University of Southampton

Faculty of Environmental and Life Sciences

School of Psychology

Exploring Factors Associated with Maths Performance in Children and Adolescents Aged 11-18 Years

by

Gemma Chelsea Muncer

Thesis for the degree of Doctorate in Educational Psychology

June 2020

University of Southampton

Abstract

Faculty of Environmental and Life
Sciences School of Psychology
Thesis for the degree of Doctorate in Educational Psychology

Exploring Factors Associated with Maths Performance in Children and Adolescents Aged 11-18 Years Old

by

Gemma Chelsea Muncer

Poor maths and numeracy skills are associated with a range of adverse outcomes, including reduced employability and poorer physical and mental health. In the UK, a grade 4 pass in GCSE Maths is a prerequisite to access many training courses and job opportunities. It is currently mandatory for UK students aged 16-18 who have not achieved a pass in GCSE Maths to re-sit the qualification (or equivalent). However, in 2019, over 80% of these students did not attain a pass grade by the age of 18.

The first chapter of this thesis presents a systematic literature review and meta-analysis that investigated the association between metacognition and maths performance in 11-16-year-olds. A systematic search in electronic databases, and of the grey literature, revealed 31 relevant studies. The quantitative synthesis of 82 effect sizes from 29 of these studies indicated a significantly positive correlation (r = .37, 95% CI = [.26, .47]) between metacognition and maths performance. Additional subgroup analyses revealed that the strength of association was stronger when (1) studies measured metacognition using think-aloud protocols and/or behavioural observation (online measures) rather than a self-reported offline questionnaire, judgment of learning score, confidence judgment or calibration score, and (2) the measure of maths performance required complex (versus simple) mathematical skill. However, there was very high unexplained heterogeneity between studies. These findings, alongside the existing literature, indicate the importance of metacognition for maths achievement in adolescents.

The empirical paper explored the outcomes of 12 online successive relearning sessions (versus restudying or teaching as usual) for students aged 16-18 years who were re-sitting GCSE Maths. The impact of this intervention was investigated in relation to (1) maths test performance, (2) state and trait measures of cognitive and affective indices associated with learning, and (3) student qualitative feedback about the sessions. Engagement in intervention sessions was low; 43% of students in the relearn or restudy groups, who completed time 1 measures, completed at least one (/12) session (relearn = 22, restudy = 29). Of these, 36 students also completed state measures for at least two sessions (/3) in one week (relearn = 17, restudy = 19). Including students in the relearn and restudy groups only where they had completed at least one relearn/restudy session, 68 students completed the maths post-test (relearn = 19, restudy = 26, control = 23) and 48 students completed trait questionnaire measures both pre- and post-intervention (relearn = 12, restudy = 24, control = 12). Students who predicted, at T1, that they would achieve a higher score in the T2 maths test completed more relearn sessions, and students who reported more positive attitudes towards maths and higher maths anxiety completed more relearn sessions. Furthermore, students who reported higher academic efficacy were more likely to engage in at least one relearn or restudy session (versus no sessions). Analysis of the impact of intervention revealed that students in the relearn and restudy groups scored higher on the maths test relative to the control group. Furthermore, there was a positive correlation between the number of relearn or restudy sessions completed and test score. Considering only students who completed at least one revision session, the correlation between number of sessions completed and test score was significant in the relearn group but was not significant in the restudy group. Students in both the relearn and restudy groups became more accurate in their predicted test score over time (pre-intervention to post-intervention), whereas those in the control group became less accurate. Within-week analysis of state indices showed a reduction in anxiety and an increase in predicted test score in both relearn and restudy groups, and a significantly higher increase in mastery and attention in the relearn (versus restudy) group. There were no differences between relearn, restudy and control groups in trait questionnaire measures. Student-reported barriers to engagement in sessions included time constraints, difficulty accessing sessions online and low motivation and monitoring and awareness.

Keywords: maths, metacognition, successive relearning, post-16 education, distributed practice, retrieval practice

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name: GEMMA CHELSEA MUNCER

Title of thesis: Exploring Factors Associated With Maths Performance in Adolescents Aged 11-18 Years

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature: Date: 04.09.2020

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to my supervisors, Dr Julie Hadwin and Dr Philip Higham. I am extremely grateful for your generosity with your time and knowledge, and I have learnt a great deal throughout this process due to the support you have given me.

Thank you to Amy Harrison, Hayden Dewis and Callum Houlker who gave up their time as Voluntary Research Assistants during the screening, data extraction and quality assessment stages of the systematic review.

I would additionally like to express my gratitude to Dr Henry Wood-Downie, Dr Jin Zhang and Professor Samuele Cortese for providing support and consultation during the meta-analysis. Thank you also to Dr Bettina Zengel, who programmed the original intervention used in the empirical study, and for providing guidance on adapting the intervention to be used in the current study.

I would also like to thank the students who took part and for the teaching staff for being so enthusiastic and welcoming. This study would not have been possible without your involvement.

A personal thank you to Vicky Mullan for being an excellent accountability partner, for the many weekend thesis sessions in Starbucks and for gracing me with your beautiful singing voice in the car journeys to and from Southampton. I am very grateful to have a friend like you.

Finally, a heartfelt thank you to my husband, Jonny Muncer. I really can’t thank you enough for your encouragement, humour, positivity and IT support during the past three years.

Definitions and Abbreviations

±	plus or minus
Λ	Wilk's lambda (index of probability distribution)
α	Cronbach's alpha (index of internal consistency)
A-Level	Advanced level (academic qualification)
ACS	Attentional control scale (Derryberry & Reed, 2002)
ACT	American college testing
ADHD	Attention deficit hyperactivity disorder
AIMS	Assessment in mathematics
AMAS	Abbreviated maths anxiety scale (Hopko et al., 2003)
ANOVA	Analysis of variance
AQA	Assessment and qualifications alliance (exam board)
ASC	Autism spectrum condition
ATMI	Attitudes toward mathematics inventory (Tapia, 1996)
β	Beta (regression) coefficient
BTEC	Business and technology education council (work-related qualification)
C	Control group (in this instance, business as usual)
CASP	Critical appraisal skills programme
CI	Confidence interval
CMA	Comprehensive meta-analysis
CYP	Child or young person/ children or young people
d	Cohen's d (effect size)

DfE	Department for Education
EBSCO	Elton B. Stephens Company
ERIC	Education Resources Information Centre
F	F statistic (from ANOVA)
FE	Further education (usually 16-18 years in the UK)
g	Hedge's g (effect size)
GAD	General anxiety disorder
GCSE	General certificate of secondary education
GPA	Grade point average
I^2	I-squared (index of heterogeneity)
ILS	Inventory of learning styles (Vermunt & van Rijswijk, 1987)
JOL	Judgment of learning
Jr. MAI	The junior metacognitive awareness inventory (Dennison et al., 1996; Sperling et al., 2002)
κ	Cohen's kappa (index of inter-rater agreement)
M	Mean
MAA	Metacognition awareness assessment (Fitzpatrick, 1994)
MAI	Metacognitive awareness inventory (Schraw & Dennison, 1994)
MANOVA	Multivariate analysis of variance
MC	Metacognition
MDT	Mathematics diagnostic test (Mathematics Diagnostic Testing Project, 2006)
MES	Metacognitive experiences scale (Efklides, 2006)
MG	Maths grade

MK	Metacognitive knowledge
MKMQ	Metacognitive knowledge in mathematics questionnaire (Efklides & Vlachopoulos, 2012).
MMAT	Missouri mastery and achievement test
MPSAT	Preliminary scholastic aptitude test
MSA-TR	Metacognitive skills and knowledge assessment (Desoete et al., 2001).
MSI	Metacognitive skills inventory (Çetinkaya & Erktin, 2002).
MSLQ	Motivated strategies for learning questionnaire (Pintrich, 1991; Wolters et al., 2006)
η^2	Eta-squared (effect size)
N	Total number of cases
n	number of cases (sub-set of the total sample)
NAEP	National Association of Education Programme
PALS	Patterns of adaptive learning survey (Midgley et al., 2000)
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
p	Probability (significance of a test statistic)
ρ	Rho (estimate of the underlying relationship between observations)
PISA	Programme for international student assessment
Q	Q statistic (index of difference in effect sizes)
Q-Q plots	Quantile-Quantile plots (index of normal distribution)
r	Pearson product-moment correlation coefficient
r_s	Spearman rank-order correlation coefficient
RGC	Research grants council

RL	Successive relearning
RS	Restudying
RVE	Robust variance estimation
SAT	Scholastic aptitude test (in the USA)
SCG	Summer course grade
SD	Standard deviation
SES	Socioeconomic status
SPSS	Statistical package for the social sciences
t	T statistic (from t-test)
T^2	Tau-squared (variation of effect sizes in different studies)
T1	Time 1 (baseline/ pre-intervention)
T2	Time 2 (post-intervention)
UK	United Kingdom
USA	United States of America
VAS	Visual analogue scale
vs.	Versus
WRAT4	Wide-ranging achievement test, 4th Edition (Wilkinson & Robertson, 2006)
YAM-5	Youth anxiety measure for the diagnostic statistics manual version 5 (Muris et al., 2017)
YPMAiM	Young pupils’ metacognitive abilities in mathematics (Panaoura & Philippou, 2003)
Z	Z score (from meta-analysis)

Chapter 1 Does Psychological Research Support an Association Between Metacognition and Maths Performance in Children and Adolescents Aged 11- 16? A Systematic Review and Meta-Analysis

1.1 Introduction

Maths and numeracy skills (the ability to use numbers and solve mathematical problems in everyday life; National Numeracy, 2020) are often used in everyday tasks, including, for example, managing money and finances, using travel timetables or following a recipe (Price & Ansari, 2013). Studies have highlighted the societal implications of poor numeracy skills. For example, Martin et al. (2014) estimated the cost of poor numeracy skills to the UK economy to be £20.2 billion per year in 2012 (approximately 1.3% of Gross Domestic Product). Further studies have found that poor numerical skills in childhood and adolescence were associated with adverse outcomes including reduced future employability and earning potential (Crawford & Cribb, 2013; Wolf, 2011) and youth offending and criminality (Meltzer et al., 1984; Parsons, 2002).

In the UK context, around one quarter (28.5%) of students who attended a state-funded school in England and sat GCSE Maths at the end of Key Stage 4 (aged 15-16 years), did not achieve a standard Grade 4 pass in 2019 (Ofqual, 2019). A GCSE Maths pass grade (or equivalent) is required to access most higher education courses and employment positions. This requirement, alongside research that has identified an association between maths skills and later outcomes, has led to the development of theoretical frameworks and research to understand factors associated with maths achievement, including the role of metacognition.

1.1.1 Defining Metacognition

Metacognition (MC) is the awareness and regulation an individual has of their own learning and thoughts (Flavell, 1979). This includes an understanding of their knowledge and strengths and weaknesses, a recognition of the strategies that may be useful to progress in a task, the monitoring of progress during a task, and behaviour change necessary to reach an outcome (Hacker et al., 1998). Metacognition has been divided into two or three parts in the literature. The dyadic model (Nelson & Narens, 1990) includes two components of MC linked to an individual’s (1) awareness of their own knowledge associated with memory and learning (so-called “monitoring”, p. 127), and (2) ability to control and manipulate cognitive behaviour (so-called “control”, p. 127). The ternary model divides metacognition into three components (e.g., Efklides, 2008). The first two components fit with those proposed in the dyadic model, namely (1) knowledge (the extent to which a person is aware of their cognition) and (2) skill (the use of strategies to monitor and regulate cognition and attain cognitive objectives). The ternary model additionally includes a third component of MC which has been termed “experience” (Flavell, 1979, p. 906). This component reflects the feelings that emerge when an individual engages in a task such as satisfaction and confidence (Efklides, 2008).

Many researchers have proposed that metacognitive monitoring/knowledge and control/skill (also termed regulation) are closely related, e.g., that monitoring is a necessary pre-requisite to regulate cognitive behaviour and behavioural responses (Baker, 1989). For example, if an individual is aware of their knowledge, then this awareness may increase their focus on what is still to be learnt (Metcalfe & Finn, 2008). Some research has found a moderate positive correlation between knowledge and regulation (e.g., Schraw et al., 2012). However, other research has identified no association in younger children (aged 8-9 years; Roebers et al., 2014), or in undergraduates who reported below-average monitoring ability (Schraw, 1994).

1.1.2 Measuring Metacognition

A systematic review of MC assessments used with 4-16-year-olds identified 84 different measures across 149 papers (Gascoine et al., 2017). Measures of MC can be categorised as online or offline. Online measures typically capture an individual’s MC through their ongoing behaviour and performance during a task (Saraç & Karakelle, 2012; Veenman & van Cleef, 2019). An example of an online measure, think-aloud protocols, is where participants verbalise their thoughts while engaging in a task (e.g., working out a maths problem). These verbalisations are recorded and are later coded according to the quality and/or quantity of metacognitive activity (e.g., see Veenman et al., 2005).

Offline measures, as defined by Saraç and Karakelle (2012), are typically questionnaires that aim to capture self-reported MC ability based on previous experiences. For example, the metacognition awareness inventory (MAI, Schraw & Dennison, 1994) asks respondents to indicate whether each of 52 statements related to learning is *true* (1 point) or *false* (0 points) for them. An example statement is, “I try to use strategies that have worked in the past”. A higher score is considered to indicate better MC. Veenman and van Cleef (2019) further defined offline measures more broadly as those that rely on self-report.

There is some uncertainty in the literature about the distinction between online and offline measures, and particularly around whether self-reported MC measures that are relevant to a specific task at hand should be classed as online or offline. These measures include Judgment of Learning scores (also known as JOLs; where an individual reports how likely they believe that they will be able to recall learnt information in the future), confidence judgments, and accuracy measures (where the difference between an individual's predicted score and their actual score is calculated; also known as calibration accuracy). Saraç and Karakelle (2012) classed these measures as online as they pertain to a specific task at hand, whereas Veenman and van Cleef (2019) considered them as offline as they are self-reported. More recently, Craig et al. (2020) categorised confidence judgments during a task (where participants reported their confidence in their answer immediately after each item before completing further items) as online and confidence judgments that were made following the completion of an entire task as offline.

Some researchers have found poor correspondence between different measures of MC. For example, Sperling et al. (2002) found a non-significant correlation between 10-15-year-olds' MC, as self-reported using the Jr MAI and as assessed using teacher-reported rating scales (scored 1-6). The discrepancy between online and offline measures has been particularly highlighted. In a later paper, Sperling et al. (2004) identified a significant correlation between undergraduates' scores on two offline self-reported questionnaires: the MAI (Schraw & Dennison, 1994) and the MC self-regulation scale of the motivated strategies for learning questionnaire (MSLQ, Pintrich, 1991, whereby respondents indicated on a 1-7 scale from *not at all true* (score 1) to *very true* (score 7) how true 12 statements, such as "I ask myself questions to make sure I know the material I have been studying", are for them). In this study, both offline questionnaires were not correlated with self-reported MC relevant to a specific task at hand, namely the accuracy of students' confidence and predicted test scores (categorised as online by Saraç & Karakelle, 2012).

A further study conducted by Tobias et al. (1999) identified a non-significant correlation (r = .17) between high school students' scores on the MAI (an offline questionnaire) and scores on a MC assessment where students were asked whether they could solve a maths question and their responses were compared with their performance. Saraç and Karakelle (2012) reported a significant positive correlation (r = .5) between two offline measures of MC: a teacher rating scale and the self-reported Jr MAI (Sperling et al., 2002), for 9-11-year-olds. However, there was a significant negative correlation between two online measures: think-aloud protocols and accuracy scores (related to a text comprehension task). In addition, all correlations between online and offline measures were non-significant (*r*s .07-.21). The lack of correspondence between online and offline measures indicates that they may be measuring different facets of MC. Sperling et al. (2002) suggested that the Jr MAI (an offline measure) is a broader measure as compared to some existing measures that focused on MC regulation. Furthermore, Saraç and Karakelle (2012) suggested that online measures may capture experience-based judgements which are implicit and unconscious, whereas offline measures may capture knowledge-based explicit and conscious judgements.

1.1.3 The Role of Metacognition in Academic Performance

Several systematic reviews/meta-analyses have explored the association between MC and academic performance. Credé and Phillips (2011) examined the association between undergraduates' responses to the MSLQ (Wolters et al., 2006) and academic performance. Student scores on the MC self-regulation scale of the MSLQ were correlated with both Grade Point Average (GPA; 98 correlations from 24 independent samples, N = 9,696, r = .22, 90% CI = [.03, .47]) and current course grade (431 correlations from 53 samples, N = 15,321, r =.23, 90% CI = [.02, .45]). Richardson et al. (2012) conducted a meta-analysis and, likewise, identified a small correlation (r = .18, 95% CI = [.10, .26]) between students' MC (as assessed using a wider range of measures than just the MSLQ) and GPA (N = 6,205 across 9 studies).

Considering research with school-aged children, Higgins et al. (2005) conducted a meta-analysis to examine the impact of thinking skills interventions (including those relating to MC) on teaching and learning outcome for 5-16-year-old children. The synthesis of effect sizes from five studies revealed a large effect size between MC intervention and cognitive outcome (e.g., test score; g = .96, 95% CI = [0.76, 1.16]). In another meta-analysis, Dent and Koenka (2016) considered the association between MC and academic performance in school-aged children aged 6-19 years. However, unlike Higgins et al. (2005), studies were included where MC was measured, rather than studies that included MC interventions. In Dent and Koenka (2016), the synthesis of 61 studies conducted in the US or Canada resulted in a small correlation between MC/self-regulated learning and academic achievement (r = .20, 95% CI = [.16, .24]). Further analysis showed that the correlation with academic achievement was moderate for online measures of MC (e.g., think-aloud protocols or behaviour observations; r = .39, 95% CI = [.34, .43]) and was small for offline MC questionnaires (r = .15, 95% CI =

[.12, .18]).

More recently, Ohtani and Hisasaka (2018) synthesised 149 effect sizes, from 118 independent samples of children and adults, to examine the association of MC with academic performance and intelligence. The authors identified a small correlation between MC and both academic performance (r = .28, 95% CI = [.24, .31]) and intelligence (r = .33, 95% CI = [.26, .39]). However, as in Dent and Koenka (2016), the measure of MC moderated the relationship between MC and academic performance. Specifically, there was a strong correlation between online MC measures and academic performance (r = .53, 95% CI = [.45, .61]) and a weak correlation between offline MC measures and academic performance (r = .23, 95% CI = [.20, .26]), with the relationship between accuracy measures and interviews and academic performance being moderate (r = .43, 95% CI = [.32, .53]). Collectively, these reviews indicate that MC and academic achievement are associated, and the strength of this association is moderated by MC measure (online vs. offline).

1.1.4 Measuring Maths Performance

Measuring maths performance in CYP has involved the use of national examinations and standardised assessments (e.g., National Foundation for Educational Research, NFER, tests), school assessment (e.g., GPA), or performance on specific maths tasks. Campbell (2005) proposed that mathematical ability is made up of two key elements: numerical ability (basic number representation and simple arithmetic and operations) and mathematical problem-solving (the generation of solutions from abstract representations of mathematical relations in context-rich problems). Other researchers have divided mathematical problems into routine (i.e., questions that test the student's knowledge of what was recently covered) and non-routine problems (i.e., those that cannot be solved immediately and often require complex multi-step problem-solving; e.g., see Mayer, 1998). Non-routine problems go beyond existing knowledge and skills, requiring the solver to plan, monitor and review their solution, all processes necessitating MC (Mayer, 1998; Verschaffel et al., 2010).

Considering the association between task complexity and MC, Stahl et al. (2006) found that undergraduate students utilised different MC skills (as assessed via a self-reported questionnaire) for questions of differing complexity. In this study, questions were categorised according to Bloom's (1956) revised taxonomy (see Anderson & Krathwhol, 2001), whereby six skill levels increase in complexity from learning that requires simple recall or recognition of relevant knowledge from long term memory to learning that requires more complex and abstract cognition (e.g., reorganising elements into a new pattern or structure).

1.1.5 The Role of Metacognition in Maths Performance

Several researchers have explored the relationship between MC and maths performance via a systematic review (Walker, 2013) or meta-analysis (Dignath & Büttner, 2008; Perry et al., 2019). The focus in these papers was on studies that investigated the outcomes of a MC intervention. For example, Kramarski & Mizrachi (2006) worked with 13-14-year-olds, in four classes, who practised mathematical problem-solving in class over four weeks. During problem-solving, the students in two (/4) classes also answered a series of self-addressed questions designed to prompt MC activity. For example, pupils described the task in their own words and considered possible useful strategies. In the post-test, which consisted of two maths word problems, pupils who engaged with MC prompts significantly out-performed pupils who did not encounter the prompts.

Walker (2013) conducted a systematic review of the effectiveness of MC interventions for maths performance. Walker (2013) concluded that, across 22 studies, students aged 18 or younger who experienced a MC intervention scored higher (vs. control) on a maths post-test. Dignath and Büttner (2008) investigated the impact of school-based self-regulated learning programmes on 5-16-year-old children’s academic performance in maths, reading and writing. Across 35 studies, there was a large synthesised effect size for the impact of self-regulated programmes based on MC strategy on academic performance. The effect size was particularly large for mathematics performance (.96 for primary school children, .23 for secondary school children). Perry et al. (2019) similarly conducted a meta- analysis to explore the relationship between the teaching of MC/self-regulation/thinking skills in schools and pupil outcomes (primarily in academic progress). The synthesised effect size from 29 studies was large (.65), and the authors concluded that teaching MC, self-regulation or thinking skills can lead to improvements in academic performance. While these review papers provide some evidence for an association between MC-related interventions and maths performance, it remains unclear whether improvements were due to a positive change in MC. For example, in the Walker (2013) review, approximately half of the included studies did not measure MC, and only one study (/22) reported the correlation between MC and maths achievement.

Several studies have reported an association between MC and maths achievement in children and adolescents (e.g., see Özsoy, 2011; van der Walt et al., 2008) though others have reported a non-significant association (e.g., see Maras et al., 2019; Young & Worrell, 2018). Some researchers have suggested that the disparity in findings across MC studies may be due to differences in how MC is conceptualised and assessed (Desoete & Roeyers, 2006; Veenman et al., 2006), with existing evidence indicating that online (vs. offline) measures are more closely linked to task performance (e.g., see reviews by Dent & Koenka, 2016; Ohtani & Hisasaka, 2018). Moreover, differences in results between studies may also be a function of how studies measure maths performance, with some researchers suggesting greater MC activity in more complex tasks (e.g., Verschaffel et al., 2010).

1.1.6 Objectives of the Current Systematic Review

A considerable body of research has explored the association between MC and maths performance. While many studies provide evidence for a significant association between MC and maths performance, some studies have not found a link. Moreover, it is unclear whether reliable associations reflect some element of individual studies concerning the specific measurements used. No review has systematically explored the strength of association between MC and maths performance from the existing evidence base.

The current systematic review and meta-analysis investigated the association between MC and maths performance in children and young people (CYP) aged 11-16. The focus in this age range reflects evidence that more complex MC skills (e.g., those associated with monitoring and evaluation) are most evident from late childhood (Dermitzaki, 2005; Veenman et al., 2006). It extends existing reviews and meta-analyses that have investigated the impact of MC interventions on maths performance (Dignath & Büttner, 2008; Perry et al., 2019; Walker, 2013) by considering studies that have directly measured the association between MC and maths performance. In addition, the meta-analysis considered whether the strength of association was different for online and offline measures of MC. Based on existing research (e.g., Dent & Koenka, 2016; Ohtani & Hisasaka, 2018), we anticipated that the association with maths performance would be stronger when MC is measured using online (vs. offline) measures. Also, because existing research has argued that the relationship between MC and performance in maths tasks would be greater for more complex problems (e.g., Verschaffel et al., 2010), we also investigated whether the complexity of the maths assessment task (simple vs. complex) was important. We anticipated that the association would be most evident with complex problem-solving.

