014). *Promoting Effective Group Work in the Primary Classroom: A handbook for teachers and practitioners*. London: Routledge.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, 47(1), 133-180. doi:10.3102/0002831209345157
- Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47-63). New York: Holt, Rinehart & Winston.
- Borko, H., & Wildman, T. (1986). *Recent research on instruction. Beginning teacher assistance program*. Richmond, VA: Department of Education, Commonwealth of Virginia.
- Brown, A. L., & Campione, J. C. (1990). Interactive learning environments and the teaching of science and mathematics. In M. Gardner et al., (Eds.), *Toward a scientific practice of science education*. Hillsdale, NJ: Erlbaum.
- Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. London: The Sutton Trust.
- Cohen Kadosh, R., Dowker, A., Heine, A., Kaufmann, L., & Kucian, K. (2013). Interventions for improving numerical abilities: Present and future. *Trends in Neuroscience and Education*, 2(2), 85-93. doi:http://dx.doi.org/10.1016/j.tine.2013.04.001
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 336(7650), 924-926. doi: 10.1136/bmj.39489.470347.AD
- Hattie, J. (2011). *Visible Learning for teachers*. Abingdon, UK: Routledge.
- Hiebert, J., & Grouws, D. A. (2007). The effect of classroom mathematics teaching on student learning. In F. K. J. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Greenwich, CT: Information Age Publishing.
- Higgins, S., & Katsipatakis, M. (2016). Communicating comparative findings from meta-analysis in educational research: some examples and suggestions. *International Journal of Research & Method in Education*, 39(3), 237-254. doi:10.1080/1743727X.2016.1166486
- Higgins, S., & Katsipatakis, M. (2016). Communicating comparative findings from meta-analysis in educational research: some examples and suggestions. *International Journal of Research & Method in Education*, 39(3), 237-254. doi:10.1080/1743727X.2016.1166486
- Hoyles, C. (2018). Transforming the mathematical practices of learners and teachers through digital technology. *Research in Mathematics Education*, 20(3), 209-228. doi:10.1080/14794802.2018.1484799
- McMullen, F., & Madelaine, A. (2014). Why is there so much resistance to Direct Instruction? *Australian Journal of Learning Difficulties*, 19(2), 137-151.
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of educational Psychology*, 81(2), 247.
- NCETM (2016). *The Essence of Maths Teaching for Mastery*. Retrieved from <https://www.ncetm.org.uk/files/37086535/The+Essence+of+Maths+Teaching+for+Mastery+june+2016.pdf>

- Pampaka, M., Williams, J., Hutcheson, G., Wake, G., Black, L., Davis, P., & Hernandez-Martinez, P. (2012). The association between mathematics pedagogy and learners' dispositions for university study. *British Educational Research Journal*, 38(3), 473-496. doi:10.1080/01411926.2011.555518
- Rosenshine, B. (2008). Five meanings of direct instruction. *Center on Innovation & Improvement, Lincoln*.
- Rosenshine, B. (2012). Principles of instruction: research-based strategies that all teachers should know. *American Educator*, 36(1), 12.
- Ruthven, K. (2001). British research on developing numeracy with technology. In M. Askew & M. Brown (Eds.), *Teaching and Learning Primary Numeracy: Policy, Practice and Effectiveness. A review of British research for the British Educational Research Association in conjunction with the British Society for Research into Learning of Mathematics* (pp. 29-33). Southwell, Notts: British Educational Research Association (BERA).
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., . . . Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), 1013-1020. doi:10.1016/j.jclinepi.2008.10.009
- Simpson, A. (2017). The misdirection of public policy: comparing and combining standardised effect sizes. *Journal of Education Policy*, 1-17. doi:10.1080/02680939.2017.1280183
- Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research*, 50, 315-342.
- Slavin, R. E. (1984). Team-Assisted individualization; Cooperative learning and individualized instruction in the mainstreamed classroom. *Remedial and Special Education*, 5, 33-42.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57(2), 175-213. doi:10.3102/00346543057002175
- Slavin, R. E. (1999). Comprehensive approaches to cooperative learning. *Theory into Practice*, 38(2), 74-79.
- Slavin, R. E., Sheard, M., Hanley, P., Elliott, L., & Chambers, B. (2013). Effects of Cooperative Learning and Embedded Multimedia on Mathematics Learning in Key Stage 2: Final Report. York: Institute for Effective Education.
- Slavin, R. E., Sheard, M., Hanley, P., Elliott, L., & Chambers, B. (2013). *Effects of Cooperative Learning and Embedded Multimedia on Mathematics Learning in Key Stage 2: Final Report*. York: Institute for Effective Education.
- Swan, M. (2005). *Improving learning in mathematics*. London: Department for Education and Skills.
- Swan, M. (2006). *Collaborative Learning in Mathematics: A challenge to our beliefs and practices*. Leicester: NIACE/NRDC.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Examples of excluded meta-analyses