1.2 Method

This review was carried out following the best practice guidelines for conducting a systematic review published by Siddaway et al. (2019) and the Preferred Reporting Items for Systematic Review and Meta-analysis guidelines (PRISMA; Moher et al., 2015). The protocol was determined before starting the review, and a title registration was pre-registered with the Campbell Collaboration (review number 19-009).

1.2.1 Search Strategy

Table 1 shows the terms used to search the titles, abstracts and keywords of records in four databases: Education Resources Information Centre (ERIC; 1966-2019; n = 542), Web of Science Core Collection (1990-2019; n = 880), and PsycINFO and PsycARTICLES via EBSCO (1887-2019; n = 628). Searches were initially conducted up to 15.07.19 and were repeated on 04.01.20, following data extraction, to identify papers that had become available since initial searches (n = 28). No limiters were imposed on publications (e.g., relating to date or language). The syntax was adapted to meet the requirements of each database (see Appendix A for an example search). To include unpublished research, we additionally searched ProQuest Dissertations and Theses Global (using the terms in Table 1; n = 327) and OpenGrey (n = 11). Due to input restrictions, the keywords metacogniti* AND math* were used to search OpenGrey. Finally, the reference lists of papers included in the final sample were manually screened for additional potentially relevant studies (n = 93). Two researchers independently carried out all database searches and yielded identical results. Pilot searches included three additional terms for MC (resolution, calibration and self-regulation) that were subsequently removed due to producing a high number of irrelevant papers.

Table 1

The Search Terms Inputted into Databases to Identify Relevant Studies

Metacognition	Maths	Performance
Metacogniti*	Math*	Performance
"Meta-cogniti*"	Arithmetic	Attainment
"Judgment* learn*"	Numeracy	Achievement
Metamemor*	Statistics	Grade
"Meta-memor*"		Score
Metacomprehen*		Mark
"Meta-comprehen*"		
Metaknowledge		
"Meta-knowledge"		
"Metacognitive monitoring"		
"Meta-cognitive monitoring"		
Overconfiden*		
"Over-confiden*"		
"Under-confiden*"		
"Self-assessment"		

Note. The Boolean operator "OR" was applied to the words within each column and the operator "AND" was applied to combine the three columns of words.

1.2.2 Inclusion and Exclusion Criteria

The titles and abstracts of all records that were retrieved via the systematic search (n = 1,985 after duplicates were removed) were screened against the pre-determined inclusion criteria. Studies were included if: (1) the researcher(s) reported the strength of association between MC and maths performance (e.g., by reporting the Pearson correlation coefficient). Where studies investigated the impact of a MC intervention, these were only included if the statistical relationship between MC and maths performance was reported before participants took part in the intervention (at time 1, T1 baseline) or in a control group, (2) participants were aged 11-16 (± two years if ≥ 80% of the sample were aged 11-16), (3) the study included an objective measure of maths performance (e.g., school assessment or standardised score), and (4) the study included a measure of MC. Studies were excluded if: (1) they did not include primary data, (2) the only measure of maths performance was self-reported, (3) participants were reported to have a complex neurodevelopmental disorder such as Autism Spectrum Condition (ASC), or (4) the only measure of MC was a measure of self-regulation as defined by Zimmerman (1989; due to this being a broader measure that also includes other variables such as motivation and effort). No studies were excluded based on publication status, date or language.

1.2.3 Study Selection

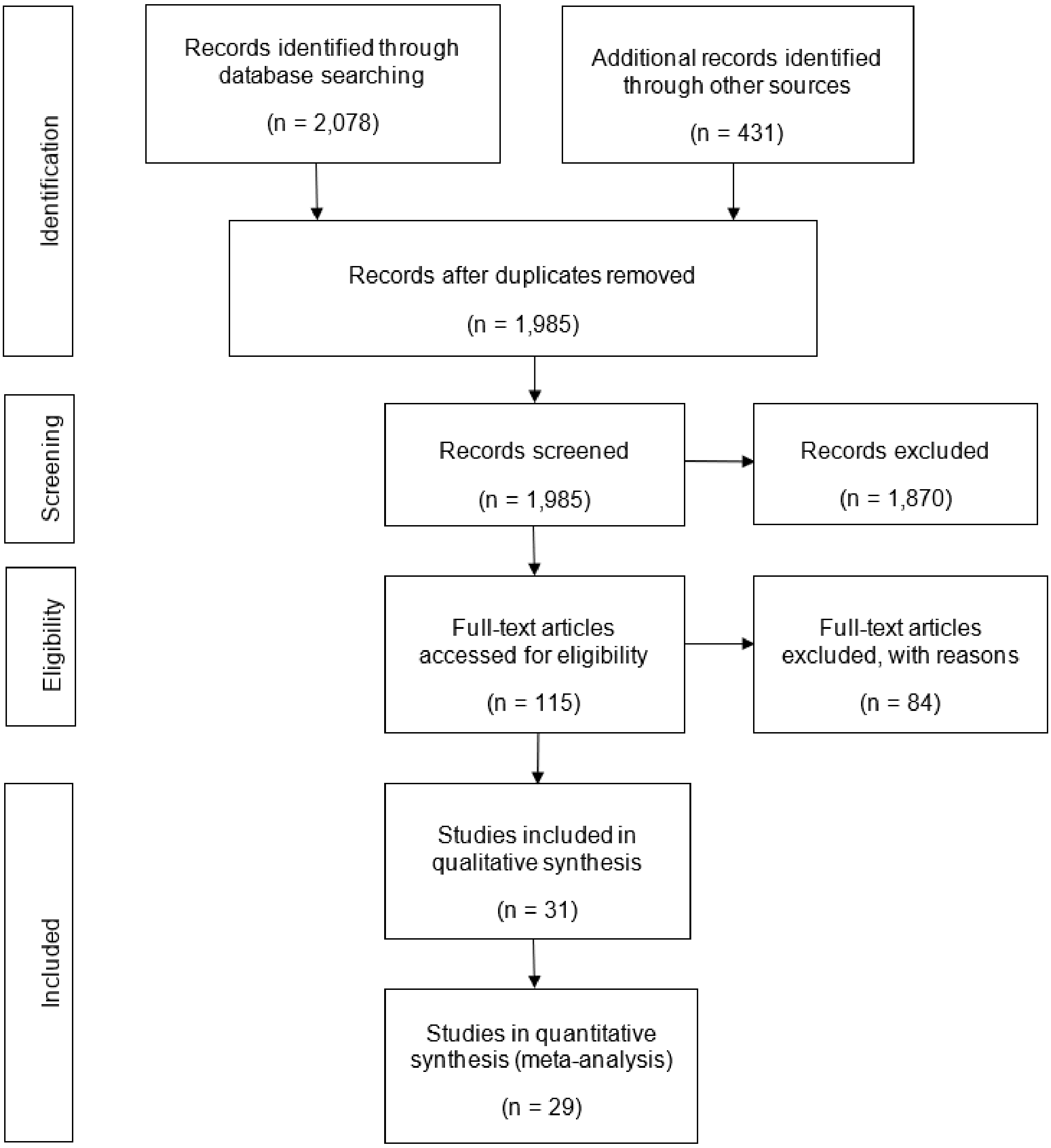
Searches yielded 2,509 records. These were exported into EndNote Desktop, and 524 duplicates were removed. Two researchers independently screened the titles, abstracts, and keywords of the remaining 1,985 records for relevance by applying the inclusion criteria stated above. This process was carried out using the web application, Rayaan. Cohen's Kappa indicated substantial agreement between the two researchers regarding the inclusion of records ($\kappa = .77$). Conflicts were resolved using the consensus model with reference to the inclusion criteria. Following this process, the full texts of 115 papers were retained for secondary screening.

Where the full text of a study was unavailable (n = 16), we contacted the corresponding author to request the paper. The authors of four studies were contacted, and two replied by sending the relevant paper. Where a contact address was not available, or the author did not reply, the paper was requested via the University of Southampton Inter-Library Loan Service (n = 14 requested, n = 10 received). Six of the retrieved papers were not in English. Two of these were translated with confidence using the online translation programme, Google Translate, and four papers were read and screened by native speakers.

Two researchers independently, and blind to the decision of the other, read the full texts of the 115 records to further consider each record's eligibility to the current review. Cohen's Kappa indicated substantial agreement between the two researchers regarding the inclusion of studies at this stage ($\kappa = .61$). As previously, any disagreements were resolved using the consensus model, and on two occasions, discussions took place with a third researcher to further consider a study's inclusion (see Appendix B). To avoid duplication of samples, where data was reported from the same participants in more than one study, the paper that reported the largest number of participants was included. If the number of participants was equal, the earliest study was included. Where the author(s) had measured MC and maths performance but had not reported the association between the two (n = 6), we contacted the author(s) to request this information. Two authors responded, one author provided the required data, and one reported that this information was not available. In total, 84 papers were excluded during secondary screening (see Appendix C). The procedure of how the final sample of studies was reached in the qualitative synthesis (n = 31) and quantitative synthesis (n = 29) is illustrated in the PRISMA flow diagram (Moher et al., 2009) in Figure 1.

Figure 1

The Process by Which the Final Sample of Studies was Reached in the Current Systematic Review and Meta-Analysis



1.2.4 Data Extraction

The data of included papers were extracted by the first researcher. A second researcher checked the extracted data from 35% of studies (11 of 31) and agreed that this was accurate. Where only some participants within a paper fitted the inclusion criteria (for example, a typically developing control group in a study primarily focused on individuals with ASC), data were extracted for these participants only. For longitudinal studies (n = 3), T1 data was extracted.

Data extracted included: (1) publication information (author(s), year, record type, the country in which the study took place, and funding received), (2) participant characteristics (number of participants, age, gender, and any other defining characteristics, e.g., learning disability), (3) study design (aim, design and setting), (4) measures (of maths performance, MC and other variables), and (5) key results (the statistical association between MC and maths performance and other key findings).

1.2.5 Quality Assessment

Two researchers independently assessed the methodological quality of each included study using the Critical Appraisal Skills Programme (CASP) Cohort Study Checklist (2018). This checklist consists of 12 items to consider the validity of the study (section A), the results (section B), and the usefulness of the results (section C). As not all studies were longitudinal/ follow-up studies, the checklist was adapted by removing both parts of question six, which asked whether the follow-up of subjects was complete and long enough. The adapted questionnaire contained 11 questions; one of these questions was comprised of two parts which meant that there were 12 items in total. Nine of these items required a *yes*, *can't tell* or *no* response, with *yes* indicating high-quality research. The remaining three items required open-ended responses. Although the authors of the checklist have not published a scoring system, the number of possible *yes* responses in the adapted checklist was 0-9. In the current review, this was used as an arbitrary scoring system. A table of the adapted checklist items and how each item was interpreted is in Appendix D.

1.2.6 Analytic Strategy

We used comprehensive meta-analysis software (CMA; Borenstein et al., 2014) to calculate the pooled correlation coefficient of the association between MC and maths performance across studies. All analyses used a random-effects model due to the assumption that the true effect sizes of the included studies were not identical. The analysis was weighted by the reciprocal of the variance of the Fisher's Z effect size, so that larger studies contributed greater weight to the summary effect size.

For 13 (/29) studies, more than one correlation between MC and maths performance was reported. The reporting of multiple associations typically reflected multiple measures of MC (e.g., self-reported questionnaire and JOL) or maths performance (e.g., school assessment and standardised test). To ensure that the assumption of independence of effect sizes was met, the average of the effect sizes was calculated for these 13 studies. The primary analysis was also repeated to examine any differences using a method developed specifically for dependent effect sizes- robust variance estimation (RVE; Hedges et al., 2010), using an SPSS macro (Tanner-Smith & Tipton, 2014).

Further subgroup analyses were conducted in CMA to explore the impact of moderators on the association between MC and maths performance. For these analyses, the within-group estimates of Tau-squared (T^2) were pooled, and subgroups were combined using a random-effects model. The moderators investigated included the measurement of MC (online vs. offline) and the complexity of the maths assessment task (complex vs. simple). Where there were studies that combined two codes for a single moderator (e.g., the participants within a study completed both online and offline measures of MC), these studies were removed from the analysis.

1.3 Results

Of 1,985 papers (115 full texts) screened for eligibility, 31 studies met the inclusion criteria. Details of the included studies, including quality assessment ratings, are displayed in Table 2.

Table 2

Details of the Studies That met the Inclusion Criteria Including Publication Information, Participant Characteristics, Design and Setting, Measures and Findings, and Quality Assessment Ratings

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
1	Ahmed et al., (2013)	T1 data only *Not reported*	Journal article	The Netherlands	495 (252, 51%) *M = 12.8 years*	N/A	Longitudinal *Two secondary schools in two middle-income suburban communities (21 classes*)	School assessment (graded 1-10)	Items from the motivated strategies for learning questionnaire (MSLQ; Wolters et al., 2006)	*r* = .36 (sig., *p* < .01)	9
2	Aşık & Erktin (2019)	N/A *Not reported*	Journal article	Turkey	406 (195, 48%) *M = 14 years*	N/A	Correlational *Three public and two private inner-city schools*	Three word problems on algebra/ arithmetical operations. Two were taken from the 7th-grade textbook. Responses were	Metacognitive skills inventory (MSI; Çetinkaya & Erktin, 2002) The metacognitive experiences scale (MES; Efklides, 2006) before the	MSI: *r* = .40 (sig., *p* < .01) MES: *r* = .53 (sig., *p* < .01) *M* = .47	6

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
								scored from 0-4 based on the Holistic Scoring Rubric (Aschbacher et al., 1995) where 0 indicates an entirely incorrect response and 4 indicates an entirely correct response	three maths problems		
3	Bishara & Kaplan (2018)	N/A *Not reported*	Journal article	Israel	60 (26, 43.3%) *Not reported but assumed to be 13-14 years*	30 adolescents who had learning disabilities and were enrolled in mixed classes in a mainstream school	Correlational (group design) *One public middle school*	Math Aptitude Test (Haddad Center, 2012) consisting of 10 questions set by the Ministry of Education's curriculum for 7th-grade. Graded 0 to 10.	Questionnaire to examine metacognitive knowledge of math (Kramarski et al., 2005, based on the questionnaire by Montague & Bos, 1990)	Before problem-solving: r = .69 (sig., p < .001) During problem-solving: r = .83 (sig., p < .001) After problem-solving: r = .69 (sig., p <	6

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
										.001) Beliefs about solving maths problems: *r* = .84 (sig., *p* < .001) *M r* = .76	
4	Callan & Cleary (2019)	N/A *Dissertation award from the Society for the Study of School of Psychology*	Journal article	The USA	96 (54, 56.2%) *Not reported but assumed to be 13-14 years*	90.7% met the criteria for a free or reduced lunch	Correlational *One urban school*	Three multi-step algebra word problems from a National Association of Education Programme (NAEP) past assessment	After three maths problems, pupils responded on a 1-7 scale to, “How sure are you that you solved this problem correctly?” This was compared with their actual performance to calculate an accuracy score	*r* = .37 (sig., *p* < .001)	8
5	Callan et al., (2016)	N/A *Not reported*	Journal article	Used PISA data (2009) from 63 countries	475,460 (239,156, 50.3%) *15 years*	N/A	Correlational *Schools in 63 countries*	The PISA multiple-choice international achievement test	The PISA metacognitive indexes (the index of understanding and remembering and the index of	*r* = .46 (sig., *p* < .001)	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
									summarising)- students were asked how useful they thought various reading strategies were to solve a reading text		
6	Chiu et al., (2007)	N/A *Partly funded by a University of Hong Kong direct grant and RGC grant*	Journal article	Uses PISA data (2000) from 34 countries	88,590 (not reported but sample included males and females) *15 years*	N/A	Correlational *Schools in 34 countries*	PISA multiple-choice achievement test	PISA self-reported metacognitive strategy use questionnaire	*r* = .04 (not sig.)	9
7	Erktin (2004)	Study 2 data for high school students only *Partly*	Journal article	Turkey	100 (not reported but approximately 50% of the total	N/A	Correlational *One inner-city private high school*	A researcher-designed multiple-choice test on probability	Metacognition Inventory (Çetinkaya, 2000; Çetinkaya & Erktin, 2002)	*r* = .42 (sig., *p* < .05)	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
		funded by a Bogazici University research fund grant			sample including university students were reported to be female) *Not reported but assumed to be 15-16 years*						
8	Fadlelmula et al., (2015)	N/A *Not reported*	Journal article	Turkey	1019 (481, 47.2%) *Assumed to be 13 years*	N/A	Correlational *11 inner-city public schools (34 classrooms)*	A 10-item multiple-choice test consisting of items covered within topics students had previously studied (numbers, geometry & algebra)	Items (planning, monitoring, regulating) from the metacognitive self-regulation scale of the motivated strategies for learning questionnaire (MSLQ; Pintrich, 1991)- participants were asked to think specifically about	$\beta = -1.41$ (not sig.)	6

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
									maths when answering items		
9	Fitzpatrick (1994)	N/A *Not reported*	Conference paper	The USA	100 (50, 50%) *Not reported but assumed to be 14-17 years*	Students were selected based on their Preliminary Scholastic Aptitude Test (MPSAT) score: 50 participants' scores fell on or below the 40th centile, and 50 participants' scores fell on or above the 60th centile	Correlational (group design) *Two urban public and three private high schools*	Six multiple-choice questions from the 1987 SAT maths test Three questions were word problems	The researcher-designed metacognition awareness assessment (MAA)	*r* = .28 (not sig.)	5
10	Fusco (1995)	N/A *Not reported*	Doctoral thesis	The USA	30 (30, 100%) *13-16 years, most were 14-15 years*	Students were selected based on how they attributed maths problem-solving performance:	Correlational (group design) *One urban Catholic high school*	One non-routine word problem. Scored from 1 to 5 (5 indicating a completely correct answer)	Think-aloud protocols and behaviour observations during problem-solving	*r* = .68 (sig., *p* < .001)	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
						ten attributed strategy, ten attributed effort and ten attributed unknown causes					
11	Harris (2015)	Data from students in grades 6, 7 and 8 only *Not reported*	Doctoral thesis	The USA	27 (not reported but sample included males and females) *11-14 years*	Six students had learning disabilities. Students with language impairments, autism and intellectual giftedness were excluded	Correlational *One public Montessori school*	A standardised grade-level skills assessment in mathematics (AIMS web; Pearson Education, 2008)	The junior metacognitive awareness inventory (Jr. MAI; Dennison et al., 1996)	β = -.30 (not sig.)	7
12	Hassan & Rahman (2017)	N/A *Not reported*	Journal article	Malaysia	333 (not reported) *Not reported but assumed to be 15-16 years*	N/A	Correlational *Ten secondary schools*	Not reported	Questionnaire adapted from the metacognitive awareness inventory (MAI; Schraw & Dennison, 1994)	β = .48 (sig., $p < .001$)	2
13	Ichihara	N/A	Journal	Japan	543 (264,	N/A	Correlation	School	Metacognitive	$r = .31$ (sig.)	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
	& Arai (2006) (*Japanese*)	*Not reported*	article		48.6%) *Not reported but assumed to be 12-14 years*		al *One public junior high school*	assessment (end of term marks, 0-100 points)	questionnaire (Sato & Arai, 1998).		
14	Maras et al., (2019)	Data for typically developing students in the control condition only *Funded by a grant from The Economic and Social Research Council*	Journal article	The UK	49 (18, 36.7%) *M = 13.4 years (11-15 years)*	Participants were working at age-related expectations in maths	Intervention experiment *One secondary school*	Three mental maths questions (block one) selected from past papers of national UK examinations or revision workbooks. Questions were scored as correct or incorrect	Four questions relating to awareness of own performance, confidence, and strategies used during the maths task	r_s = -.12 (not sig.)	8

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
15	Martín et al., (2008)	Data at time 1 only *Not reported*	Journal article	Spain	965 (435, 45.1%) *Most were 12-13 years old*	N/A	Longitudinal *17 private and ten public inner-city secondary schools*	A multiple-choice researcher-designed test	A test that included four scales: meta-comprehension (accuracy of predicted score), verification of one's results, the consciousness of the strategies one uses and consciousness of one's own comprehension (Moreno, 2002)	β = 3.24 (sig., $p < .001$)	9
16	Ning (2016)	N/A *Funded by The Office of Educational Research National Institute of Education*	Journal article	Singapore	873 (441, 50.52%) *M = 15.36 years*	N/A	Correlational *10 schools*	A multiple-choice standardised test.	The junior metacognitive awareness inventory (Jr. MAI; Sperling et al., 2002)	Knowledge of cognition scale: $r = .00$ (not sig.) Regulation of cognition scale $r = .07$ (not sig.) $M\ r = .04$.	6

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
		Nanyang Technological University									
17	Özcan (2016)	N/A *Not reported*	Journal article	Turkey	268 (not reported after attrition but 145 of 323 of the original sample, 45%, were female) *Not reported but assumed to be 11-14 years*	N/A	Correlational *Two inner-city public schools*	Six maths problems related to problems in the students' course textbooks. Each problem was scored from 0 to 4 with 4 indicating a wholly correct and clear answer	The young pupils' metacognitive abilities in mathematics (YPMAiM; Panaoura & Philippou, 2003) The metacognitive experience scale (MES; Efklides, 2006)	YPMAiM: $r = .17$ (sig., $p < .05$) MES: $r = .33$ (sig., $p < .01$) $M\ r = .25$	7
18	Özcan & Eren Gümüş (2019)	N/A *Not reported*	Journal article	Turkey	517 (265, 51%) *Not reported*	N/A	Correlational *Two inner-city public*	Four multi-step word problems on linear equations taken	The metacognitive experience scale (MES; Efklides, 2006)	$r = .5$ (sig., $p < .01$)	8

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
					but assumed to be 12-13 years		*middle schools*	from the seventh-grade coursebook. Responses were scored using the Holistic Scoring Rubric (0-4 where 4 is a completely correct answer)			
19	Özsoy (2011)	N/A *Not reported*	Journal article	Turkey	242 (134, 55.4%) *M = 11.3 years*	N/A	Correlational *Six urban public schools*	Mathematics Achievement Test designed by the researcher (Özsoy, 2005)	Metacognitive skills and knowledge assessment (MSA-TR; Desoete et al., 2001)	r = .65 (sig., p < .01)	6
20	Peng et al., (2014)	N/A *Not reported*	Journal article	China	438 (256, 58.4%) *15-16 years*	N/A	Correlational *One inner-city high school*	Final test score (not reported but assumed to be an end of year school-administered test)	Items from the test-taking strategies questionnaire (Hong & Peng, 2004)	Planning scale: r = .11 (sig., p < .05) Self-checking scale: r = .05 (not sig.) Strategy selection scale: r = .04 (not sig.)	6

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
21	Sink et al., (1991)	N/A *Not reported*	Conference paper	The USA	62 (34, 55%) *M = 11.6 years (11-13 years)*	N/A	Correlational *One middle school in a small town*	School assessment (teacher-designed maths test) The Missouri Mastery and Achievement Test (MMAT; Missouri Department for Elementary and Secondary Education, 1990)	Accuracy of predicted test score (to actual achieved score)	*M* r = .07 School assessment: r = .29 (sig., p < .05) MMAT: r = .43 (sig., p < .01) *M* r = .36	8
22	Tian et al., (2018)	N/A *Received funding from the Young Scholars Foundation of*	Journal article	China	569 (324, 56.9%) *M = 16.39 years*	N/A	Correlational *One high school*	Three successive mathematics examinations	Metacognitive knowledge in mathematics questionnaire (MKMQ; Efklides & Vlachopoulos, 2012)	Separate correlations reported for each of the three maths exams. MK of self (easiness/fluency): rs ≥	8

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
		Beijing Science 12th Five-Year Plan (2015)								.19, all sig., *p*s < .05 MK of self (difficulty/lack of fluency): *r*s ≥ .14, all sig., *p*s < .05 MK of tasks (easy/low demands): *r*s -.05 to -.06, all not sig., *p*s > .05 MK of tasks (difficult/high demands): *r*s ≤ -.17, all sig., *p*s < .05 MK of strategies (cognitive/metacognitive strategies): *r*s ≥ .026, all sig., *p*s < .05 MK of	

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
										strategies (competence-enhancing strategies): *r*s ≥ .16, all sig., *p*s < .05 MK of strategies (avoidance strategies): *r*s ≤ -.15, all sig., *p*s < .05 *M r* = .18 *(scores reversed where necessary)*	
23	van der Stel & Veenman (2014)	Time 1 data only *Not reported*	Journal article	The Netherlands	25 (not reported but sample included both males and females) *13 years*	Students with known learning or conduct disorders were excluded	Longitudinal *One urban secondary school*	Five word problems adapted from an age-appropriate maths textbook. Participants were awarded one point for using the correct	Think-aloud protocols were analysed for use of metacognitive skills, according to the quantity (frequency) and quality of utterances	(Semi-partial correlations account for intellectual ability) Quality of utterances: *r* = .70 (semi-partial = .30)	9

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
								procedure and one point for a correct answer		(sig., $p < .01$) Quantity of utterances: $r = .73$ (semi-partial = .30) (sig., $p < .01$) *M* $r = .72$	
24	van der Stel et al., (2010)	N/A *Not reported*	Journal article	The Netherlands	59 (36, 61%) *13-15 years*	N/A	Correlational (group design) *Two suburban schools*	Five (2nd years) or six (3rd years) word problems adapted from a commonly used math textbook. Students could achieve a maximum of 10 points per question	Think-aloud protocols were analysed on the use of metacognitive skills, according to the quantity (frequency) and quality of utterances	*Second-year participants* Quality of utterances: $r = .53$ (sig., $p < .01$) Quantity of utterances: $r = .29$) (not sig.) *Third-year participants* Quality of utterances: $r = .78$ (sig., $p < .01$) Quantity of utterances: r	9

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
										= .40 (sig., $p < .01$) For total population: $r = .51$	
25	van der Walt et al., (2008)	N/A *Not reported*	Journal article	South Africa	339 (199, 58.7%) *M = 13.64 years (12-17 years)*	N/A	Correlational *Six urban schools*	School assessment (exam score) A geometry problem (calculate the surface area of a parallelogram within a rectangle)	The Lucangeli-Cornoldi instrument (Lucangeli & Cornoldi, 1997) was used while pupils were solving the geometry problem	School assessment: Prediction of success: $r_s = .26$ (sig., $p < .05$) Degree to which learner could monitor steps in the solution: $r_s = .21$ (sig., $p < .05$) Evaluation of success: $r_s = .30$ (sig., $p < .05$) Reflection on solution: $r_s = .11$ (sig., $p <$	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
										.05) Geometry problem: Prediction of success: r_s = .37 (sig., p < .05) Degree to which learner could monitor steps in the solution: r_s = .33 (sig., p < .05) Evaluation of success: r_s = .39 (sig., p < .05) Reflection on solution: r_s = .04 (not sig.) *M r* = .25	
26	Veenman et al.,	Data from the	Journal article	The Netherlan	30 (not reported)	Students were selected based	Interventio n	Three mathematical	Systematic behaviour	Systematic observation:	5

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
	(2000)	control condition only *Not reported*		ds	*12-13 years*	on anxiety questionnaire scores: 20 reported high anxiety and 10 reported low anxiety	experiment *One secondary school*	word problems, adapted from Henfi (1990).Scored as correct (1 point) or incorrect (0 points)	observations Analysis of think-aloud protocols for quality of metacognitive skilfulness during problem-solving	r = .41 (.38 corrected for extreme anxiety groups) (sig., p < .05) Think-aloud protocols: r = .52 (.50 corrected) (sig., p < .01) M r = .47	
27	Veenman et al., (2005)	Data from the control condition only *Not reported*	Journal article	The Netherlands	41 (approximately 50%) *12-13 years*	N/A	Intervention experiment *Two urban secondary schools*	Performance on three mathematical word problems, adapted from Henfi (1990). Scored as correct (1 point) or incorrect (0 points) Grade Point Average (GPA)	Systematic behaviour observations and analysis of think-aloud protocols	(Semi-partial correlations account for intellectual ability) Word problems: r = .48 (semi-partial = .47) (sig., p < .01) Maths GPA: r = .40	9

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
								for maths at the end of the previous school year		(semi-partial= .30) (sig., p < .01) *M r* = .44	
28	Veenman & Spaans (2005)	N/A *Not reported*	Journal article	The Netherlands	31 (18, 58.1%) *M = 13.9 years*	Pupils were selected based on intelligence test score: 16 pupils scored as low in intelligence and 15 scored highly	Correlational (group design) *One urban secondary school*	Six maths problems adapted from Henfi (1990). Scored as correct (1 point) or incorrect (0 points	Systematic behaviour observations	(Semi-partial correlations account for intelligence) r = .75 (semi-partial = .35, sig., p < .01) (corrected for extreme intelligence groups, r = .66, semi-partial= .45)	7
29	Walker (2013)	Data from time 1 only *Not reported*	Doctoral thesis	The UK	18 (11, 61%) *M = 13.15 years (13-14 years)*	All had made age-appropriate progress in maths by the end of primary school but had not made	Intervention experiment *One secondary school*	The oral maths scale and computation scale from the wide-ranging achievement test, 4th Edition (WRAT4;	The junior metacognitive awareness inventory (Jr MAI; Sperling et al., 2002)	r = -.14 (not sig.)	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
						expected progress at secondary school		Wilkinson & Robertson, 2006)			
30	Yap (1993)	Data for grade 8 students only. Data were taken from NAEP motivation studies *Not reported*	Doctoral thesis	The USA	591 (285, 48.2%) *Not reported but assumed to be 13-14 years*	N/A	Correlational *18 schools*	The National Assessment of Educational Progress math tests (standardised)- 41 multiple-choice items	The self-checking subscale from the state self-regulatory inventory (mainly developed by O'Neil, later published as O'Neil & Abedi, 1996)	r = .21 (sig., p < .01)	9
31	Young & Worrell (2018)	Data from study 1 only *No funding received*	Journal article	The USA	179 (maths grade and GPA available) / 183 (MDT and summer	Attending a university summer programme for academically talented youth	Correlational *One university summer programme*	Most recent mathematics school grade (MG) Academic Grade Point Average for mathematics (GPA)	The junior metacognitive awareness inventory (Jr. MAI; Sperling et al., 2002)	MG: r = .05 (not sig.) GPA: r = .00 (not sig.) MDT: r = -.12 (not sig.)	7

	Publication Information				Participant Characteristics		Design & Setting	Measures		Findings	Quality Rating
Study	Author(s) (Year)	To Note *Funding*	Publication Type	Country	N (Females) *Age*	Defining Characteristics	Design *Setting*	Maths Performance	Metacognition (MC)	Association Between MC and Maths Performance (Sig.)	Score (/9)
					course data available) (97, 54%) *M = 13.29 years (11-17 years)*			Mathematics diagnostic test (MDT; Mathematics Diagnostic Testing Project, 2006) Final course grade in a mathematics course at the end of the summer program (SCG)		SCG: $r = .01$ (not sig.) M $r = -.02$	

Note. N = number of participants, MC = metacognition, M = mean, T1 = time 1 (pre-intervention) , T2 = time 2 (post-intervention), sig = statistical significance/statistically significant, $p = .05$ = 95% confidence in significance, $p < .01$ = 99% confidence in significance, r = Pearson correlation coefficient, β = beta coefficient

1.3.1 Qualitative Results

1.3.1.1 Publication Information

1.3.1.1.1 Type

Studies included in this review are peer-reviewed journal articles (n = 25), unpublished doctoral theses (n = 4) and conference research papers (n = 2).

1.3.1.1.2 Date

Included studies were published/made available between 1991 and 2019. Four papers were published from 1991-1995; all originated from the USA and were either doctoral dissertations (n = 2) or conference papers (n = 2). Eight papers were published from 2000-2009. Nineteen papers were published from 2010-2019, and most of these were published in the second half of the decade (n = 13), reflecting the increasing interest in this area of research.

1.3.1.1.3 Country

Two studies used data collected by the Programme for International Student Assessment (PISA). One included data from students in 34 countries using the PISA 2000 database (6[1]), and the other used data from 63 countries using the 2009 database (5). Other studies were conducted in 11 countries including the Netherlands (n = 6), Turkey (n = 6), Israel (n = 1), the United States of America (USA; n = 7), Malaysia (n = 1), the United Kingdom (UK; n = 2), Japan (n = 1), Spain (n = 1), Singapore (n = 1), China (n = 2) and South Africa (n = 1).

1.3.1.1.4 Study Aim

Aims are recorded in Appendix F. Twenty studies specifically investigated the relationship between MC and maths performance (1, 2, 4-11, 13, 15, 17-22, 25, 31). Alternative aims included, for example, to explore the development of MC skills over time (23) and the role of MC in test anxiety (26).

1.3.1.1.5 Funding

The authors of 25 studies did not report whether the study received funding. One study reported that no funding was received (31). For three studies, the researchers reported that they received university research grants (4, 6, 7) and in two studies, funding was received

[1] Identifying study number as shown in Table 2.

from public organisations (14, 16).

1.3.1.2 Participant Characteristics

1.3.1.2.1 Sample Size

Collectively, the 31 papers included 572,559 participants. Aside from the two studies that used PISA data and involved high numbers of participants (88,590 and 475,460), the number of participants ranged from 18 (29) to 1,019 (8).

1.3.1.2.2 Age

All participants were aged 11-17 years. Most studies (n = 22) included CYP who were within the same academic year group (1, 2-8, 10, 12, 15, 16, 18-23, 26, 27, 29, 30). Nine studies included participants of a broader age range (9, 11, 13, 14, 17, 24, 25, 28, 31). The largest reported age range was 11-17 years (31). There was inconsistent reporting of age; some studies reported mean age and/or age range, and some studies did not report age (n = 9). In these instances, due to the participants' stage of schooling, it was assumed that they were aged between 11 and 16 years, with a minority being 17 years old. The lowest reported mean age was 11.3 years (19), and the highest was 16.39 years (22).

1.3.1.2.3 Sex/Gender

Twenty-six (/31) studies reported the sex/gender split of participants. In 24 studies, an exact number or percentage of males/females was reported, and in two studies, it was reported that the gender split was approximately equal (7, 27). One study included females only (9). Apart from this study, the lowest reported percentage of female participants was 43.3% (3), and the highest was 61% (24, 29). Six studies did not report participant gender (6,11, 12, 17, 23, 26). Including only those studies that reported sex/gender, the total participant sample was 483,145, and of these, 243,061 (50.3%) were female.

1.3.1.2.4 Socioeconomic Status

Socioeconomic status (SES) was inconsistently reported. Where information was available (n = 18; 1, 2, 4-7, 9, 11, 13, 15, 20, 21, 23, 24, 27, 29-31), SES was varied. One study reported that 90.7% of CYP met the criteria for a free or reduced school lunch (4), suggesting a relatively disadvantaged SES. Further studies (n = 6) included CYP who were attending private schools (2, 5-7, 9, 15), indicating advantaged SES.

1.3.1.2.5 Specific Characteristics

Some studies included participants with specific characteristics, e.g., CYP who had made below age-related mathematics progress at secondary school (30) or who were "academically talented" (31, p. 263). Some participants were selected based on further characteristics such as whether or not they had a learning disability (3), were of below-average or above-average intelligence (9, 28), had low or high levels of anxiety (26) or attributed their problem-solving performance to strategy, effort or unknown causes (10).

1.3.1.3 Design and Setting

1.3.1.3.1 Research Design

Most studies used a correlational design (n = 24). Five of these studies further incorporated a between-groups aspect to the design to examine the differences in the association between MC and maths performance between groups, according to intelligence (high vs. low; 3, 9, 28), age (2^{nd} years vs. 3^{rd} years; 24) or attribution style of maths performance (attribution to strategy vs. effort vs. unknown causes; 10). Three studies were longitudinal (1, 15, 23) and four studies investigated the impact of interventions (14, 26, 27, 29).

1.3.1.3.2 Setting

Data were collected in schools, except for one study where data were collected during a university summer programme (31). Studies were conducted in state (n = 8; 3, 4, 8, 11, 13, 17-19) and private (n = 1; 8) schools, or in a combination of state and private schools (n= 5; 2, 5, 6, 9, 15). Fifteen studies did not report whether the school(s) that participants attended were state or private (1, 10, 14, 16, 20-30). Most studies included participants from one or two schools (n = 19), although several studies (n = 12) included participants across more than two schools. One study included participants from 27 schools (15). The two studies using PISA data included participants across considerably more schools (5, 6), though the number was not reported. Of those studies that reported school setting, all were in urban or suburban settings (n = 18; 1, 2, 4, 7-10, 15, 17-21, 23-25, 27, 28).

1.3.1.4 Measures

1.3.1.4.1 Maths Performance

Measures of maths performance included performance on word problems (n = 11), standardised assessment (n = 10), school assessment (n = 8), mental maths questions (n = 1) and course grading (n = 1). One study did not report how maths performance was measured (12). Four studies measured maths performance using more than one method (21, 25, 27, 31), and one study used three successive maths examinations (22).

Eleven studies assessed participants' mathematical performance using word problems taken or adapted from age-appropriate textbooks or past national assessments (2, 4, 10, 17, 18, 23-28) and two studies utilised measures in which participants had to complete some word problems alongside other less complex calculations (8, 9). The number of word problems used ranged from one to six. Participant responses were either scored as *correct* (1 point) or *incorrect* (0 points) or scores were awarded based on a Holistic Scoring Rubric (Aschbacher et al., 1995) or a similar method that involved awarding points on a scale, depending on evidence of correct procedure, and accuracy and clarity of answers.

Ten studies used a standardised assessment to measure maths performance (3, 5, 6, 9, 11, 16, 21, 29-31). Examples include the oral maths and computation scales from the wide-ranging achievement test (WRAT 4^{th} Edition, Wilkinson & Robertson, 2006; 29) and the mathematics diagnostic test (Mathematics Diagnostic Testing Project, 2006; 31). Eight studies measured maths performance using school assessment (1, 13, 20-22, 25, 27, 31); in six of these, this was performance in a school examination (1, 13, 20-22, 25) and in two instances, this was GPA (27, 31). Two studies used a series of mental maths questions from past examinations or revision workbooks (14) or performance on a university-facilitated maths course (31).

The measures used differed in their complexity from questions requiring the use of complex problem-solving skills (e.g., word problems) to those requiring the use of learnt algorithms or simple recall. In some studies, there was insufficient reported information to identify the complexity of maths items. However, four studies included sufficient information to indicate that the measure of maths performance did not require complex problem-solving or reasoning (11, 14, 29, 30), e.g., participants selected from four options to complete sentences based on recall of knowledge from long-term memory.

Several studies included items from a single mathematical topic such as probability (7), linear equations (18), algebra (4) or geometry (25). Some studies included measures covering two (2, 29), three (8, 9, 26-28) or four or more topics (5, 6, 19). Several studies did not report which topics were covered (1, 3, 10-17, 20-24, 31).

1.3.1.4.2 Metacognition

Studies used a range of tools to measure MC; at least 19 distinct instruments/measures were reported. Sixteen studies used self-report questionnaires. In nine studies, participants were asked to make a judgment of their learning (e.g., to predict whether they had answered a maths question correctly or incorrectly, and this was compared with their performance). Five studies used think-aloud protocols, and in four studies, participants were observed during maths problem-solving (and observers awarded a score based on evidence of MC).

Studies employed a mixture of online and offline measures. Using Saraç and Karakelle's (2012) distinction (i.e., that online measures pertain to a specific task at hand), 14 studies used online measures only (4, 9, 10, 14, 15, 18, 19, 21, 23-28) and 15 studies used offline measures only (1, 3, 5-8, 11-13, 16, 20, 22, 29-31). A further two studies used a combination of online and offline measures (2, 17). Most offline measures included general MC questionnaires (1, 6, 7, 11-13, 16, 20, 29, 30). Three of these studies (3, 17, 22) used questionnaires that were domain-specific to assess MC specifically relating to maths (e.g., questionnaire to examine metacognitive knowledge of math; Kramarksi et al., 2005; 3). One study adapted the wording of a general questionnaire measure (the Jr MAI, Sperling et al., 2002) to relate items to MC thinking around maths (31). A further study asked participants to think about maths while responding to a general questionnaire (8). One study used a domain-specific questionnaire relating to reading (5).

Of the 17 studies that used offline measures of MC, 13 different instruments were used. The Jr MAI (Dennison et al., 1996; Sperling et al., 2002) was used in four studies (11, 16, 29, 31), although the question wording was altered to relate specifically to maths in one study (31). One study additionally used items adapted from the MAI (Schraw & Dennison 1994; 12), which the Jr MAI was adapted from for use with CYP, and another used the young pupils' metacognitive abilities in mathematics (Panaoura & Philippou, 2003) which was partly based on the MAI and Jr MAI (17). Two studies used items from the MSLQ (Pintrich, 1991; Wolters et al., 2006; 1, 8).

Of the 16 studies that used online measures, six studies measured MC by participants either vocalising their thought processes (think-aloud protocols) and/or being observed during a maths task (10, 23, 24, 26-28). An additional six studies measured MC during a maths task and CYP were, e.g., asked to predict their performance immediately before and/or after each maths question during a series of maths questions (2, 4, 14, 15, 17, 18). In another two studies, participants made a JOL or confidence judgment specifically relating to either an upcoming maths task (21) or following the completion of an entire maths task (9). A further two studies used online measures of MC that incorporated both prospective and retrospective elements (e.g., CYP predicted their performance both before and after engaging in a maths task; 19, 25).

Considering the chronology of MC measures, online measures were used in most papers from 1990-1999 (3/4) and 2000-2009 (5/8). More recently (2010-2019), most studies used offline measures (11/19, with an additional two studies that used both an online and offline measure).

1.3.1.4.3 Other Variables Measured

In addition to MC and maths performance, studies measured a range of further variables. Examples include maths-related emotions (1), locus of control (3, 21), homework behaviour (17), motivation (18, 20, 22) test anxiety (20), maths anxiety (18), self-efficacy (18, 20, 22) and working memory (29).

1.3.1.5 The Association Between Metacognition and Maths Performance

Nineteen of the 31 papers reported only a significant positive association(s) between MC and maths performance (*p* <. 05; 1-5, 7, 10, 12, 13, 15, 17-19, 21, 23, 26-28, 30). Eight studies reported only an association(s) that was not statistically significant (6, 8, 9, 11, 14, 16, 29, 31). Four studies reported mixed findings (i.e., more than one correlation was reported due to measuring MC and/or maths performance using more than one measure/scale, and at least one correlation was significant, and one correlation was non- significant; 20, 22, 24, 25). Medium effect sizes were most commonly reported (1, 2, 4, 5, 7, 12, 13, 17, 21, 24-27).

1.3.1.5.1 The Association Between Metacognition and Maths Performance According to Measure of Metacognition

From the 16 studies that used online measures (including confidence judgments, JOLs and calibration scores), 13 reported significant correlations, two reported mixed findings and one reported a non-significant correlation. The study that measured MC prospectively to the maths task reported a significant correlation (21) and the study that measured MC retrospectively to the maths task reported a non-significant correlation (9). Eleven (/12) studies that measured MC during a maths task reported significant correlations (2, 4, 10, 14, 15, 17, 18, 23, 26-28) and one reported mixed findings (24). One (/2) study that used an online measure that included both prospective and retrospective elements identified a significant correlation (19), and one reported mixed results (25).

Nine of the 17 studies that used offline questionnaires reported significant positive correlations between MC and maths performance (1-3, 5, 7, 12, 13, 17, 30), six reported non-significant correlations (6, 8, 11, 16, 29, 31) and two reported mixed findings (20, 22). Of the five studies that either included a maths-specific questionnaire or a questionnaire that had been adapted to be maths-specific, two reported a significant correlation (3, 17), two reported a non-significant correlation (8, 31), and one reported mixed findings (22).

1.3.1.5.2 The Association Between Metacognition and Maths Performance According to Measure of Maths Performance

Of the 11 studies that measured maths performance using word problems, nine reported that MC and maths performance were significantly positively related (2, 4, 10, 17, 18, 23, 26-28). The other two studies reported mixed, although mostly positive, correlations (24, 25). Of the ten studies that used standardised assessment to measure maths performance, seven produced non-significant correlations (6, 9, 11, 15, 20, 28, 30) and three produced significant correlations (3, 5, 29). Of the eight studies that used school assessment, five reported significant positive correlations (1, 13, 21, 25, 27), one reported a non- significant correlation (31) and two reported mixed findings (20, 22). Two studies used a series of mental maths questions from past examinations or revision workbooks (14) or performance on a university-facilitated maths course (31). Both studies reported non- significant correlations between MC and maths performance.

As aforementioned, nine of the 11 studies that used word problems (requiring complex problem-solving) reported significant correlations. Of the four studies that required less complex mathematic skill (e.g., simple recall or calculation), three reported a non-significant association between MC and maths performance (11, 14, 29) and one reported a significant correlation (30).

1.3.1.6 Quality Assessment

There were nine questions within the adapted CASP Cohort Study checklist (2018) that required a response of *yes*, *no* or *can't tell*, with a *yes* response indicative of higher-quality research. In seven studies, we rated all nine questions as *yes*, six studies were given eight *yes* responses, ten were given seven *yes* responses, five were given six, two were given five, and one study was awarded two *yes* responses only (see Appendix F for responses for each item).

All studies addressed a focused issue which was evidenced by clear research aims and rationales, and all were considered to be sufficiently precise. Correlations/associations were reported with at least 95% confidence ($p < .05$), and in most cases, with greater confidence (e.g., 99%, $p < .01$). Generally, participants were recruited in a way that meant they were likely to be representative of their cohort. However, in some studies (n = 7), it was not reported how participants were recruited (2, 4, 7, 12, 18, 19, 26).

Most MC measures were considered to be valid and reliable. Most studies (n = 18) used pre-published and validated questionnaires. One study used a measure designed by the researcher for the study (9). In this case, the measure had high inter-rater reliability, but there was no reference to validity testing. Most studies (n = 25) used acceptable measures of maths performance. In two studies, the measure of maths performance was unclear. For example, one study did not report how maths performance was measured (12), and another gave vague information (22). Additionally, one study reported coefficient alphas that fell within the poor/questionable range (αs < .70), one study used a measure on a single topic within mathematics that was felt to be particularly difficult for teachers to teach (7) and in one study, the questions were selected on the basis that males had previously out-performed females on those questions (9).

The impact of confounding factors on the relationship between MC and maths performance was not consistently considered. In six studies, factors that previous studies have linked to MC and/or maths performance were considered, such as gender, socioeconomic status and intelligence, but were not incorporated into the study's design or analysis. Three studies (23, 27, 28) reported semi-partial correlations between MC and maths performance to control for the contribution of general intelligence.

Most studies included participants who were considered to be relevant to the population of interest in the current review (11-16-year-olds). It was unclear whether some studies that selected participants based on specific characteristics, e.g., learning disability, having made below- expected progress in maths, having below or above average anxiety, or being academically talented, were representative of "typical" 11-16-year-olds. Furthermore, one study included only female participants.