Examples of meta-analyses excluded from Levels 1 and 2 of the dataset are set out below:

Reasons for exclusion from Levels 1 & 2	Example / explanation
Not a meta-analysis	Systematic or narrative reviews with no overall, or very minimally reported, meta-analysis, e.g. Timperley et al., (2007) excluded; Slavin (1987) included as Level 3 because relevant to explicit and direct instruction.
Not located	E.g., Parham (1983), a US doctoral thesis.
Not relevant to interventions or strategies targeted at mathematics	E.g., Peng et al., (2016), meta-analysis related to mathematics learning and working memory, excluded; Fischer et al., (2013) a framework for assessing interventions, excluded.
Not specific to LAs	E.g., Ellington (2006) Meta-analysis of the effects of graphing calculators, studies largely conducted with older or college students, excluded.
Superseded by another study	Either alternative version, e.g., Slavin et al., (2007), included as Level 4 because it provided some additional detail, or complete overlap with another study, e.g., Smith (1996), all studies included in Ellington's (2003) calculator use meta-analysis.
Not relevant to the UK context	Studies entirely, or very largely, consisting of studies in very different educational systems, e.g., Faramarzi et al., (2015), excluded.
Not specific to mathematics education	E.g., Chauhan (2017), generic use of technology tools, excluded.
Subjects too old	E.g., Lake et al., (2016), pedagogies with college students, excluded
Subjects too young	E.g., Malofeev (2005), pre-school.
Insufficient data to extract ES relevant to LAs	E.g., Li & Ma (2008), unclear how attainment scaled, excluded; Templeton et al. (2008), only PND reported, included as Level 3.

References for examples of excluded studies

- Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education*, 105, 14-30.
- Ellington, A. J. (2006). The Effects of Non-CAS Graphing Calculators on Student Achievement and Attitude Levels in Mathematics: A Meta-Analysis. *School Science and Mathematics*, 106(1), 16-26.
- Faramarzi, S., Shamsi, A., Samadi, M., & Ahmadzade, M. (2015). Meta-analysis of the efficacy of psychological and educational interventions to improve academic performance of students with learning disabilities in Iran. *Journal of education and health promotion*, 4.
- Fischer, U., Moeller, K., Cress, U., Nuerk, C. H. (2013) Interventions supporting children's mathematics school success: A meta-analytic review. *European Psychologist*, Vol 18(2), 2013, 89-113. <http://dx.doi.org/10.1027/1016-9040/a000141>
- Lake, W., Wallin, M., Woolcott, G., Boyd, W., Foster, A., Markopoulos, C. & Boyd, W. (2016): Applying an alternative mathematics pedagogy for students with weak

- mathematics: meta-analysis of alternative pedagogies, *International Journal of Mathematical Education in Science and Technology*, DOI: 10.1080/0020739X.2016.1245876
- Li, Q., & Ma, X. (2008). Technology in Mathematics Classrooms: A Meta-Analysis of the Recent Literature. *Journal on School Educational Technology*, 3(4), 34-54.
- Malofeeva, E. V. (2005). Meta-analysis of mathematics instruction with young children. PhD Dissertation, University of Notre Dame
- Parham, J. L. (1983). A meta-analysis of the use of manipulative material and student achievement in elementary school mathematics. Unpublished doctoral dissertation, Auburn University. (AAG8312477)
- Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*.
- Slavin, R. E., Lake, C., & Groff, C. (2007). *Effective Programs in Middle and High School Mathematics: A Best-Evidence Synthesis*. Baltimore: The Best Evidence Encyclopedia, Center for Data-Driven Reform in Education, Johns Hopkins University, 2007.
- Smith, B.A. (1996). A meta-analysis of outcomes from the use of calculators in mathematics education. (Doctoral dissertation, Texas A & M University-Commerce, 1996). Dissertation Abstracts International, 58(03), 787.all studies included in Ellington (2003)]
- Templeton, T. N., Neel, R. S., & Blood, E. (2008). Meta-analysis of math interventions for students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 16(4), 226-239. <http://dx.doi.org/10.1177/1063426608321691>
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development. Best evidence synthesis iteration (BES)*. Wellington, New Zealand: Ministry of Education)

UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL

www.ucl.ac.uk/ioe