1.3.2 Meta-Analysis Results

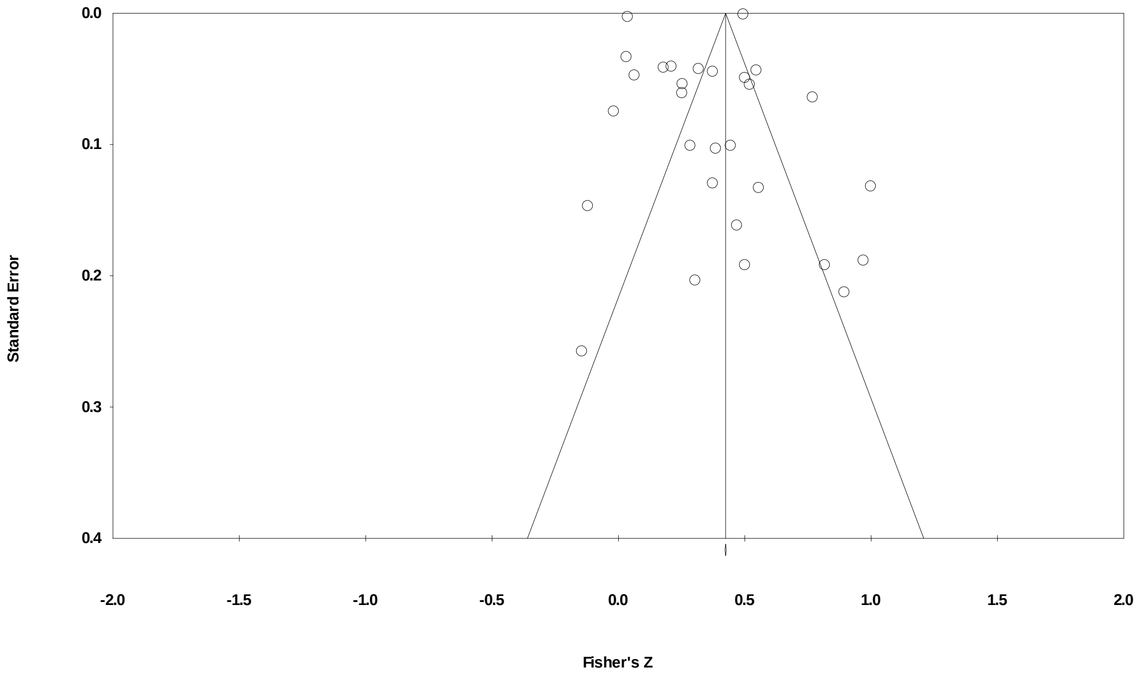
Correlations between MC and maths performance were available for 29 (/31) studies, and these were included in the meta-analysis (participant N = 570,575).

1.3.2.1 Publication Bias

Publication bias was assessed using Egger's test (Egger et al., 1997) and a visual inspection of a funnel plot created in CMA. The funnel plot showed a reasonable level of symmetry (see Figure 2), and the result of Egger's test showed no evidence for significant publication bias ($t(27) = .75$, $p = .23$). This was supported by the results of Beggs's test (Begg & Mazumdar, 1994; Kendall's tau = .26, $p = .02$).

Figure 2

A Funnel Plot to Inspect Publication Bias.



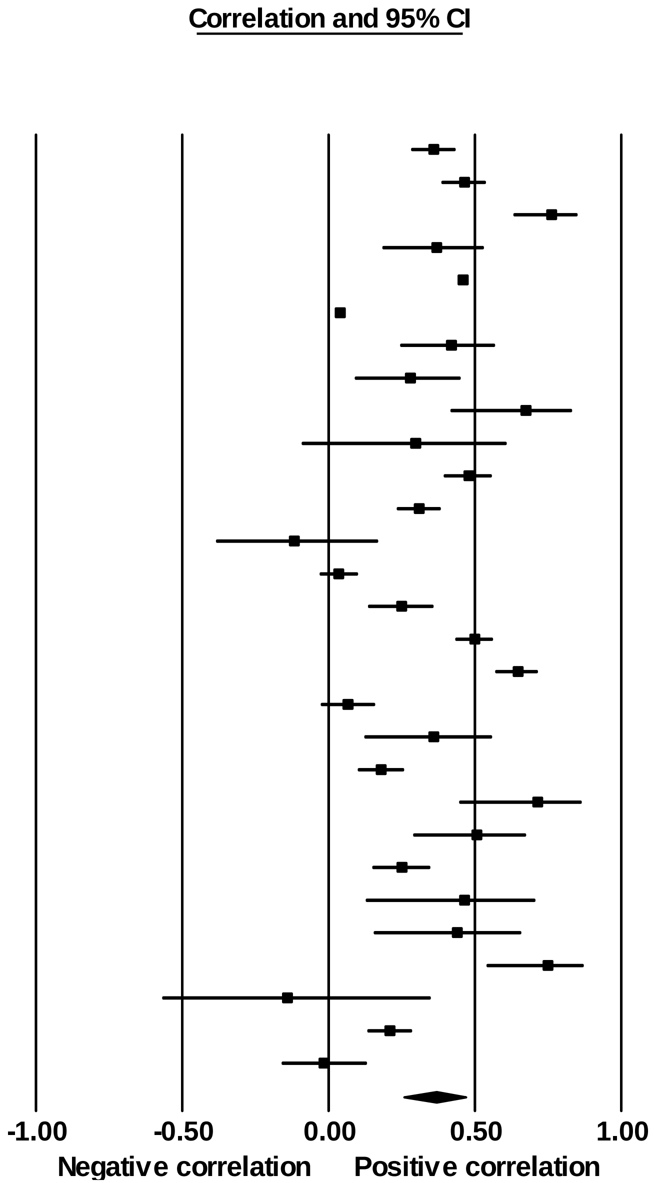
Note. Each plotted point represents the effect size from each study. Where studies reported more than one effect size (n = 13), the average effect size is represented. The Fisher's z effect size is plotted on the x-axis, and the standard error for each effect size is plotted on the y-axis

1.3.2.2 Primary Analysis

The primary analysis revealed a positive and significant association between MC and maths performance ($r = .37$, 95% CI = [.26, .47], Z = 6.15, $p < .001$; see Figure 3). Sensitivity analysis showed that removing any individual study did not significantly change the results (all rs ≥ .34, all ps < .001). There were no studentised residuals ± 1.96, suggesting that no individual study was a significant outlier. The primary analysis using RVE also indicated that the association between MC and maths performance was significantly positive ($r = .35$, $p < .001$). For this analysis, ρ was set to 0.8; however, sensitivity analysis indicated that the findings were robust across other reasonable estimates of ρ.

Figure 3. *A Forest Plot of the Effect Sizes for Each Study and the Overall Effect Size, as Calculated in CMA. The Boxes Represent the Effect Size (r) for Each Study, the Lines Represent 95% Confidence Intervals, and the Diamond Represents the Synthesised Effect Size.*

Study name	Statistics for each study				
	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Ahmed et al. (2013)	0.360	0.281	0.434	8.360	0.000
Asik & Erktin (2019)	0.465	0.385	0.538	10.111	0.000
Bishara, & Kaplan (2018)	0.763	0.631	0.852	7.566	0.000
Callan & Cleary (2019)	0.370	0.183	0.531	3.746	0.000
Callan et al. (2016)	0.460	0.458	0.462	342.913	0.000
Chiu et al. (2007)	0.040	0.033	0.047	11.912	0.000
Erktin (2004)	0.420	0.244	0.569	4.409	0.000
Fitzpatrick (1994)	0.280	0.088	0.452	2.833	0.005
Fusco (1995)	0.675	0.416	0.833	4.260	0.000
Harris (2015)	0.298	-0.092	0.609	1.506	0.132
Hassan & Rahman (2017)	0.480	0.393	0.559	9.500	0.000
Ichihara & Arai (2006)	0.310	0.232	0.384	7.449	0.000
Maras et al. (2019)	-0.117	-0.386	0.170	-0.797	0.425
Ning (2016)	0.035	-0.031	0.101	1.033	0.302
Özcan (2016)	0.250	0.134	0.359	4.158	0.000
Özcan & Gümüs (2019)	0.500	0.432	0.562	12.454	0.000
Ozsoy (2011)	0.648	0.568	0.716	11.932	0.000
Peng et al. (2014)	0.067	-0.027	0.159	1.393	0.164
Sink et al. (1991)	0.360	0.121	0.559	2.895	0.004
Tian et al. (2018)	0.180	0.099	0.258	4.330	0.000
van der Stel & Veenman (2014)	0.715	0.446	0.866	4.209	0.000
van der Stel et al. (2010)	0.507	0.288	0.675	4.181	0.000
van der Walt et al. (2008)	0.251	0.149	0.348	4.706	0.000
Veenman et al. (2000)	0.465	0.126	0.707	2.617	0.009
Veenman et al. (2005)	0.440	0.153	0.659	2.911	0.004
Veenman & Spaans (2005)	0.750	0.539	0.872	5.148	0.000
Walker (2013)	-0.140	-0.570	0.350	-0.546	0.585
Yap (1993)	0.210	0.132	0.286	5.169	0.000
Young & Worrell (2018)	-0.015	-0.161	0.132	-0.199	0.842
	0.365	0.255	0.466	6.146	0.000



1.3.2.3 Moderator Analysis

1.3.2.3.1 Online vs. Offline Measurement of Metacognition

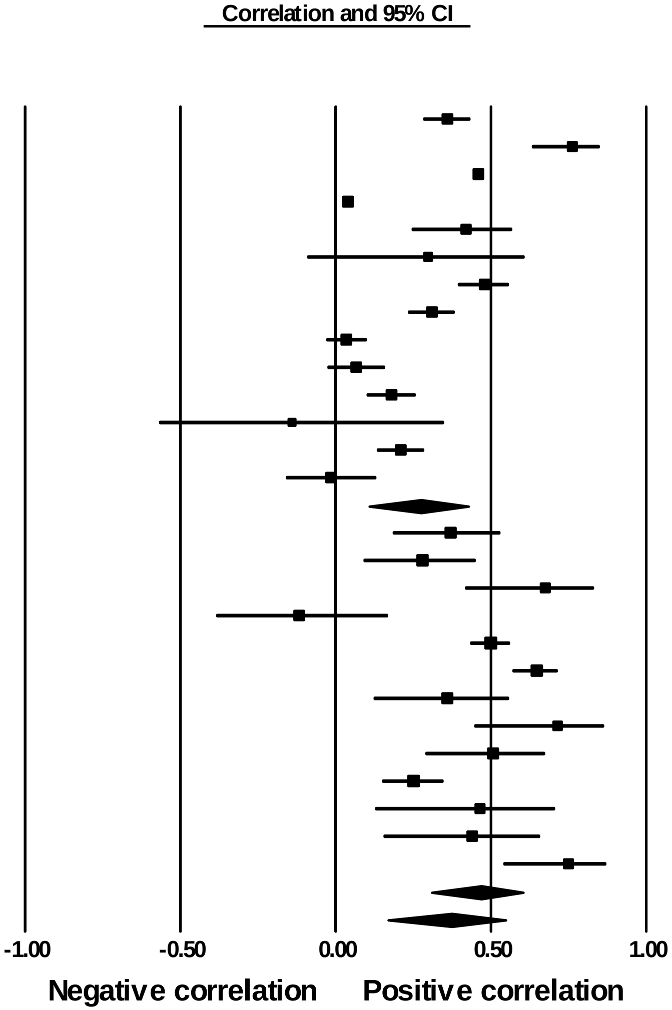
Subgroup analysis explored whether measure of MC (online vs. offline) moderated the relationship between MC and maths performance. To consider the different distinctions in the literature between online and offline measures, this analysis was carried out four times. In all four analyses self-report questionnaires were classed as offline and think-aloud protocols and behaviour observations were classed as online, however JOLs, confidence judgments and calibration scores were classed differently between analyses. The four analyses carried out were: (1) online defined as pertaining to a specific task at hand (i.e., JOLs, confidence judgments and calibration scores classed as online), (2) online defined as not self-reported (JOLs, confidence judgments and calibration scores classed as offline), (3) online defined as taking place during a task (JOLs, confidence judgments and calibration scores classed as online where they were carried out during a task, those carried out before/after a task classed as offline) and (4) JOLs, confidence judgments and calibration scores (i.e., student-reported MC score relevant to a specific task at hand) categorised separately from other online and offline measures.

(1) When JOLs, confidence judgments and calibration scores were classed as online (JOLs, confidence judgments, calibration scores, think-aloud protocols and observations vs. offline questionnaires), the effect size was descriptively (but not significantly) larger in studies that used online measures (n = 13, $r = .47$, 95% CI = [.31, .60]), than in studies that used offline measures (n = 14, $r = .27$, 95% CI = [.11, .42]; $Q(1) = 2.98$, $p = .08$; see Figure 4). (2) When JOLs, confidence judgments and calibration scores were classed as offline (think-aloud protocols and observations vs. offline questionnaires, JOLs, confidence judgments and calibration scores), the effect size was significantly larger for studies that used online measures (n = 6, $r = .60$, 95% CI = [.38, 76]) than those that used offline measures (n = 21, $r = .30$, 95% CI = [.17 .42]; $Q(1) = 5.46$, $p = .02$; see Figure 5). (3) When JOLs, confidence judgments and calibration scores that were completed before or after each individual item during a longer maths task were classed as online, and those that were completed before or after an entire maths task were classed as offline, the difference between online and offline studies approached significance, $Q(1) = 3.57$, $p = .06$ (see Figure 6); the effect sizes were descriptively larger for studies that measured MC during a task (online; n = 11, $r = .49$, 95% CI = [.32, .63]) than studies that measured MC not during a task (offline; n = 16, $r = .28$, 95% CI = [.12, .42]). (4) Finally, when JOLs, confidence judgments and

calibration scores were coded separately to other online and offline measures, the difference in effect size between groups (online vs. offline vs. JOLs, confidence judgments and calibration scores) approached significance (*Q*(2) = 5.77, *p* = .056; see Figure 7); effect sizes were descriptively largest in studies that used online measures (n = 6, *r* = .60, 95% CI = [.38, .76]), followed by studies that used JOLs, confidence judgments or calibration (n = 7, *r* = .35, 95% CI = [.12, .55]) and those that used offline measures (n = 14, *r* = .27, 95% CI = [.11, .42]).

Figure 4. *A Forest Plot to Show the Effect Sizes for Studies Included in the Online Vs. Offline Subgroup Analysis When Online Measures are Defined as Those That Pertain to a Specific Task at Hand and Offline Measures do not Relate to a Specific Task at Hand (Self-Report Questionnaires)*

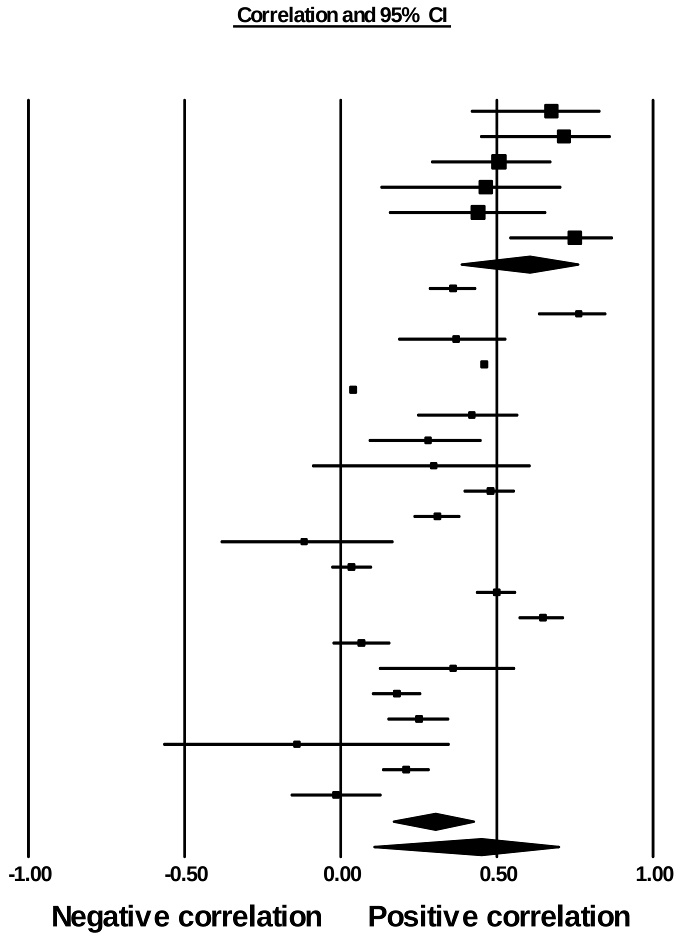
Group by Online vs. offline	Studyname	Statistics for each study				
		Correlation	Lower limit	Upper limit	Z-Value	p-Value
offline	Ahmed et al. (2013)	0.360	0.281	0.434	8.360	0.000
offline	Bishara, & Kaplan (2018)	0.763	0.631	0.852	7.566	0.000
offline	Callan et al. (2016)	0.460	0.458	0.462	342.913	0.000
offline	Chiu et al. (2007)	0.040	0.033	0.047	11.912	0.000
offline	Erktin (2004)	0.420	0.244	0.569	4.409	0.000
offline	Harris (2015)	0.298	-0.092	0.609	1.506	0.132
offline	Hassan & Rahman (2017)	0.480	0.393	0.559	9.500	0.000
offline	Ichihara & Arai (2006)	0.310	0.232	0.384	7.449	0.000
offline	Ning (2016)	0.035	-0.031	0.101	1.033	0.302
offline	Peng et al. (2014)	0.067	-0.027	0.159	1.393	0.164
offline	Tian et al. (2018)	0.180	0.099	0.258	4.330	0.000
offline	Walker (2013)	-0.140	-0.570	0.350	-0.546	0.585
offline	Yap (1993)	0.210	0.132	0.286	5.169	0.000
offline	Young & Worrell (2018)	-0.015	-0.161	0.132	-0.199	0.842
offline		0.272	0.105	0.424	3.150	0.002
online	Callan & Cleary (2019)	0.370	0.183	0.531	3.746	0.000
online	Fitzpatrick (1994)	0.280	0.088	0.452	2.833	0.005
online	Fusco (1995)	0.675	0.416	0.833	4.260	0.000
online	Maras et al. (2019)	-0.117	-0.386	0.170	-0.797	0.425
online	Özcan & Gümüs (2019)	0.500	0.432	0.562	12.454	0.000
online	Ozsoy (2011)	0.648	0.568	0.716	11.932	0.000
online	Sink et al. (1991)	0.360	0.121	0.559	2.895	0.004
online	van der Stel & Veenman (2014)	0.715	0.446	0.866	4.209	0.000
online	van der Stel et al. (2010)	0.507	0.288	0.675	4.181	0.000
online	van der Walt et al. (2008)	0.251	0.149	0.348	4.706	0.000
online	Veenman et al. (2000)	0.465	0.126	0.707	2.617	0.009
online	Veenman et al. (2005)	0.440	0.153	0.659	2.911	0.004
online	Veenman & Spaans (2005)	0.750	0.539	0.872	5.148	0.000
online		0.466	0.306	0.600	5.250	0.000
Overall		0.370	0.166	0.544	3.447	0.001



Note. The boxes and lines represent the effect size (r) and 95% confidence intervals for each study, and the diamonds represent the synthesised effect sizes for each subgroup (offline, online) and overall

Figure 5. *A Forest Plot to Show the Effect Sizes for Studies Included in the Online Vs. Offline Subgroup Analysis When Online Measures are Defined as Those That are not Self-Reported (Think-Aloud Protocols and Behavioural Observations) and Offline Measures are Those That are Self-Reported.*

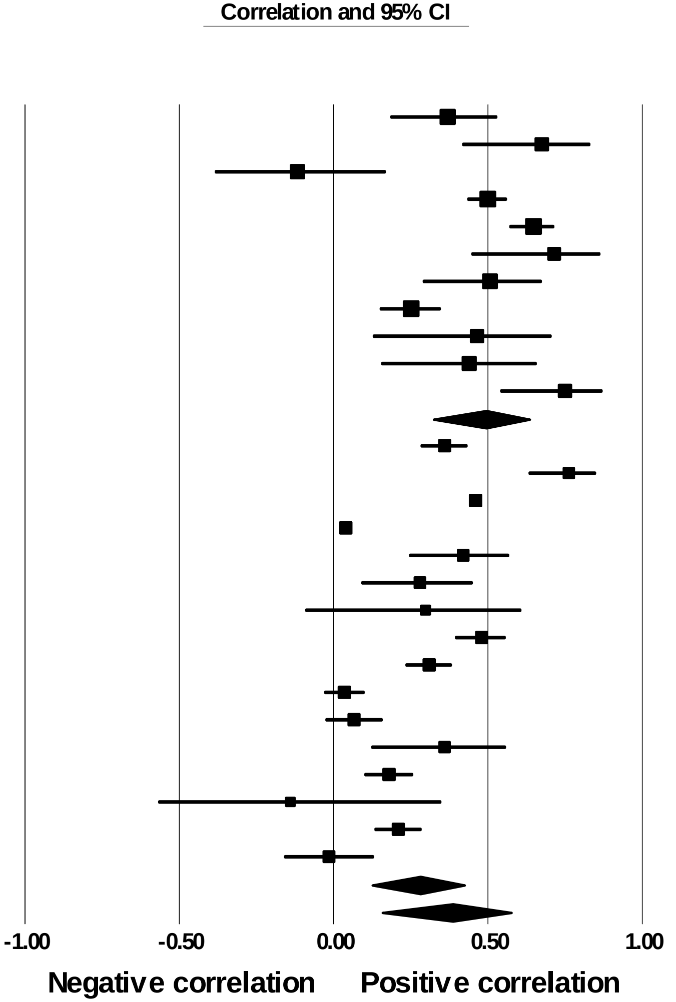
Group by Self-reported vs. not self-reported	Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
not	Fusco (1995)	0.675	0.416	0.833	4.260	0.000
not	van der Stel & Veenman (2014)	0.715	0.446	0.866	4.209	0.000
not	van der Stel et al. (2010)	0.507	0.288	0.675	4.181	0.000
not	Veenman et al. (2000)	0.465	0.126	0.707	2.617	0.009
not	Veenman et al. (2005)	0.440	0.153	0.659	2.911	0.004
not	Veenman & Spaans (2005)	0.750	0.539	0.872	5.148	0.000
not		0.602	0.383	0.756	4.673	0.000
self-reported	Ahmed et al. (2013)	0.360	0.281	0.434	8.360	0.000
self-reported	Bishara, & Kaplan (2018)	0.763	0.631	0.852	7.566	0.000
self-reported	Callan & Cleary (2019)	0.370	0.183	0.531	3.746	0.000
self-reported	Callan et al. (2016)	0.460	0.458	0.462	342.913	0.000
self-reported	Chiu et al. (2007)	0.040	0.033	0.047	11.912	0.000
self-reported	Erktin (2004)	0.420	0.244	0.569	4.409	0.000
self-reported	Fitzpatrick (1994)	0.280	0.088	0.452	2.833	0.005
self-reported	Harris (2015)	0.298	-0.092	0.609	1.506	0.132
self-reported	Hassan & Rahman (2017)	0.480	0.393	0.559	9.500	0.000
self-reported	Ichihara & Arai (2006)	0.310	0.232	0.384	7.449	0.000
self-reported	Maras et al. (2019)	-0.117	-0.386	0.170	-0.797	0.425
self-reported	Ning (2016)	0.035	-0.031	0.101	1.033	0.302
self-reported	Özcan & Gümüs (2019)	0.500	0.432	0.562	12.454	0.000
self-reported	Ozsoy (2011)	0.648	0.568	0.716	11.932	0.000
self-reported	Peng et al. (2014)	0.067	-0.027	0.159	1.393	0.164
self-reported	Sink et al. (1991)	0.360	0.121	0.559	2.895	0.004
self-reported	Tian et al. (2018)	0.180	0.099	0.258	4.330	0.000
self-reported	van der Walt et al. (2008)	0.251	0.149	0.348	4.706	0.000
self-reported	Walker (2013)	-0.140	-0.570	0.350	-0.546	0.585
self-reported	Yap (1993)	0.210	0.132	0.286	5.169	0.000
self-reported	Young & Worrell (2018)	-0.015	-0.161	0.132	-0.199	0.842
self-reported		0.300	0.166	0.422	4.286	0.000
Overall		0.447	0.104	0.695	2.503	0.012



Note. The boxes and lines represent the effect size (r) and 95% confidence intervals for each Study, and the diamonds represent the synthesised effect sizes for each subgroup (not self-reported, self-reported) and overall

Figure 6. *A Forest Plot to Show the Effect Sizes for Studies Included in the Online Vs. Offline Subgroup Analysis When Online Measures are Defined as Those That are Completed During the Maths Task and Offline Measures are Those That are not Completed During the Assessment Task.*

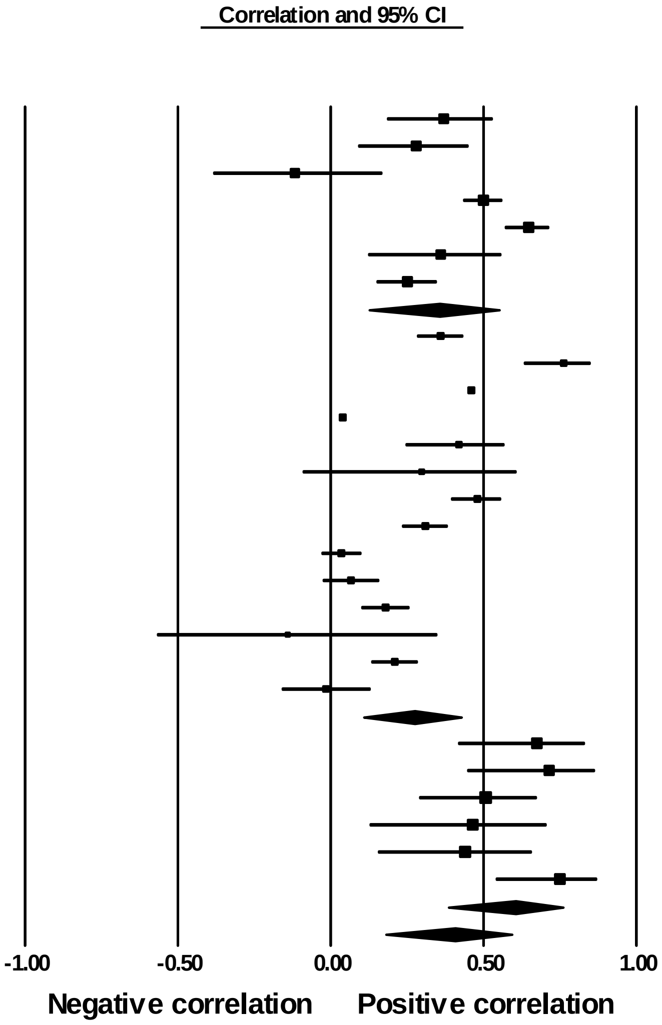
Group by When MC measured	Studyname	Statistics for each study Correlation	Lower limit	Upper limit	Z-Value	p-Value
during	Callan & Cleary (2019)	0.370	0.183	0.531	3.746	0.000
during	Fusco (1995)	0.675	0.416	0.833	4.260	0.000
during	Maras et al. (2019)	-0.117	-0.386	0.170	-0.797	0.425
during	Özcan & Gümüs (2019)	0.500	0.432	0.562	12.454	0.000
during	Ozsoy (2011)	0.648	0.568	0.716	11.932	0.000
during	van der Stel & Veenman (2014)	0.715	0.446	0.866	4.209	0.000
during	van der Stel et al. (2010)	0.507	0.288	0.675	4.181	0.000
during	van der Walt et al. (2008)	0.251	0.149	0.348	4.706	0.000
during	Veenman et al. (2000)	0.465	0.126	0.707	2.617	0.009
during	Veenman et al. (2005)	0.440	0.153	0.659	2.911	0.004
during	Veenman & Spaans (2005)	0.750	0.539	0.872	5.148	0.000
during		0.492	0.321	0.632	5.128	0.000
not during	Ahmed et al. (2013)	0.360	0.281	0.434	8.360	0.000
not during	Bishara, & Kaplan (2018)	0.763	0.631	0.852	7.566	0.000
not during	Callan et al. (2016)	0.460	0.458	0.462	342.913	0.000
not during	Chiu et al. (2007)	0.040	0.033	0.047	11.912	0.000
not during	Erktin (2004)	0.420	0.244	0.569	4.409	0.000
not during	Fitzpatrick (1994)	0.280	0.088	0.452	2.833	0.005
not during	Harris (2015)	0.298	-0.092	0.609	1.506	0.132
not during	Hassan & Rahman (2017)	0.480	0.393	0.559	9.500	0.000
not during	Ichihara & Arai (2006)	0.310	0.232	0.384	7.449	0.000
not during	Ning (2016)	0.035	-0.031	0.101	1.033	0.302
not during	Peng et al. (2014)	0.067	-0.027	0.159	1.393	0.164
not during	Sink et al. (1991)	0.360	0.121	0.559	2.895	0.004
not during	Tian et al. (2018)	0.180	0.099	0.258	4.330	0.000
not during	Walker (2013)	-0.140	-0.570	0.350	-0.546	0.585
not during	Yap (1993)	0.210	0.132	0.286	5.169	0.000
not during	Young & Worrell (2018)	-0.015	-0.161	0.132	-0.199	0.842
not during		0.278	0.122	0.420	3.435	0.001
Overall		0.383	0.155	0.572	3.199	0.001



Note. The boxes and lines represent the effect size (r) and 95% confidence intervals for each study, and the diamonds represent the synthesised effect sizes for each subgroup (during maths task, not during maths task) and overall

Figure 7. *A Forest Plot to Show the Effect Sizes for Studies That Measured Metacognition Using Self-Reported MC Scores Relevant to a Task at Hand (JOLs, Confidence Judgments, Calibration Scores), Offline Questionnaires, and Online Think-Aloud Protocols and/or Behaviour Observations*

Group by Online vs. offline vs. JOL	Studyname	Statistics for each study: Correlation	Lower limit	Upper limit	Z-Value	p-Value
JOL	Callan & Cleary (2019)	0.370	0.183	0.531	3.746	0.000
JOL	Fitzpatrick (1994)	0.280	0.088	0.452	2.833	0.005
JOL	Maras et al. (2019)	-0.117	-0.386	0.170	-0.797	0.425
JOL	Özcan & Gümüs (2019)	0.500	0.432	0.562	12.454	0.000
JOL	Ozsoy (2011)	0.648	0.568	0.716	11.932	0.000
JOL	Sink et al. (1991)	0.360	0.121	0.559	2.895	0.004
JOL	van der Walt et al. (2008)	0.251	0.149	0.348	4.706	0.000
JOL		0.354	0.123	0.548	2.941	0.003
offline	Ahmed et al. (2013)	0.360	0.281	0.434	8.360	0.000
offline	Bishara, & Kaplan (2018)	0.763	0.631	0.852	7.566	0.000
offline	Callan et al. (2016)	0.460	0.458	0.462	342.913	0.000
offline	Chiu et al. (2007)	0.040	0.033	0.047	11.912	0.000
offline	Erktin (2004)	0.420	0.244	0.569	4.409	0.000
offline	Harris (2015)	0.298	-0.092	0.609	1.506	0.132
offline	Hassan & Rahman (2017)	0.480	0.393	0.559	9.500	0.000
offline	Ichihara & Arai (2006)	0.310	0.232	0.384	7.449	0.000
offline	Ning (2016)	0.035	-0.031	0.101	1.033	0.302
offline	Peng et al. (2014)	0.067	-0.027	0.159	1.393	0.164
offline	Tian et al. (2018)	0.180	0.099	0.258	4.330	0.000
offline	Walker (2013)	-0.140	-0.570	0.350	-0.546	0.585
offline	Yap (1993)	0.210	0.132	0.286	5.169	0.000
offline	Young & Worrell (2018)	-0.015	-0.161	0.132	-0.199	0.842
offline		0.272	0.105	0.424	3.149	0.002
online	Fusco (1995)	0.675	0.416	0.833	4.260	0.000
online	van der Stel & Veenman (2014)	0.715	0.446	0.866	4.209	0.000
online	van der Stel et al. (2010)	0.507	0.288	0.675	4.181	0.000
online	Veenman et al. (2000)	0.465	0.126	0.707	2.617	0.009
online	Veenman et al. (2005)	0.440	0.153	0.659	2.911	0.004
online	Veenman & Spaans (2005)	0.750	0.539	0.872	5.148	0.000
online		0.602	0.383	0.757	4.659	0.000
Overall		0.404	0.178	0.590	3.372	0.001



Note. The boxes and lines represent the effect size (r) and 95% confidence intervals for each study, and the diamonds represent the synthesised effect sizes for each subgroup and overall. In the figure, "JOL" represents not just Judgment of Learning scores, but also confidence judgments and calibration scores.

1.3.2.3.2 Simple vs. Complex Maths Performance

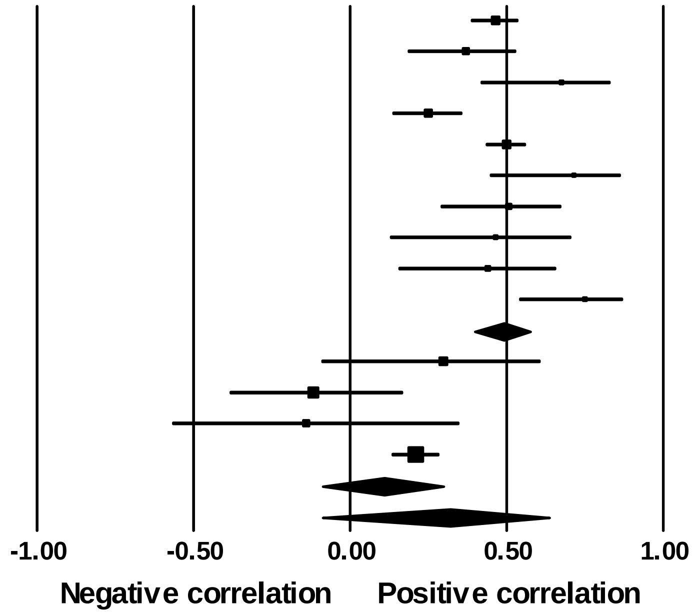
Considering measure of maths performance, descriptively, effect sizes were highest in studies that used researcher-designed tests (n = 2, *r* = .55, 95% CI = [.15,.79]), followed by word- problems (n = 10, *r* = .52, 95% CI = [.34, .66]), studies that did not specify how maths performance was measured (n = 2, *r* = .34, 95% CI = [-..10, .66]), standardised assessment (n = 8, *r* = .28, 95% CI = [.05, .48]), school assessment and a geometry problem (n = 1, *r* = .25, 95% CI = [-.36, .71]), school assessment (n = 3, *r* = .25, 95% CI = [-.11, .55]), school and standardised assessment (n = 2, *r* = .17, 95% CI = [-.29, .56]) and mental-maths questions (n = 1, *r* = -.12, 95% CI = [-.67, .52]). There was not a significant difference between these groups ($Q(7) = 7.98$, $p = 3.4$).

However, the effect size was significantly larger in studies that required participants to use complex problem-solving skills during assessment of maths performance (n = 10, *r* = .49, 95% CI = [.40, .57]) than simple recall and learnt algorithms (n = 4, *r* = .11, 95% CI = [-.09, .30]; $Q(1) = 13.35$, $p < .001$; see Figure 8).

Figure 8. *A Forest Plot to Show the Effect Sizes for Studies Included in the Subgroup Analysis to Compare Studies Which Required Complex Versus Simple Mathematical Skill*

Group by Maths task	Study name	Statistics for each study				
		Correlation	Lower limit	Upper limit	Z-Value	p-Value
complex	Asik & Erktin (2019)	0.465	0.385	0.538	10.111	0.000
complex	Callan & Cleary (2019)	0.370	0.183	0.531	3.746	0.000
complex	Fusco (1995)	0.675	0.416	0.833	4.260	0.000
complex	Özcan (2016)	0.250	0.134	0.359	4.158	0.000
complex	Özcan & Gümüs (2019)	0.500	0.432	0.562	12.454	0.000
complex	van der Stel & Veenman (2014)	0.715	0.446	0.866	4.209	0.000
complex	van der Stel et al. (2010)	0.507	0.288	0.675	4.181	0.000
complex	Veenman et al. (2000)	0.465	0.126	0.707	2.617	0.009
complex	Veenman et al. (2005)	0.440	0.153	0.659	2.911	0.004
complex	Veenman & Spaans (2005)	0.750	0.539	0.872	5.148	0.000
complex		0.489	0.395	0.572	8.971	0.000
simple	Harris (2015)	0.298	-0.092	0.609	1.506	0.132
simple	Maras et al. (2019)	-0.117	-0.386	0.170	-0.797	0.425
simple	Walker (2013)	-0.140	-0.570	0.350	-0.546	0.585
simple	Yap (1993)	0.210	0.132	0.286	5.169	0.000
simple		0.106	-0.091	0.295	1.051	0.293
Overall		0.317	-0.091	0.634	1.533	0.125

Correlation and 95% CI



Note. The boxes and lines represent the effect size (r) and 95% confidence intervals for each study, and the diamonds represent the synthesised effect sizes for each subgroup (complex, simple) and overall

1.3.2.4 Heterogeneity

Heterogeneity was significant and very high in the primary analysis (*Q*(28) = 16022.16, $p < .001$, $I^2 = 99.83$), indicating significant variation across included studies (Higgins et al., 2003). Subgroup analyses were appropriate to explore this further. However, the variance in effect sizes was not accounted for by moderators (all *Q*s > 15904, all *p*s< .001). Heterogeneity was also significant within most subgroups analysed, with two exceptions; when JOLs, confidence judgments and calibration scores were classed as an offline measure, although the heterogeneity in offline measures was significant and very high ($I^2 = 99.99$), the dispersion in studies that used online measures was not significant (*Q*(5)= 7.24, $p = .20$, $I^2 = 30.92$). Heterogeneity was also not significant in studies that used maths assessment tasks which required simple recall/learnt algorithms (*Q*(3) = 6.63, $p = .09$, $I^2 = 54.74$).

1.3.2.4.1 Power Analysis

To further explore heterogeneity, a power analysis was conducted to investigate whether statistical power contributed to the results. Power analysis was undertaken using G*Power software (Faul et al., 2007) using the synthesised effect size in CMA ($r = .37$). The findings indicated that a sample of 89 participants is necessary to detect significance with 95% power, and 55 participants to detect significance with 80% power ($\alpha = .05$, two-tailed). Based on this, eight of the 29 included studies (10, 11, 14, 23, 26-29) did not have sufficient statistical power (at 80%). However, heterogeneity remained significant when these eight studies were removed from the analysis (*Q*(20) = 15985.70, $p < .001$; $I^2 = 99.88$). In this analysis, the overall correlation decreased slightly, although it was still significantly positive ($r = .34$, 95% CI = [.22, .46], Z = 5.06, $p < .001$).

1.4 Discussion

The current systematic review investigated the association between MC and maths performance in CYP aged 11-16. In addition, we conducted a meta-analysis to test the hypothesis that there is a significant positive association between MC and maths performance in 11-16-year-olds across studies, and to consider whether measurements of MC (online vs. offline) and maths performance (simple vs. complex) moderated this association. A systematic search for published and unpublished records yielded 31 relevant studies. The synthesis of 82 effect sizes from 29 of these studies (N = 570,575) indicated a significantly positive, medium-sized correlation between MC and maths performance (r = .37, p < .001). This indicates that CYP who reported or demonstrated better MC also performed better in maths tasks. These findings fit with previous systematic reviews/meta-analyses that have identified a relationship between MC and academic performance (Dent & Koenka, 2016; Ohtani & Hisasaka, 2018), and student engagement in MC interventions and maths performance (Dignath & Büttner, 2008; Perry et al., 2019; Walker, 2013).

Online and offline measures of MC are defined differently in the literature. To try to address this in the current review, subgroup analyses were undertaken according to three popular distinctions: JOLs, confidence judgments and calibration scores classed as (1) online, (2) offline, and (3) those completed during a task as online and those completed before or after a task as offline. We also ran this analysis a fourth time where JOLs, confidence judgments and calibration scores (self-reported MC score relevant to a task at hand) were categorised separately from other online and offline measures. The difference in effect sizes between online and offline measures was significant when JOLs, confidence judgments and calibration scores were categorised as offline (p = .02). This indicates that the association between MC and maths performance was significantly greater when MC was measured using online think-aloud protocols and/or behavioural observations (r = .60) rather than self-report measures (offline questionnaires or JOLs, confidence judgments or calibration scores; r = .30). The difference between groups also approached significance when JOLs, confidence judgments and calibration scores that were completed during a task were classed as online alongside think-aloud protocols and observations (r = .49), and those that were completed before or after an entire task (rather than individual items) were classed as offline measures alongside self-report questionnaires (r = .28, p = .06). When JOLs, confidence judgments and calibration scores were categorised separately from online measures (think-aloud protocols and behaviour observations) and offline measures (self-report questionnaires), the difference between subgroups (online, offline, JOLs) approached significance (*p* = .056); studies that used online measures reported the largest effect sizes (*r* = .60), followed by those that used JOLs, confidence judgments or calibration scores (*r* = .35) and those that used offline questionnaires (*r* = .27).

The current findings are consistent with previous studies that have identified a greater correlation between maths performance and MC when MC is measured using online think-aloud protocols and/or behaviour observations, than self-reported measures. For example, Veenman and van Cleef (2019) asked 30 adolescents to complete five MC instruments before, during or after completing maths word problems. Student scores on two offline questionnaires (the MSLQ and inventory of learning styles; ILS, Vermunt & van Rijswijk, 1987) completed before the maths task were not correlated with maths performance (as measured by post-test and GPA). JOL following the maths task was significantly correlated with the post-test score but not with GPA. On the other hand, there were significant positive correlations between online think-aloud protocols and observations during the maths tasks and both measures of maths performance. The current findings also fit with meta-analyses (Dent & Koenka, 2016; Ohtani & Hisaka, 2018) that have also identified larger effect sizes for online (vs. offline) measures of MC and academic performance.

The difference in the strength of association between online and offline MC measures with maths performance may be explained by the idea that, as some researchers have previously suggested (e.g., Saraç & Karakelle, 2012), online and offline measures capture different aspects of MC. In the current review, think-aloud protocols and behaviour observations (online measures) primarily measured MC control and JOLs, confidence judgments and calibration scores primarily measured MC monitoring. Most self-report questionnaires (n = 6) asked participants questions relating to both MC monitoring and control.

As discussed, whether self-reported MC scores relevant to a specific task (including JOLs, confidence judgments and calibration scores) should be classed as online or offline has been debated in the literature. In a previous meta-analysis, Ohtani and Hisaka (2018) identified that maths performance was most associated with online think-aloud protocols and/or observations, followed by self-reported MC relevant to a specific task, and then offline self-report questionnaires. This order was replicated in the current study. Moreover, as the differences between online, offline and self-reported MC relevant to a specific task (JOLs, confidence judgments, calibration scores) approached significance (*p* = .056), it could be considered that JOLs, confidence judgments and calibration scores are not well-placed in either category and the online/offline division may be too simplistic.

Further subgroup analysis identified a significantly greater correlation between MC and maths performance when the measure of maths performance required complex (vs. simple) mathematical skill. This suggests that MC may be more important for complex problem-solving and may be less critical to generate solutions to simple questions. This fits with Verschaffel et al. (2010) who considered that complex problem-solving necessitates MC. However, it should be acknowledged that, of the four studies that used measures of maths performance which required "simple" mathematical skill, two of these used the Jr. MAI and a third used another offline self-report questionnaire. Therefore, using offline questionnaires may also account for, or partially account for, lower reported correlations between MC and maths performance.

1.4.1 Strengths and Limitations of the Current Review and Directions for Future Research

A strength of this review is the thoroughness of the systematic search. The search strategy was developed after consulting a researcher who has a specific interest in MC and has written several published papers in the field, in addition to a research librarian. To reduce the impact of publication bias, both published and unpublished research in all languages were included. Unpublished research and/or research that was not written in English were omitted from some of the previous reviews on this topic (Dignath & Büttner, 2008; Higgins et al., 2005; Perry et al., 2019; Walker, 2013). The breadth of literature reviewed in the current review incites confidence that all relevant records have been included, and that the conclusions drawn are based on the synthesis of all available, relevant evidence.

On the other hand, a fundamental limitation is the small number of studies in subgroup analyses and the unexplained high heterogeneity. This signifies the importance of further research to consider other factors that are recognised to influence achievement in school, which may moderate the relationship between MC and maths performance, such as intelligence and attention (Steinmayr et al., 2010) and anxiety (Moran, 2016). This would also allow more specific recommendations and would address a methodological weakness of studies regarding the absence of consideration and measurement of confounding variables.

In addition to the high heterogeneity across studies, heterogeneity was also significant within most subgroups, and in particular, was extremely high for offline measures. This is likely reflective of the wide range of questionnaires that are used to assess MC; 13 different self-report questionnaires were used in included studies. Moreover, studies that used the same instrument reached similar conclusions; all of the studies (4 of 4) that used the Jr MAI (Dennison et al., 1996) reported non-significant correlations and all of the studies (3 of 3) that used the MES (Efklides, 2006) identified significant correlations. This warrants further consideration regarding the validity and reliability of offline measures, particularly as the current review identified that these are becoming increasingly used over time. As the correlation between online (vs. offline) MC measures and maths performance was larger, it could be hypothesised that if this meta-analysis were repeated in the future, the synthesised effect size might decrease.

Another limitation of the current review was that, despite efforts to retrieve all relevant papers, four papers were unable to be accessed online either freely, using the databases available within the University or through an Inter-Library Loan. There was also a considerable drop-out of irrelevant papers from those identified in initial searches to those that met the inclusion criteria, indicating that the search criteria were possibly too broad. The main reasons for eliminating papers during full-paper screening were that participants were not within the target age range, followed by the absence of a reported association between MC and maths performance.

Moreover, although the methodological quality of most included papers was acceptable or good (as indicated by receiving affirmative responses in the CASP questionnaire), subgroup analysis based on maths performance measure revealed that the two studies that had used researcher-designed tests to measure maths performance reported the highest correlation with MC.

1.4.2 Conclusion and Practical Implications

In summary, the synthesised effect size is consistent with the notion that CYP aged 11-16 years who have better MC, also perform more highly on maths assessments. The association between MC and maths performance was greater in studies that used think-aloud protocols and behavioural observation to measure MC (vs. self-report measures), and also in studies in which the measure of maths performance required complex (vs. simple) mathematical skill.

Previous systematic reviews and meta-analyses have identified that engaging in a MC intervention is associated with increased mathematics performance in CYP. The current findings, taken alongside this previous literature, suggest that it is likely to be helpful to encourage MC in schools to support achievement. This is particularly important both given that under-achievement in maths is associated with a range of adverse outcomes, and that a considerable number of young people achieve below expectations in maths (e.g., at GCSE). As the association with maths performance is greater for online than for offline measures, it may be most helpful to support students to develop their MC skills during a task.

Chapter 2 A Pilot Study Exploring the Impact of a Successive Relearning Intervention for Post-16 GCSE Mathematics Re-Sit Students

2.1 Introduction

Adults with poor maths/numeracy skills are more likely to be in unskilled employment, to experience low self-esteem, and to report poorer physical health (Parsons & Bynner, 2005). Most students in the United Kingdom (UK) take GCSE (General Certificates of Secondary Education) qualifications aged 15-16 years. A standard pass grade in GCSE Maths is typically a prerequisite to access training courses and job opportunities, as well as to take further and higher education qualifications. The UK government have implemented several reforms to increase the number of students achieving pass grades in core subjects. These include raising the age of compulsory education to 18 years and requiring students to re-sit GCSE (or equivalent) Maths if they have not achieved pass grades (Education & Skills Funding Agency, 2019). However, in 2019, only 18.2% of the 142,488 16-18-year-olds who re-sat GCSE Maths achieved a pass grade by the end of further education aged 18 years (DfE, 2020). Moreover, compared with their original point score aged 15-16 years, 36.2% of students showed some improvement, 25.2% achieved the same point score, and 38.1% achieved a lower point score aged 16-18 years (DfE, 2020). Researchers have argued that the requirement to continue to re-sit GCSE Maths is harmful and demotivating to young people (e.g., Johnston-Wilder et al., 2015).

Educational professionals are increasingly considering the evidence-base behind teaching approaches and techniques to achieve the best outcomes for students (The Sutton Trust, 2018). The evidence for two techniques (distributed practice and retrieval practice) is particularly robust (Dunlosky et al., 2013). These strategies can be described as “desirable difficulties” (Bjork & Bjork, 2014; p. 59) because they require more effort than more passive techniques such as re-studying (i.e., re-reading material), but result in more durable learning (Bjork, 2017).

2.1.1 Distributed Practice

Distributed practice involves the spreading of learning across multiple sessions, with intervals in between (vs. studying the same material in a single longer session- massed practice; Carpenter et al., 2012). Studies have found evidence for the spacing effect, showing that distributed practice leads to better learning when compared to massed practice, despite the total time spent studying being equal in both revision methods (see reviews by Benjamin & Tullis, 2010; Cepeda et al., 2006; Donovan & Radosevich, 1999).

Rohrer and Taylor (2006, 2007) compared the test performance of undergraduates who learnt maths problems in one session (massed practice) or two shorter sessions separated by a one-week interval (distributed practice). The results showed that students who engaged in distributed practice significantly outperformed students who engaged in massed practice both one week and four weeks later. The learning benefits of distributed practice for maths have also been identified for children; Chen et al. (2018) found that children in China (*M* age = 11 years) who studied topics (e.g., how to calculate with negative numbers) that were distributed across three sessions on three consecutive days, significantly improved their performance compared to children who studied the same topics in a single session.

Although the spacing effect is well-established, most research has involved undergraduate students, and has not investigated how individual differences are influenced by distributed practice (Delaney et al., 2010). Some research has also identified conflicting results. For example, Barzagar Nazari and Ebersbach (2018) randomly assigned 44 students aged 15-17 years to distributed practice or massed practice. Students in the distributed practice condition completed three sets of online statistics exercises at home on three different days of the week. Students in the massed condition completed the same three sets of exercises but all on the same day. Two weeks later, unexpectedly, students in the massed condition (vs. distributed condition) performed better on a maths post-test. As significantly fewer students in the distributed condition completed all of the exercises (29% vs. 65% in the massed condition), the authors suggested that the increased effort required in distributed practice may have resulted in fewer students completing all learning tasks. Further analysis showed that females with better self-reported concentration skills (vs. males and females with weaker concentration skills) were significantly more likely to complete all of the distributed practice exercises.

2.1.2 Retrieval Practice

Retrieval practice (also known as practice testing) is where a learner attempts to recall learnt information from memory. There is a large body of evidence that confirms the testing effect- that retrieval practice leads to increased and more enduring learning when compared with a passive learning strategy such as restudying (e.g., see Adesope et al., 2017).

Roediger and Karpicke (2006a) investigated the testing effect with undergraduates in the US who were asked to recall a passage of text. All students took part in four 7-minute sessions, and all studied the text during the first session. In the subsequent three sessions, students either restudied the text during all three sessions, restudied the text for two sessions followed by a 7-minute test, restudied the text for a further session followed by two 7-minute tests or took part in three 7-minute tests. No feedback was provided during or following tests. Five minutes after the final session, the group who had restudied the text three times (with no test) recalled the most aspects of the text and the group who participated in three tests recalled the fewest. However, in delayed tests at two days later and one week later, students who had taken part in repeated testing performed best, and students who had taken part in repeated restudying performed the worst. Students who repeatedly restudied the text (vs. those who were tested), reported greater confidence that they would remember the text in a week, despite them remembering the least when tested. Other researchers have also found an association between restudying and over-confidence (e.g., Kornell & Son, 2009; Potts & Shanks, 2014). Hughes et al. (2018) found that both adults and children who studied word pairs followed by an immediate test with feedback (retrieval practice) were more accurate when predicting the number of word pairs that they would remember when tested, as compared to children and adults who restudied the word pairs.

Researchers have suggested that retrieval practice also has benefits beyond the recall of material learnt during sessions. For example, if a student frequently engages in tests, they may study more (Roediger & Karpicke, 2006b) or pay more attention to taught material (Lyle & Crawford, 2011). Other studies have identified support for the forward testing effect- that retrieval practice increases long-term recall of subsequently studied new information, even when new information is unrelated to previously tested information (Pastötter et al., 2013; Szpunar et al., 2013).

Further research has highlighted conditions under which the testing effect is most optimal for learning. These conditions include providing corrective feedback after a participant recalls an answer (Bangert-Drowns et al., 1991), or repeating retrieval practice at intervals over time (i.e., distributed practice; Binks, 2017; Rawson & Dunlosky, 2013). Hopkins et al. (2016) investigated the outcomes of spaced retrieval practice for students who enrolled in a pre-calculus course. These students either had ACT (American College Testing) scores below the calculus course requirement or felt unprepared for the course. Following the attempted retrieval of each answer, students saw the questions that they had answered incorrectly and could choose to view the correct answers. The findings indicated that both pre-calculus exam score and also calculus exam score the following year were higher in a spaced retrieval practice condition (vs. massed retrieval practice).

2.1.3 Successive Relearning

Successive relearning is a technique in which practice testing takes place to some level of mastery over multiple spaced sessions. Learners receive immediate feedback. If they answer an item correctly, this item is removed from the session. If they answer incorrectly, they see the correct response so that they can restudy it and return to it later in the same session (Bahrick, 1979). A similar method to successive relearning, interpolated testing, involves several short tests (retrieval practice) interpolated within a lecture, lesson or period of study. Previous research suggests that interpolated testing leads to improved learning and test performance as compared to restudying (Szpunar et al., 2013).

Studies involving undergraduates have provided support for the benefits of successive relearning when learning foreign-language word-pairs (Bahrick, 1979; Vaughn et al., 2016), definitions (Rawson & Dunlosky, 2013) and key concepts (Janes, accepted manuscript). Rawson et al. (2013), for example, investigated the impact of a twice-weekly successive relearning virtual flashcard intervention on undergraduates' recall of key definitions. In this study, participation in successive relearning significantly improved recall three days later and 24 days later, as compared to restudy, self-regulated practice, or no intervention control.

Higham et al. (manuscript in preparation) explored the effect of a successive relearning intervention for undergraduates in the UK. After each weekly lecture, students received emailed links to either a successive relearning or a restudying intervention. In the relearning intervention, for each item, students were asked to fill in the missing keyword to complete the sentence before receiving corrective feedback. In restudying, students read the completed equivalent sentences. All items were relevant to lecture content. Students completed revision sessions three times each week over ten weeks, and sessions were spaced two, four and six days after the lecture. Each week, half of the students engaged in relearning and half engaged in restudying. Then, during the following week, each student engaged in the alternative intervention so that all students experienced both conditions. The findings indicated that the recall of lecture content was higher following successive relearning than restudying. Qualitative feedback additionally indicated that students reported greater enjoyment and value in successive relearning (vs. restudying).

2.1.4 Motivational and Affective Factors Related to Learning

Further research has highlighted motivational/affective factors associated with mathematics performance in adolescents. These factors include a more positive attitude towards maths (Hemmings & Kay, 2010), increased motivation (Robbins et al., 2004), and lower maths anxiety (Devine et al., 2012). In a study that included 6,689 16-year-olds in the UK, maths self-efficacy and interest were also significantly associated with GCSE Maths grade (Tosto et al., 2016). The relationships between motivational/affective variables and maths achievement are likely to be bi-directional. For example, Pekrun et al. (2017) analysed longitudinal data from a sample of children in Germany from 5th grade through to 9th grade (11-16 years old). Positive emotions such as enjoyment and pride positively predicted maths performance and maths performance positively predicted these emotions. The relationship between anxiety and maths performance is also likely to be reciprocal; Ma and Xu (2004) identified that secondary school students' low performance in maths was significantly associated with future high maths anxiety.

The learning benefits of retrieval practice and distributed practice are well established, and there is an emerging evidence base for successive relearning. There is a further small body of research on the motivational and affective outcomes associated with these techniques. For example, in one study, 72% of 11-18-year-olds reported that retrieval practice with immediate feedback made them less nervous for exams, 22% reported no difference and 6% reported increased nervousness (Agarwal et al., 2014). Other research has concluded that interpolated testing (vs. interpolated restudying) has cognitive and affective benefits for undergraduates including increased attention, and decreased exam anxiety and negative affect towards the final exam (Szpunar et al., 2013). Higham et al. (manuscript in preparation) similarly identified that successive relearning (vs. restudying) was associated with increased metacognition, reduced anxiety, improved sense of mastery, and improved attention, across three sessions within one week that were on the same topic.

2.1.5 The Present Study

The present study aimed to replicate Higham et al. (manuscript in preparation) to investigate whether successive relearning may be beneficial for 16-18-year-olds who were re-sitting GCSE Maths. Specifically, it investigated the impact of an online successive relearning intervention (RL vs. restudying RS intervention vs. no intervention control C) for post-16 GCSE Maths students to consider: (1) Does engagement in an RL intervention (vs. RS vs. C) predict maths test performance? (2) Is engagement in an RL intervention (vs. RS vs. C) associated with positive affective outcomes in trait-like measures of attitudes to mathematics, academic efficacy, attentional control, general anxiety, maths anxiety and metacognition? (3) Is engagement in an RL intervention (vs. RS vs. C) associated with positive changes in state indices of anxiety, mastery, attention and judgements of predicted test score as reported immediately following intervention sessions? Furthermore, (4) What are participants' experiences of the interventions?

It was hypothesised that students who participate in the RL intervention would score highest on a maths test, followed by students in the RS condition, with students in the C condition scoring lowest. It was additionally hypothesised that, as compared to students in the RS or C conditions, students in the RL condition would show a more positive change from time 1 (pre-intervention) to time 2 (post-intervention) in (1) trait indices of attitudes towards maths, academic-efficacy, attentional control and anxiety, and accuracy of predicted test score (judgment of learning) and (2) state indices of anxiety, mastery and attention, and predicted test score. This study was pre-registered on the Open Science Framework (osf.io/agxtw).

2.2 Method

2.2.1 Participants

Students were recruited from one Further Education (FE) college. All students had previously achieved a grade 3 in GCSE Maths and were enrolled onto compulsory re-sit classes. From the 281 students aged 16-18 years and due to re-sit GCSE Maths at the college, 157 students (56%; n=94 female, n=1 chose not to report gender) gave informed consent and completed baseline measures. Alongside GCSE Maths, participants were studying a range of academic (GCSEs/ A-levels) and/or vocational courses (Diplomas/ BTECs). All students were aged between 16 and 18 years (n=91 were 16, n=47 were 17, n=18 were 18, n=1 not reported).

In the intervention analysis, we planned to include students in the RL and RS groups who completed at least 80% of sessions. However, due to low engagement, we revised this to include students who completed at least one session (n=51; RL=22, RS=29). Sixty-eight students completed the post-test and also either completed at least one session or were in the control group (RL=19, RS=26, C=23), and 48 students completed both pre- and post-intervention questionnaires and completed at least one session or were in the control group (RL=12, RS=24, C=12).

2.2.2 Design

The present study included one between-subject factor that included three groups: successive relearning intervention (RL) vs. restudying intervention (RS) vs. business as usual control (C). There was additionally one within-subject factor that included two time-points: pre- (T1) and post- (T2) intervention.

2.2.3 Intervention

The RL and RS groups were based on Higham et al. (manuscript in preparation). There were three sessions each week over four weeks (12 sessions total). The three sessions within the week were identical. Each session lasted a maximum of 25 minutes. Sessions were accessed online on a computer, tablet or smartphone via the software, Qualtrics. Students received email reminders to complete sessions, and links to the sessions were also uploaded onto the college's online student area.

During each session, the RL group completed 20 maths statements based on the material they had learnt in class the previous week, for each of four topics across the four weeks (see Table 3). The statements were developed by the first researcher after consulting the AQA GCSE specification, and a maths teacher external to the college agreed that all questions were appropriate. Each statement had one word missing, so students had to type in the correct missing word and were given corrective feedback if the answer was incorrect. There was a requirement that each question was answered correctly once, using a dropout methodology whereby questions that were answered incorrectly were seen and answered again in the same session until answered correctly, up to a maximum of three times. Students in the RS condition viewed the same 20 statements that corresponded to those in the RL intervention. However, in RS, the statements were complete, and students were required to read the statement, rather than recall the missing word (see Appendix G).

Table 3

Example Items From the Relearn and Restudy Sessions. The Second Column of the Table Shows the Maths Topic That the Questions Within the Week Related to

Week	Maths topic	Relearn group	Restudy group
1	Number operations and estimating	In the expression y < x, the symbol < shows that the value of y is____than x.	In the expression y < x, the symbol < shows that the value of y is less than x.
2	Fractions, decimals and percentages	A number that contains both a whole part and a fraction, for example, 3½, is called a number.	A number that contains both a whole part and a fraction, for example, 3,½, is called a mixed number.
3	Ratio estimation and rounding	The number 13.437 to 2 significant figures is_.	The number 13.437 to 2 significant figures is 13.
4	Conversions, exchange rates and negative numbers	40 000 in___form is 4 x 10^4.	40 000 in standard form is 4 x 10^4.

To control for exposure to revision material, RL students had 15 seconds to type their answer to each question the first time it was seen, followed by 5 seconds of feedback (20 seconds exposure per item). On subsequent attempts, RL students had 10 seconds to type their answer, followed by 5 seconds of feedback (15 seconds per item). As it was estimated that the average number of attempts required to answer correctly would be two in the RL group, RS students saw the complete sentences twice in each session- for 20 seconds on the first occasion, and 15 seconds on the second occasion.

2.2.4 Measures

Students completed the same questionnaires at T1 and T2. At T1, students additionally reported their age, gender and the course(s) that they were enrolled onto (in addition to GCSE Maths). At T2, students also reported the number of GCSE Maths re-sits they had taken. School staff additionally provided data on student attendance at maths lessons during the six weeks of data collection.

2.2.4.1 Maths performance

Learning was assessed using a 15-minute paper test (see Appendix H) that students completed in class in exam conditions. Like the intervention items, the test was written by the first researcher after consulting the appropriate exam specification and a GCSE Maths teacher external to the college. The test consisted of 36 questions (scored 0-36). Of these questions, 12 were the same as items used in RL sessions (same missing word; i.e., exact questions), 12 assessed near-transfer effects (i.e., the questions comprised of the same sentences used in RL items but the missing word within the sentence was different) and 12 assessed far-transfer effects (i.e., the questions covered material within the same topic as those within the sessions but students had not seen the questions before; see Table 4). As in the RL intervention, students were required to fill in the missing word to complete each sentence.

Table 4

An Example of the Original Question as Taken From a Relearn Session, Alongside How the Question Would be Adapted as a Near-Transfer or Far-Transfer Question

Original question	Near-transfer	Far-transfer
A number that can only be divided by [1] and itself is called a prime number.	A number that can only be divided by 1 and itself is called a [prime] number.	When rounding to the nearest 10, you would round down if the last number is [4] or less.

Note. Students were asked to recall the word in parentheses.

2.2.4.2 Pre- and Post-Intervention Questionnaire Measures

2.2.4.2.1 Attitudes to Maths

We used the attitudes toward mathematics inventory (ATMI, Tapia, 1996; Appendix I) to measure attitudes about maths. The ATMI is a 40-item measure consisting of four subscales: sense of security, enjoyment, perceived value and motivation. Respondents indicated to what extent they agree with each statement on a 5-point scale from *strongly disagree* (score 1) to *strongly agree* (score 5). We used the total score, whereby a higher score indicates more positive attitudes about mathematics (score range = 40-300). Previous research has highlighted high to excellent internal consistency for each scale and satisfactory test-retest reliabilities over four months (Tapia & Marsh, 2004). In the current study, internal consistency was excellent at both time points ($\alpha > .9$).

2.2.4.2.2 Academic Efficacy

We used the academic efficacy scale from the patterns of adaptive learning survey (PALS, Midgley et al., 2000; Appendix J). The scale includes five items (e.g., “I'm certain I can master the skills taught in class this year.”) Students indicated how true each item is of them from *not at all true* (score 1) to *very true* (score 5; score range = 5-25). In the current study, we asked students to think about maths when answering questions. The scale is a widely-used measure (Honicke & Broadbent, 2016) that is reliable and valid (Midgley et al., 2000). In a sample of 10-14-year-olds in the US, the authors found acceptable internal consistency ($\alpha = .78$; Midgley et al., 2000). In the current study, the scale had high internal consistency ($\alpha > .8$ for both time points).

2.2.4.2.3 Attentional Control

We used the attentional control scale (ACS, Derryberry & Reed, 2002, Appendix K). Students selected to what extent each of the 20 statements is true for them from *almost never* (score 1) to *always* (score 4). An example statement is, “When concentrating, I can focus my attention so that I become unaware of what's going on in the room around me.” A higher score indicates better attentional control (score range = 20-80). Researchers have identified that the ACS has good internal consistency (Derryberry & Reed, 2002, $\alpha = .88$; Ólafsson et al., 2011, $\alpha = .84$). In the current study, Cronbach’s alphas showed good internal consistency at both time points ($\alpha > .8$).

2.2.4.2.4 General Anxiety

We used the general anxiety disorder (GAD) subscale of the youth anxiety measure for DSM 5 (YAM-5, Muris et al., 2017; Appendix L). The scale consists of six statements (e.g., “I worry about a lot of things”). Students indicated how true each statement is for them from *never* (score 0) to *always* (score 3). Higher scores indicate more feelings of anxiety (score range = 0-18). Muris et al. (2017) found high internal consistency for the questionnaire in a non-clinical sample of 12-17-year-olds in The Netherlands. In the current study, internal consistency was excellent at both time points (α > .9).

2.2.4.2.5 Maths Anxiety

We used the abbreviated maths anxiety scale (AMAS, Hopko et al., 2003; Appendix M). Students read nine statements of maths events (e.g., “thinking about an upcoming mathematics test one day before”) and responded on a 5-point scale whether they would feel *low anxiety* (score 1) to *high anxiety* (score 5) about the event (score range = 9-45). Higher scores indicate more anxiety about maths. Hopko et al. (2003) found good test-retest reliability and internal consistency. In the current study, internal consistency was excellent at both time points (α > .9).

2.2.4.2.6 Judgment of Learning (JOL)

Students prospectively predicted their score on the T2 maths test (from 0-36). We also used this prediction to generate an accuracy score by calculating the difference between students’ predicted score and their later attained score. A score of 0 indicated that the student’s prediction was entirely accurate (e.g., they predicted that they would achieve 20 marks on the test, and they later achieved 20 marks). A higher number of marks difference indicates a less accurate prediction. Where students provided a range for their predicted test score (e.g., 20-24), the median value was taken (e.g., 22).

2.2.4.2.7 Feelings of Anxiety, Mastery and Attention and Judgment of Learning

After each intervention session, participants responded to nine “I feel” statements on a sliding visual analogue scale (VAS; Appendix N). Three items measured feelings of anxious affect (anxious, worried, relaxed), three items measured feelings of mastery (confused, I have mastered this topic, I understand this topic) and three items measured feelings of attention (focused, disorganised, I concentrated). For each session, the scores for each item (0-100) were reversed if necessary (e.g., “I feel relaxed”) and the mean was calculated to generate a score (0-100) for each construct (anxious affect, mastery, attention), with a higher score indicating greater anxiety, greater mastery and better attention.

Following each session, students also provided a JOL by prospectively predicting their T2 maths test score (0-36).

2.2.4.2.8 End-User Feedback

Once students had finished the T2 questionnaires, they answered two open questions online: “If you have missed any intervention sessions, what was the reason(s) for this?” and

"You have finished four weeks of the online maths intervention. What did you think about these sessions?"

2.2.5 Procedure

The researcher emailed information about the study (Appendix O), including participant (Appendix P) and teacher information sheets (Appendix Q) to the staff at five colleges. The Head of Maths at one college subsequently agreed to take part and introduced teachers to the study during a departmental staff meeting. Parents and guardians received information about the study via the college parent contact system and were given the option to opt-out of the study on behalf of their child (n=0). Teachers introduced students to the study during a maths lesson at the beginning of the 2019/20 academic year, using the participant information sheet.

A computer program was used to randomly allocate the 281 eligible students (by unique identification code) into three groups: RL (n=112), RS (n=113) and C (n=56). Students were over-allocated (40%) to the RL and RS conditions, as compared to the C group (20%), due to anticipated greater attrition from these groups. Of the students who filled in T1 measures, 60 students (38%) were in the RL condition, 62 students (39%) were in the RS condition, and 34 students (22%) were in the C condition.

Data collection took place over six weeks from September 2019 to October 2019. During a maths lesson at T1, students were informed about the study, gave consent, and completed the T1 questionnaire battery. The order of questionnaires was randomised, and the median time taken to complete questionnaires was 14.1 minutes. Over the following four weeks, students in the RL and RS conditions completed revision sessions outside of lesson time on Monday, Wednesday and Friday. Immediately following each session, students reported their feelings of anxiety, mastery and attention (VAS) and predicted their T2 maths test score (JOL). Between T1 and T2, all students attended maths lessons and received homework as usual. At T2, in class, all groups completed the same questionnaires as at T1 and also completed end-user feedback. This was followed by the maths test.

2.2.6 Ethics

Ethical approval was granted by The University Ethics and Research Governance at The University of Southampton (submission number 47186; Appendix R). As data were collected online, participants were shown a consent statement (Appendix S) at the start of the T1 and T2 questionnaires and gave informed consent by ticking the box to continue. Students were informed of their right to withdraw. Seven students emailed the researcher to withdraw from the study; five students withdrew as they were no longer taking GCSE Maths, one withdrew due to a reported lack of time to complete sessions, and one withdrew due to disinterest. At the end of the study, students received a debriefing sheet (Appendix T) by email. This was also read to them by their teacher.

Allocation to group was carried out before gaining informed consent to allow teachers to deliver all instructions relating to the study within a single lesson. Where a student did not provide consent, the information held by the researcher (identification code and group) was deleted. A list of student names against identification codes was held securely at the college and was never removed from the site by the researcher.

Participants in the RL and RS conditions received a 50p (college cafe) food voucher per completed session. Participants who completed at least ten sessions received a £20 Amazon voucher, and participants who completed all three sessions in the final week received an additional £10 Amazon voucher. Students in all three conditions who completed both T1 and T2 questionnaires were entered into a prize draw for three £50 Amazon vouchers.

2.3 Results

2.3.1 Approach to Analysis

The analysis aimed to understand group differences (RL vs. RS vs. C) on learning outcome (i.e., maths test score), trait-like measures of attitudes toward maths, academic efficacy, attentional control, anxiety, predicted test score and accuracy of test score, as well as within-week differences in anxious affect, mastery, attention and predicted test score. Further analysis considered student views of the intervention.

Figure 9 shows the analysis of the results, highlighting three broad parts: analysis of (A) T1 data (analyses 1, 2 and 3), (B) the intervention including individual session data relating to state measures (analysis 4), test data (analysis 5) questionnaire data (analysis 6), and (C) students' experiences of the interventions (analysis 7). As few students met the intended threshold for inclusion in the analysis of completing 80% of sessions (RL=7, RS=5), students in the RL and RS groups who had completed at least one session were included. Furthermore, we had intended to examine individual session data (anxious affect, mastery, attention and predicted test score), over the 12 sessions. However, due to low engagement in sessions, differences in student reporting were compared within-week where students had completed at least two sessions within a week (i.e., sessions 1 and 2, sessions 2 and 3 or sessions 1 and 3). Where students had completed all three sessions in a week, the change between sessions 1 and 3 was considered. If students completed at least two sessions per week for more than one week (/4), then a session mean was used in the analysis, e.g., if a student completed two sessions in week one and two sessions in week two, then the mean of both of the first sessions within-weeks, and the mean of the last sessions within-weeks, was calculated. We hypothesised that there would be a greater increase in mastery, attentional control and predicted test score, and a greater decrease in anxious affect in the RL group (vs. the RS group).

2.3.1.1 Approach to Analysis for Quantitative Data

To analyse T1 data and the intervention, statistical tests were conducted using IBM SPSS version 26. The assumptions of parametric testing were checked in T1 data across the sample and within groups, and for each analysis. The assumptions of independence of observations and linearity were met in all instances. All dependent variable data were continuous and of interval or ratio level. Normal distribution was assessed using the Shapiro-Wilk test, alongside a visual inspection of histograms and Q-Q plots. The spread of scores around the mean was assessed using Levene's test of equality of variances. Homogeneity of covariance matrices was assessed using Box's test, and multicollinearity was assessed using Pearson's correlation. Outliers were identified by visual inspection of box plots and an examination of studentised residuals for values greater than ± 3. Multivariate outliers were examined using Mahalanobis distance ($p > .001$).

Although the statistical analyses used are robust to deviations from normal distribution (Maxwell & Delaney, 2004), where data were not normally distributed, analyses were repeated using non-parametric tests, and any differences are reported. Where the impact of time (e.g., T1 vs. T2) was examined, we calculated change scores to carry out non- parametric tests. Where the assumption of homogeneity of variance was violated, Welch's test is reported. There were no violations of homogeneity of covariance matrices. Analyses were conducted both with and without outliers, and any differences are reported. Where the removal of an outlier did not make a difference to significance, results are reported with outliers included. All outliers were genuine values.

2.3.1.2 Approach to Analysis for Qualitative Data

End-user feedback was analysed across groups using Braun and Clarke's (2006) six-stage process of thematic analysis. Table 5 shows how this process was carried out in the present study. We considered reporting the number of endorsements of each theme per group (RL, RS) but decided against this as Braun and Clarke's method is intended to be "fully qualitative" (Clarke & Braun, 2018, p.107). Moreover, Braun and Clarke (2019) argued that frequencies do not determine the value or importance of theme or sub-themes and reflect anxiety in the validity of one's qualitative research.

Table 5

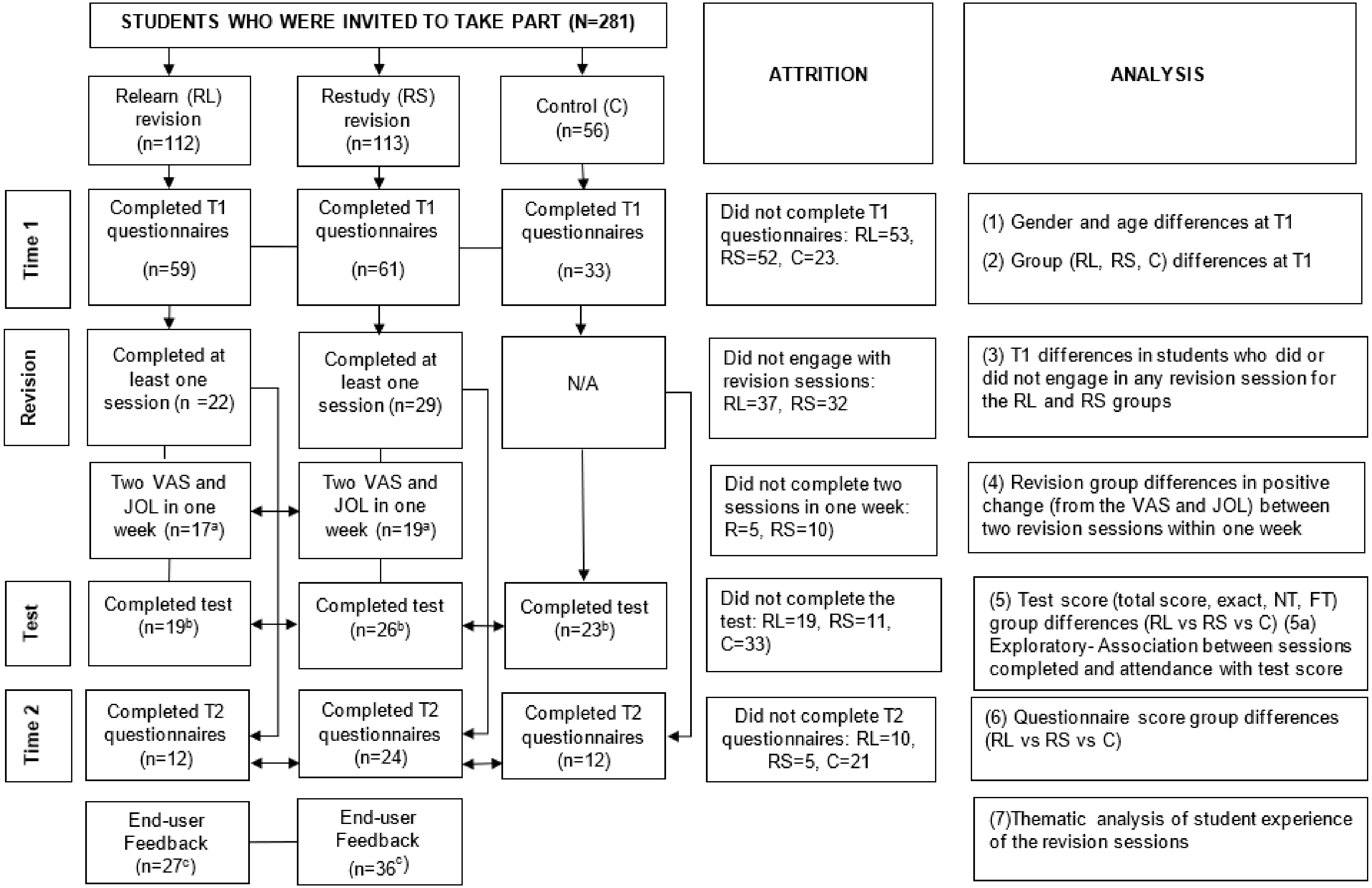
How Braun and Clarke's (2006) Stages of Thematic Analysis Were Applied to Analyse End-User Feedback in the Current Study

Stage	Process
1. Data familiarisation	• Student responses to end-user feedback questions were downloaded from Qualtrics. • The first researcher repeatedly and actively read through the questionnaire responses, recording initial ideas of codes and patterns. • The researcher discussed initial thoughts relating to the responses with the other researchers.
2. Coding	• The first researcher manually assigned codes to the data. Coding was carried out at the semantic level and was driven by the data (inductive) rather than by theory (deductive).
3. Generation of initial themes	• Codes were recorded on paper and were arranged into initial main themes (based on shared meaning), and subsequently also sub-themes. • During this process, the first researcher engaged in reflexive discussions with the second researcher regarding how they had coded and analysed the data. • An initial thematic map was developed.
4. Reviewing themes	• Themes and sub-themes were reviewed and reworked several times until the final themes and sub-themes were reached. • During this process, the researcher referred back to the original data set to ensure that the themes and sub-themes reflected students' responses. • A final thematic map was developed.
5. Defining themes	• A "coding manual" was developed to contain the name of each theme and sub-theme, their definitions and example extracts. • Themes and sub-themes were named after considering all of the codes which were included within the theme/sub-theme.
6. Reporting analysis	• A written account of the data, including themes, sub-themes and student quotes, was produced to consider students' experiences of the intervention.

2.3.2 Analysis of T1 Data

Figure 9 shows the number of students in each condition who informed each analysis, alongside participant attrition and the analyses carried out. Figure 9 shows that, of the 281 students who were invited to participate in the study, 54.5% (N=153) completed T1 questionnaires (RL=59/53%; RS=61/54%; C=33/59%). An additional four students who enrolled in the GCSE Maths course late completed the T1 questionnaire; however, these students were excluded from the intervention analysis as they had not been allocated to condition. Some students who did not complete the T1 questionnaire completed the T2 questionnaire (n=18; RL=3, RS=9, C=6) and/or the maths test (n=22; RL=7, RS=8, C=7). This variation was due to student absence when T1 questionnaires were completed, or student enrolment late in the academic year.

Figure 9. *The Number and Attrition of Students Through the Study, Measures at Each Time Point and Related Analyses*



Note. RL = relearn, RS = restudy, C = control, T1 = time 1, T2 = time 2, VAS = visual analogue scale of reported anxious affect, mastery and attention, JOL = judgment of learning (predicted test score). [a]Only students who completed at least two sessions in any given week were included in analysis. [b]Students who were either in the RL or RS group and completed at least one session or were in the control group and who completed the test. Some students in this analysis completed the test and not the T1 questionnaire (RL = 5, RS = 2, C = 11). [c]Students in the RL and RS groups who provided feedback as part of the T2 questionnaire. Some students in this analysis did not complete the T1 questionnaire (RL = 3, RS = 9) and/or did not engage with any revision session (RL = 16, RS = 15).

The means (*M*) and standard deviations (SD) for experimental groups (RL, RS, C) for each measure at T1 and T2, including every student who completed each measure, are in Appendix U. Table 6 shows the *M* and SD for each group for each questionnaire measure at T1 at T2 for students who completed the questionnaire at both time points and either engaged in at least one session (RL, RS) or were in the control group. Table 6 also shows the *M* and SD for each group for maths test data for students who engaged in at least one session or were in the control group.

Table 6

The Means and Standard Deviations of Questionnaire Measures for Students who Either Engaged in at Least One Intervention Session (RL or RS) or Were in the Control Group and a) Completed Both T1 and T2 Questionnaires and/or b) Completed the Test

	Experimental Condition											
	Relearn (RL)				Restudy (RS)				Control (C)			
	T1		T2		T1		T2		T1		T2	
	M	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD	*M*	SD
Questionnaire measures												
n	12				24				12			
Attitudes toward maths	109.92	25.10	112.58	25.89	121.21	23.32	120.50	26.73	116[a]	22.86[a]	115.10[a]	21.55[a]
Academic efficacy	16.17	3.54	16.17	4.22	17.04	4.08	17.25	4.12	16.27[b]	3.82[b]	15.82[b]	4.92[b]
Attentional control	47.25	10.64	46.67	9.03	49.46	9.53	50.92	8.77	47.30	10.78	48.60	11.78
General anxiety	10.67	5.00	11.00	4.35	9.96	5.32	9.50	4.88	8.42[a]	4.25[a]	8.50[a]	3.60[a]
Maths anxiety	23.75	9.61	22.83	9.14	22.38	11.25	21.17	10.12	19.91[b]	9.02[b]	19.91[b]	6.96[b]
Test scores												
n			19				26				23	
Total score			23.16	6.78			22.73	4.90			20.26	6.47
Exact score			8.00	3.15			7.54	2.34			6.61	2.41
Near-transfer score			7.84	2.12			8.04	2.03			7.22	2.30
Far-transfer score			7.32	2.38			7.12	1.84			6.43	2.57
Engagement indices												
n	13				27				20			
Attendance			92.24	13.14			86.76	16.12			89.47	12.87
Number of GCSE Maths resits	1.23	0.59			1.41	1.05			1.75	1.45		

Note. *M* = Mean, SD = standard deviation, n = number of participants. Participant numbers vary for some measures due to some incomplete questionnaires.
[a] n=10 [b] n=11

Table 7 shows correlations between T1 data, test score and number of sessions (RL/RS) completed. It shows that scores on all five questionnaires were significantly correlated at T1, except for the correlation between academic efficacy and general anxiety ($r(118) = -.10$, $p > .05$). Specifically, a more positive attitude towards maths was moderately positively associated with academic efficacy, attentional control, and was negatively associated with general anxiety and maths anxiety (small and moderate correlations, respectively). Increased academic efficacy was also moderately linked to greater attentional control and lower maths anxiety (small correlation). There was a small positive correlation between attentional control and attendance. In addition, students who reported greater anxiety were more likely to report poorer attentional control (moderate correlation). General anxiety and maths anxiety were strongly correlated, indicating that students who reported higher general anxiety, were also likely to report anxiety relating to maths.

With regards to test data, students' predictions of their test score were significantly and moderately correlated with all T1 questionnaire measures, indicating that students who predicted that they would attain a higher score on the test, also reported more positive attitudes towards maths, higher academic efficacy and greater attentional control, as well as fewer symptoms of anxiety. Total test score and accuracy of predicted score were uncorrelated with all other variables at T1 (rs < .25, ps > .05).

Across groups (RL, RS), the number of sessions completed (0-12) was significantly correlated with more positive attitudes toward maths (small correlation), increased maths anxiety (small correlation) and increased test score (strong correlation). The number of GCSE Maths re-sits taken was not correlated with any T1 measures, test score or engagement indices.

Table 7. *Correlations Between all T1 Measures, Test Score, and Indices of Engagement Across Groups (RL, RS)*

Variables	1	2	3	4	5	6	7	8	9	10	11
Questionnaires											
1. Attitudes toward maths	--	.49**	.33**	-.27**	-.38**	.22**	-.14	-.07	.21*	.11	-.18
2. Academic efficacy		--	.40**	-.10	-.23**	.43**	-.14	.08	.04	.02	-.19
3. Attentional Control			--	-.37**	-.45**	.34**	.08	.00	-.06	-.20*	.10
4. General anxiety				--	.65**	-.25**	.19	.10	.14	.10	-.05
5. Maths anxiety					--	-.28**	.11	.18	.19*	.05	-.09
Test											
6. Predicted test score						--	.12	.21	.11	-.14	-.09
7. Actual total test score[a]							--	.08	.54**	.08	-.08
8. Difference between predicted and actual scores								--	.07	-.18	-.17
Engagement indices											
9. Number of intervention sessions completed[b]									--	.04	-.11
10. Attendance (%) to maths lessons										--	-.18
11. Number of GCSE maths re-sits											--

Note. [a]There were strong correlations (rs = .55-.89) between total test score, exact question score, near-transfer question score and far-transfer question score. Like total test score, exact, near-transfer and far-transfer question score were not correlated with other T1 variables, except for a small correlation between exact question score and predicted test score r = .23, p = .029). [b]Excluding students in the control group. * two-tailed $p < .05$ ** two-tailed $p < .01$. RL group only : number of intervention sessions correlated with test score (r = .66, $p < .001$) and predicted test score (r = .30, p = .029); test score correlated with number of intervention sessions. RS group only: number of intervention sessions correlated with test score (r = .38, p = .008), attitudes towards maths (r = .31, p =.016) and maths anxiety (r =.29, p = .023); test score correlated with number of sessions and difference between predicted and actual scores (r = .46, p = .005).

2.3.2.1 Gender and Age Differences at T1

Four two-way ANOVAs were conducted to examine the impact of gender (60 males vs. 96 females) and age (92 16-year-olds vs. 64 17/18-year-olds) on T1 questionnaire data, number of GCSE Maths re-sits taken, attendance to maths lessons and total test score. Only significant main effects, or those approaching significance, are discussed (for all non-significant findings $Fs < 1.8$, $ps > .1$). There was no statistically significant interaction between gender and age in any of these analyses ($Fs < 2.5$, $ps > .10$).

2.3.2.1.1 Gender Differences

The main effect of gender on attitudes towards maths was significant, although the effect was very small ($F(1,150) = 6.77$, $p = .010$; partial $\eta^2 = .043$), highlighting that males (vs. females) reported more positive attitudes towards maths; *M*(SD) were 115.97(19.51) for males and 106.71(25.59) for females. The main effect of gender on academic efficacy approached significance ($F(1,148) = 3.885$, $p = .051$; partial $\eta^2 = .026$), highlighting that males (vs. females) descriptively reported higher academic efficacy; *M*(SD) were 16.23(3.79) for males and 15.08(4.03) for females. When outliers were removed (4 males), the main effect of gender on academic efficacy is significant ($F(1,144) = 4.369$, $p = .038$; partial $\eta^2 = .029$); *M*(SD) were 16.28(3.18) for males and 15.08(4.03) for females. The main effect of gender on attentional control was significant ($F(1,150) = 6.359$, $p = .013$; partial $\eta^2 = .041$), highlighting that males reported better attentional control than females; *M*(SD) were 51.18(8.15) for males and 47.28(9.13) for females. The main effect of gender on general anxiety was significant ($F(1.150) = 11.124$, $p = .001$; partial $\eta^2 = .069$), with females (vs. males) reporting higher general anxiety; *M*(SD) were 7.81(4.69) for males and 10.55(4.82) for females. The main effect of gender on maths anxiety was also significant ($F(1,149) = 10.755$, $p = .001$; partial $\eta^2 = .067$), with females reporting higher maths anxiety as compared to males; *M*(SD) were 17.90(7.86) for males and 22.93(8.85) for females.

The main effect of gender on predicted test score was significant ($F(1,139) = 4.376$, $p = .038$; partial $\eta^2 = .031$), highlighted that males (vs. females) predicted that they would attain higher marks in the maths test; *M*(SD) were 21.16(6.88) for males and 19.22(4.74) for females. When the difference between predicted test score and attained test score was calculated to generate an accuracy score, the main effect of gender on predicted test accuracy was also significant ($F(1,75) = 5.336$, $p = .025$; partial $\eta^2 = .066$), revealing that females (vs. males) were more accurate in their predicted scores (fewer marks difference between predicted and total test score); *M*(SD) were 7.22(5.96) for males and 5.00(3.85) for females.

2.3.2.1.2 Age Differences

The main effect of age on attitudes towards maths was not significant (*F*(1,150) = 3.101, *p* = .080; partial η^2 = .02), although, descriptively, 16-year-olds (vs. 17/18-year olds) reported more positive attitudes towards maths; *M*(SD) were 112.78(26.05) for 16-year-olds and 107.23(19.72) for 17/18-year-olds. The main effect of age on number of GCSE Maths re-sits was significant (*F*(1,69) = 56.476, *p* < .001; partial η^2 = .45); *M*(SD) were 0.98(0.26) for 16-year-olds and 2.38(1.12) for 17/18-year olds, showing that 17/18-year-olds had taken more re-sits than 16-year-olds. The main effect of age on attendance was also significant (*F*(1,152) = 4.717, *p* = .031; partial η^2 = .03), highlighting that the attendance of 16-year olds was higher than 17/18-year olds; *M*(SD) were 86.45(16.9) for 16-year-olds and 80.75(19.59) for 17/18-year-olds.

2.3.2.2 Experimental Group Differences at T1

A series of one-way ANOVAs were conducted to examine differences between the three experimental groups on T1 questionnaire measures (attitudes toward maths, academic efficacy, attentional control, general anxiety, maths anxiety and predicted test score), attendance and number of GCSE Maths re-sits taken. For all analyses, there were no statistically significant differences between students in the RL, RS and C groups (*Fs* < 1, *ps* >.1).

Of the students who completed at least one intervention session (/12), students in the RL group completed an average of 3.84(3.84) sessions and students in the RS group completed an average 3.70(3.35) sessions. This difference was not significant (*t*(76) = .175, *p* = .86, *d* = .040).

2.3.2.3 Differences in Engaged vs. Disengaged Students at T1

Figure 9 shows that, of the 153 students who completed the T1 questionnaire, 43% of students in the intervention conditions (RL, RS) engaged by completing at least one session (RL = 22/37%; RS = 29/48%). A series of t-tests were conducted to compare T1 measures between students across the RL and RS groups who did vs. did not engage in at least one session. This analysis showed that there were no significant differences between students who engaged and did not engage in most T1 questionnaire measures. The difference did, however, approach significance for attitudes towards maths (*t*(89.6) = 1.98, *p* = .051, *d* = .418); descriptively, students who engaged in at least one session reported more positive attitudes toward maths at T1; *M*(SD) were 115.59(27.35) for students who engaged, compared to 106.52(20.44) for those who did not engage. The difference between engaged and disengaged participants also approached significance for number of GCSE Maths re-sits taken ($t(87) = -1.925$, $p = .058$, $d = .413$), with those who engaged having taken fewer re-sits; *M*(SD) were 1.35(0.92) for engaged students and 1.78(1.12) for disengaged students. Similarly, the difference between engaged and disengaged students approached significance for academic self-efficacy ($t(115) = 1.882$, $p = .062$, $d = .351$); *M*(SD) were 16.37(4.4) for those who engaged, compared to 14.97(3.66) for those who did not engage. When the outlier in academic efficacy data was removed (1 engaged), the academic efficacy of students who engaged was significantly higher, ($t(114) = 2.206$, $p =.029$, $d = .413$); *M*(SD) were 16.58(4.19) for engaged students, compared to 14.97(3.66) for those who did not engage.

Furthermore, correlations of time 1 data (Table 7) showed that, at T1, predicted test score in the T2 maths test was positively correlated with number of RL sessions completed (0-12) and, in the RS group, reported attitudes towards maths and maths anxiety was positively correlated with number of RS sessions completed.

2.3.3 Analysis of the Intervention

2.3.2.4 Analysis of Individual Session Data

A series of two-way mixed ANOVAs were conducted to investigate the impact of group (RL, RS) and time (session 1, session 2) on reported anxious affect, mastery, attention and predicted test score. Session 1 in this analysis refers to the first session that students completed within a week (this could be session 1 or session 2 of the three possible sessions within a week), and session 2 refers to the last session that participants completed in the same week (this could be session 2 or session 3 within the week). So that each participant's data was included only once in the analysis, where a participant had completed two sessions within more than one week (e.g., two sessions in week 2 and two sessions in week 4), we used this data to calculate a single mean for all of the first sessions (within each week) that they completed, and a second single mean for all of the final sessions (within each week) that they completed. Only significant results are reported; for all other analyses (both relating to main effects and interactions), $F < 2.2$, $p > .1$. Means for each group at each time point, for each construct, are presented in Figure 10.

2.3.2.4.1 Anxious Affect

The main effect of time showed a statistically significant decrease in reported anxious affect from session 1 to session 2 ($F(1,32) = 15.048$, $p <.001$; partial $\eta^2 = .320$); *M*(SD) were

36.56(26.51) at session 1 and 27.67(21.72) at session 2.

2.3.2.4.2 Mastery

There was a statistically significant interaction between intervention group and time on reported mastery ($F(1,31) = 4.27$, $p = .047$; partial $\eta^2 = .121$). Planned comparisons showed that the difference between the RL and RS groups for session 1 approached significance ($F(1,31) = 3.435$, $p = .073$; partial $\eta^2 = .1$); *M*(SD) were 52.12(20.93) for the RL group and 66.90(24.28) for the RS group. The difference between the RL and RS groups for session 2 was not significant ($F(1,31) = .255$, $p = .617$; partial $\eta^2 = .008$); *M*(SD) were 71.84(20.56) for the RL group and 75.59(21.66) for the RS group. Considering differences between session 1 and session 2, there was a significant effect of time on reported mastery for both the RL group ($F(1,14) = 17.098$, $p = .001$; partial $\eta^2 = .550$) and the RS group ($F(1,17) = 9.340$, $p = .007$; partial $\eta^2 = .355$).

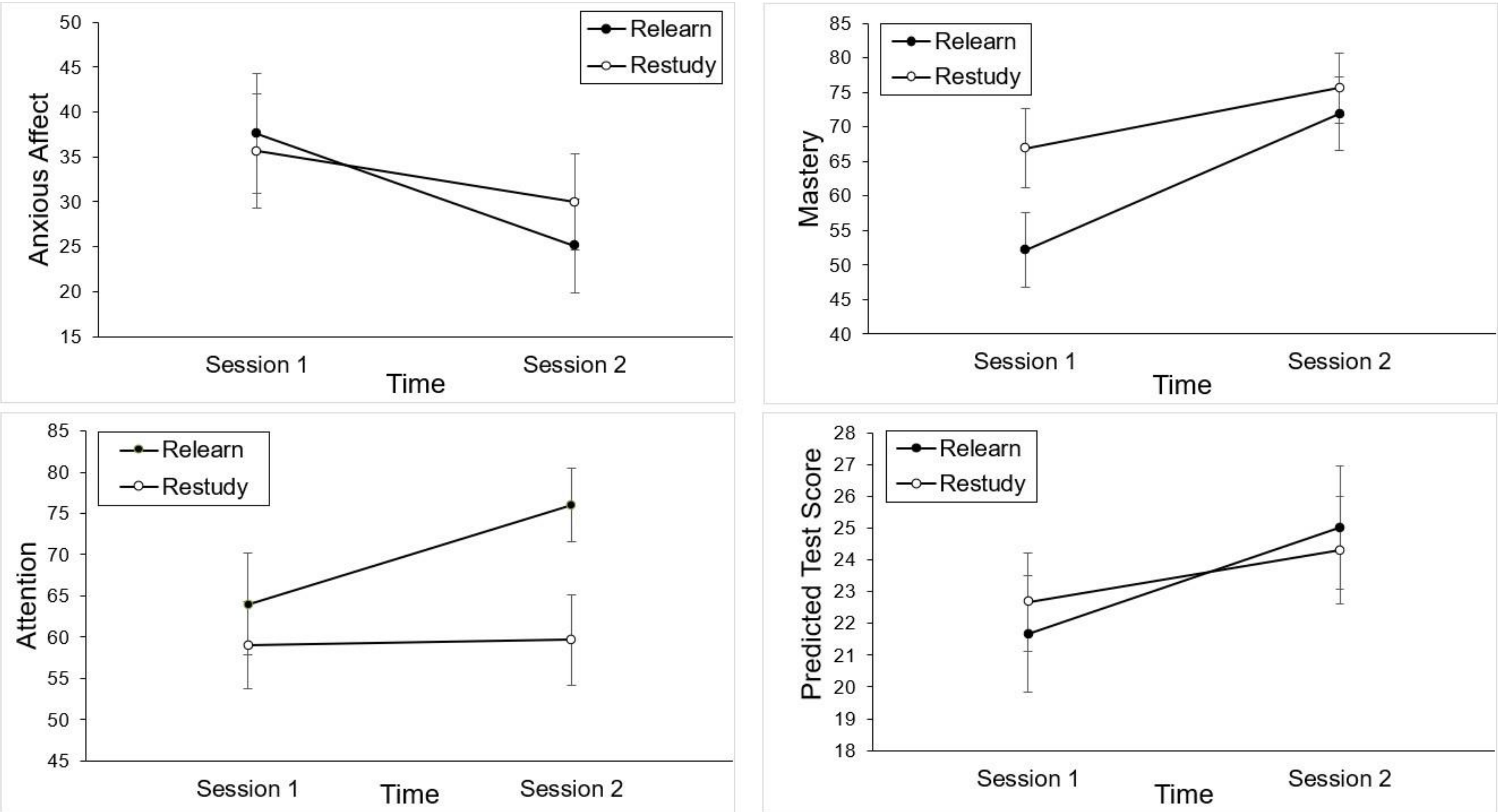
2.3.2.4.3 Attention

There was a statistically significant interaction between group and time on reported attention ($F(1,30) = 8.385$, $p = .007$; partial $\eta^2 = .218$). Planned comparisons showed that the difference between groups was not significant for session 1 ($F(1,30) < 1$, $p = .545$; partial $\eta^2 = .012$); *M*(SD) were 64.01(23.25) for RL students and 59.03(22.61) for RS students. The difference between groups was significant for session 2 ($F(1,30) = 4.96$, $p = .034$; partial $\eta^2 = .142$); *M*(SD) were 76.02(16.90) for RL students and 59.65(23.09) for RS students. Considering the effect of time, there was a statistically significant effect of time on reported attention for the RL group ($F(1,13) = 14.003$, $p = .002$; partial $\eta^2 = .519$). There was not a statistically significant effect of time on reported attention for the RS group ($F(1,17) = .067$, $p = .799$; partial $\eta^2 = .004$).

2.3.2.4.4 Predicted Test Score

The main effect of time showed a statistically significant increase in predicted test score from session 1 to session 2 ($F(1,30) = 17.966$, $p < .001$; partial $\eta^2 = .375$); *M*(SD) were 22.17(6.67) for session 1 and 24.65(7.14) for session 2.

Figure 10. *Within-Week Student Responses to Visual Analogue Scales (Anxious Affect, Mastery and Attention) and Predicted Test Score Immediately Following Intervention Sessions*



Note. Every scale reflects a judgement from 0-100. Error bars represent standard error. Session 1 refers to the first session students completed within a week, and session 2 refers to the final session the student completed within the same week. Where students had completed two sessions for more than one week, a mean value of the first and final sessions was taken.

2.3.3.2 Analysis of Test Data

Means and SDs for test data are shown in Table 6. Including outliers, there were no statistically significant differences in total test score between the experimental groups (RL, RS, C; $F(2,65) = 1.507$, $p = .229$; partial $\eta^2 = .044$); *M*(SD) were 23.16(6.78) in the RL group, 22.73(4.9) in the RS group and 20.26(6.47) in the control group. However, when the four outliers were removed (RS=3, C=1), the difference in total test score between groups was significant (Welch's $F(2, 35.841) = 3.320$, $p = .048$; partial $\eta^2 = .095$). Data are presented as *M*(SD). There was an increase in test score from 19.55(5.61) in the control group to 23.13(3.68) in the RS group to 23.16(6.78) in the RL group, in that order. Tukey post hoc analysis revealed that the increase from the control group to the RS group (3.59, 95% CI = [.11, 7.06]) was significant ($p = .042$), but no other group differences were statistically significant. Considering the impact of the intervention on exact, near-transfer and far-transfer question scores using a MANOVA, there was not a statistically significant difference between groups on the combined dependent variables (exact questions, near-transfer questions, far-transfer questions ($F(6,126) = .728$, $p = .628$, $\Lambda = .934$; partial $\eta^2 = .033$).

As shown in the correlation matrix in Table 7, there was significant positive correlation between the number of intervention sessions completed (0-12) and total test score, across the RL and RS groups ($r(88) = .54$, $p < .01$). Considering the correlations for the RL and RS groups separately, test score correlated with number of RL sessions completed ($r(42) = .66$, $p < .001$) and number of RS sessions completed ($r(46) = .38$, $p = .008$). Considering only students in RL and RS groups who completed at least one session, the correlation between number of sessions completed and test score was significant in the RL group ($r(18) = .76$, $p < .001$) but was not significant in the RS group ($r(25) = .32$, $p = .114$).

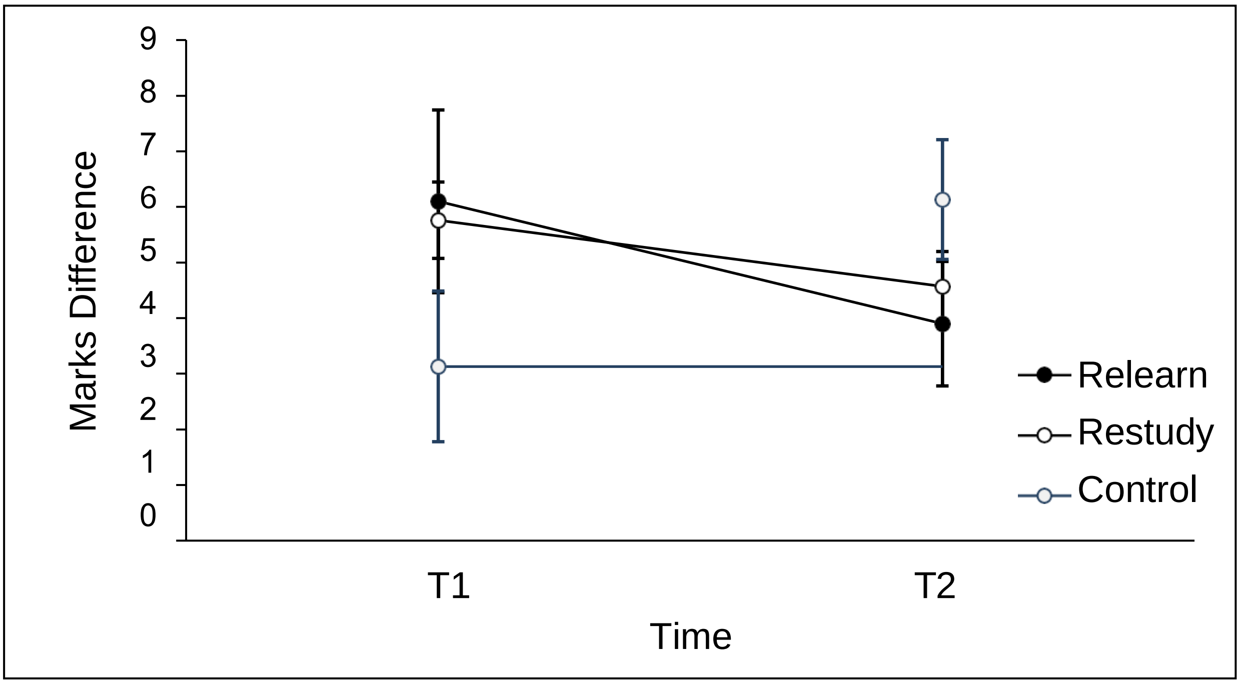
2.3.3.3 Analysis of Questionnaire Data

Means and SDs for questionnaire data are in Table 6. There was no statistically significant interaction between group (RL, RS, C) and time (T1, T2) on any questionnaire measure (attitudes towards maths, academic efficacy, attentional control, general anxiety, maths anxiety and predicted test score; (in all cases $F < 1.1$, $p > .1$). Only significant main effects are reported, in all other main effects $F < 1$, $p > .1$. Data is *M*(SD) unless stated otherwise. The main effect of group on predicted test score was significant ($F(2,41) = 5.326$, $p = .009$; partial $\eta^2 = .206$); predicted test score (across T1 and T2) increased from 17.05(3.3) in the control group to 21.30(5.15) in the RL group to 21.81(3.63) in the RS group.

There was a statistically significant interaction between group and time on predicted test score accuracy ($F(2,36) = 3.337$, $p = .047$; partial $\eta^2 = .156$; see Figure 11). At T1, the number of marks difference between predicted and attained test score was 3.13(3.83) in the control group, 5.76(3.14) in the RS group and 6.10(5.20) in the RL group, indicating that at T1, the control group were most accurate in their predicted test score, and the RL group were the least accurate. At T2, the number of marks difference between predicted and attained test score was 3.9(3.54) in the RL group, 4.57(2.87) in the RS group and 6.13(3.04) in the control group, indicating that the RL group were the most accurate in their predicted scores and the control group were the least accurate (a reversal of T1).

Figure 11

A Graph Showing That the Relearn and Restudy Groups Became More Accurate in Their Predicted Test Scores Over Time (Fewer Marks Difference Between Predicted and Attained Test Score at T2 than at T1). The Control Group Became Less Accurate Over Time (Greater Difference Between Predicted and Attained Scores at T2 Than at T1)



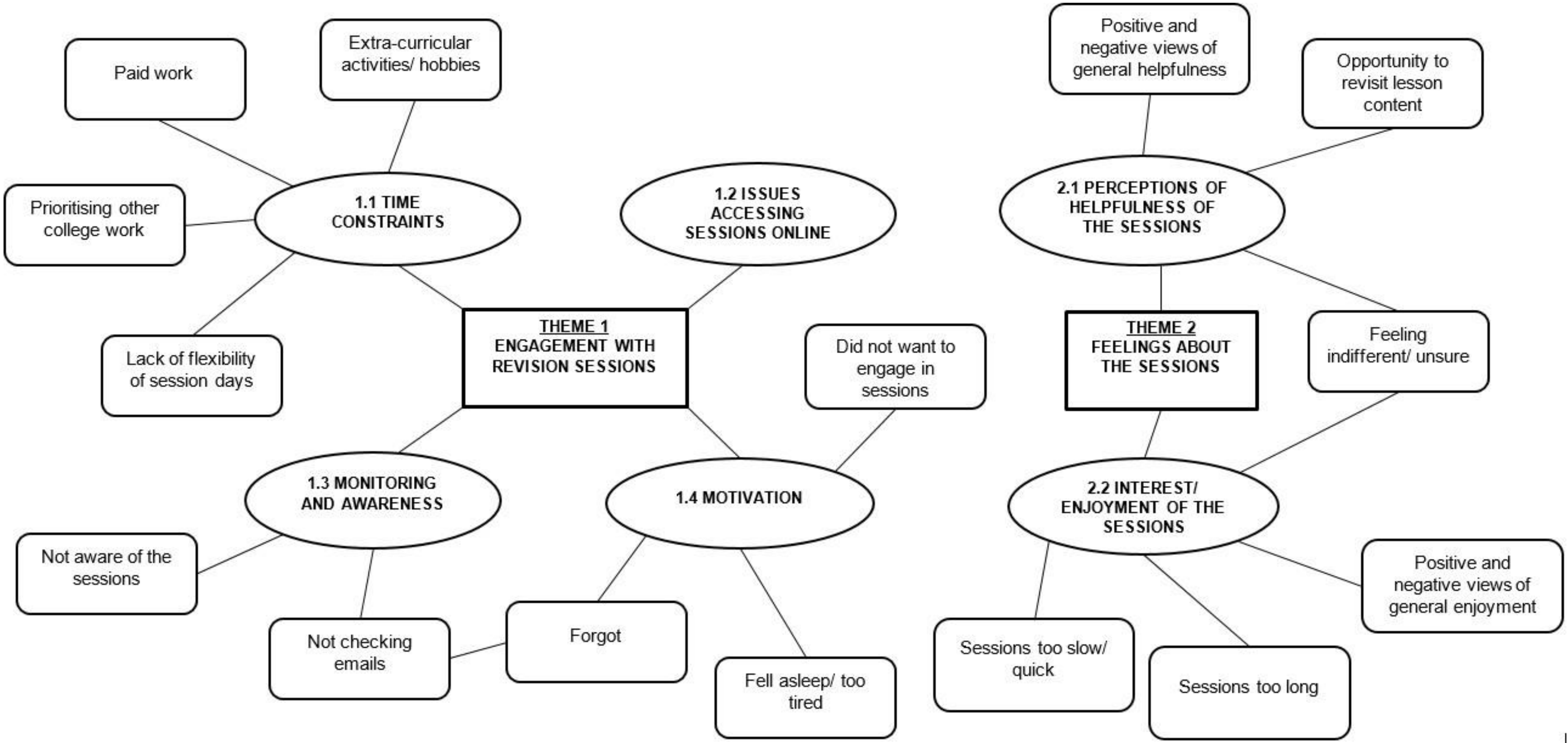
Note. T1 = time 1 (pre-intervention), T2 = time 2 (post-intervention). For each participant, at each time point, we calculated the number of marks difference between their predicted test score and the score that they attained in the maths test at T2. The plotted points represent the average marks difference for each group.

2.3.4 Analysis of End-User Feedback

Analysis of student responses was conducted across groups (RL, RS). All responses were included, regardless of whether students engaged in sessions. Thematic analysis revealed two themes: engagement with revision sessions and feelings about the sessions. There were several sub-themes and subordinate sub-themes, as shown in the final thematic map in Figure 12. Descriptions of sub-themes and subordinate sub-themes, alongside example quotes, are in the coding manual in Appendix V.

Figure 12

The Final Thematic Map Showing the Themes, Sub-Themes and Subordinate Sub-Themes of Students' Experiences of the RL and RS Interventions



2.3.4.1 Engagement with Revision Sessions

This theme referred to factors that decreased engagement in intervention sessions. Four sub-themes were identified within this theme: time constraints, issues accessing sessions online, monitoring and awareness, and motivation.

2.3.4.1.1 Time Constraints

Students frequently reported that they did not engage in sessions due to a lack of time because of other demands or commitments. This sub-theme included four subordinate sub-themes that represent these demands/commitments: extra-curricular activities/hobbies, paid work, prioritising other college work and lack of flexibility of session days. Students referenced extra-curricular activities/ hobbies as a reason for not completing sessions, e.g., “I was attending cadets”. Students also referenced paid work as a time constraint that impacted upon their engagement, e.g., “work straight after college so was unable to find the time to complete them.” Similarly, prioritising other college work, including homework and coursework, was a common theme, e.g., “too busy with assignments.” The final subordinate sub-theme related to lack of flexibility of session days. This included all references to students not being available on the particular days on which revision sessions were required to be completed (Monday, Wednesday and Friday), e.g., “not set on great days”. Some students referenced specific reasons for this, such as a regular event, e.g., visiting a family member, a special occasion such as a birthday, or an unprecedented event such as being unwell.

2.3.4.1.2 Issues Accessing Sessions Online

Several students cited difficulties accessing the sessions online as a barrier to engagement. Difficulties were described both with the hardware used to access the internet, e.g., “I often lose my phone” and also with the internet or program (software), e.g., “when I tried to log onto it my internet stopped working”.

2.3.4.1.3 Monitoring and Awareness

Several students reported that they were unaware of the sessions or they did not check their emails. This sub-theme included two subordinate subthemes: not aware of sessions, e.g., “I was not informed of these revision sessions”, and not checking emails, e.g., “I don’t check my emails and when I do the emails got pushed down to the bottom”.

2.3.4.1.4 Motivation

The final sub-theme within the first theme refers to low motivation. This included three subordinate sub-themes: fell asleep/too tired, e.g., “went to sleep”, did not want to engage, e.g., “I didn’t want to do them”, and forgot, e.g., “just forgot”. Forgot was linked to the subordinate sub-theme of not checking emails within the monitoring and awareness sub-theme.

2.3.4.2 Feelings About the Sessions

This theme includes two sub-themes: perceptions of the helpfulness of the sessions and interest/enjoyment of the sessions.

2.3.4.2.1 Perceptions of the Helpfulness of the Sessions

This sub-theme included any reference to the usefulness of the sessions and includes three subordinate sub-themes: positive and negative views of general helpfulness, opportunity to re-visit learning content, and feeling indifferent/unsure. Students reported varied feelings about the sessions being generally helpful, both positive, e.g., “very helpful”, and negative, e.g., “kinda pointless”. Some students reported that the sessions provided an opportunity to revisit lesson content and basic concepts, e.g., “I think they are useful as you get to revise certain topics and practice them.” Some students reported feeling indifferent/unsure about the sessions both generally and concerning their helpfulness, e.g., “I still don’t think I’m that much better at maths than I was before I started revising them”.

2.3.4.2.2 Interest/Enjoyment of the Sessions

The second sub-theme within the second theme included any reference to students’ interest or enjoyment of sessions. This included four subordinate subthemes: positive and negative views of general enjoyment, sessions too long, sessions too slow/too quick and feeling indifferent/unsure about sessions. Relating to their interest or enjoyment of sessions, students reported both positive views, e.g., “they were really fun and interesting”, and negative views, e.g., “boring”. Several students reported finding the sessions too long, e.g., “they took to long to do”, or disliked the session pace, either finding it too slow, e.g., “the first one I took was very slow and made it frustrating to complete”, or too quick, e.g., “some took too long to complete in time.” Some students reported feeling indifferent or unsure about the sessions, e.g., “meh, they were okay.”

2.4 Discussion

The present study investigated the impact of a successive relearning (RL) intervention on maths performance and motivational and affective outcomes in 16-18-year-olds who were re-sitting GCSE Maths. Engagement in both RL and RS sessions was low. Across groups, students who reported more positive attitudes towards maths and higher maths anxiety at T1 completed more sessions (0-12). Considering groups separately, students who predicted that they would achieve higher on the maths test completed more RL sessions, and students who reported more positive attitudes towards maths and higher maths anxiety completed more RS sessions. Students who engaged (i.e., completed at least one RL or RS session) reported greater academic efficacy at T1 (vs. students who did not engage). Furthermore, the difference between students who did and did not engage approached significance for attitudes towards maths (at T1) and number of GCSE Maths resits taken; indicating that, descriptively, students who reported more positive attitudes towards maths, and those who had taken fewer GCSE Maths re-sits, were more likely to complete at least one RL or RS session.

Previous research has identified the enhanced recall of participants who engaged in RL (vs. RS or no intervention; e.g., Higham et al., manuscript in preparation; Rawson et al., 2013). In the current study, there was evidence for the learning benefits of both the RL and RS interventions. The analysis showed that students who engaged in RL scored highest in the maths test, followed by students who engaged in RS, and students in the control group scored the lowest. The differences between groups were significant; however, test scores in the RL and RS groups were extremely similar. Furthermore, post-hoc tests revealed that only the increase from the control group to RS was significant. Further analysis identified a positive correlation between test score and number of RL sessions ($r = .66$, $p < .001$) and RS sessions ($r = .38$, $p = .008$) completed. Additionally, when only students who engaged in at least one session were included in this correlational analysis, completing a greater number of sessions was associated with increased test score in the RL group ($r = .76$, $p < .001$) but not in the RS group ($r = .32$, $p = .114$).

The analysis of state measures fits with Higham et al., who identified an increase in predicted test score and a reduction in anxiety in both RL and RS groups. However, the within-week change reported by Higham et al. was more substantial for RL (vs. RS) students. Like in Higham et al., there was a greater increase in reported mastery and attention in students who engaged in RL (vs. RS). This fits with Szpunar et al.'s (2013) finding that interpolated testing (vs. interpolated restudying) is associated with increased concentration.

Overall, the analysis of state measures suggests that engaging in an RL or RS intervention may lead to a decrease in anxiety and an increase in confidence (as measured by predicted test score). It also suggests that engaging in an RL intervention, in particular, may lead to increased feelings of mastery and attention (vs. RS).

The analysis of trait-like measures showed that RL or RS did not predict improvements in attitudes towards maths, academic self-efficacy, attentional control or anxiety. There was also no correlation between test score and trait variables at T1. This was unexpected given previous research that has identified an association between maths performance and motivational and affective factors in adolescents, including attitude towards maths (Hemmings & Kay, 2010), self-efficacy (Tosto et al., 2016) and maths anxiety (Devine et al., 2012). Moreover, previous research has identified that practice testing (vs. restudying) led to more accurate predictions of learning (Hughes et al., 2018). However, in the current study, students who engaged in RL or RS became more accurate in their predicted test score from T1 to T2, whereas students in the control group became less accurate, suggesting that both RL and RS may lead to improvements in students' awareness of their knowledge (metacognition).

Previous studies have identified gender differences in motivational and affective constructs relating to learning, e.g., Oppong Asante (2012) identified that male (vs. female) 16-21-year-old students had more positive attitudes towards maths. In the current study, male (vs. female) students reported more positive attitudes towards maths, higher academic efficacy and attentional control, and less anxiety. Males (vs. females) also predicted that they would achieve higher scores in the maths test. However, there were no gender differences in test score and females (vs. males) were more accurate in their predicted test score. Considering age, 17/18-year-olds had taken more GCSE Maths re-sits and had poorer attendance to maths lessons as compared to 16-year-olds. The difference between 16-year-olds and 17/18-year- olds additionally approached significance for attitudes towards maths, with 16-year-olds descriptively reporting more positive attitudes. This fits with Johnston-Wilder et al.'s (2015) postulation that continued GCSE Maths re-sits are associated with reduced motivation.

Considering student reflections on the intervention, reasons for not engaging in sessions included time constraints due to participation in extracurricular activities, paid work or other college work, and the inflexibility of session days. Other barriers included difficulty accessing sessions online, unawareness of sessions, not checking emails and low motivation.

Positive, negative and indifferent/unsure feelings about the helpfulness and interest and enjoyment relating to sessions were reported. Some students additionally felt that the session pace was too quick or too slow, or that the sessions were too long.

2.4.1 Limitations and Directions for Future Research

As far as we are aware, this study was the first to examine whether the benefits of successive relearning, as identified in previous studies, may extend to 16-18-year-olds who are re-sitting GCSE Maths. The findings provide preliminary evidence that engaging in RL or RS may be useful for adolescents re-sitting GCSE Maths who are learning key concepts and definitions. However, there were several limitations to the current study. Firstly, similar to Barzagar Nazari and Ebersbach (2018), a fundamental limitation was the very high level of attrition. Because of this, these findings reflect a cohort of students who, compared to their peers, reported significantly higher academic efficacy, reported descriptively, more positive attitudes towards maths and had taken fewer GCSE Maths re-sits. The high level of attrition also meant that there were fewer students within the final sample than recommended using power analysis (analysis using G*Power indicated that 53 participants per group were needed to detect a medium-sized effect when $\alpha = .05$ and power = .80). A further limitation of the current study is the lack of a baseline (T1) maths test. Unfortunately, this was not possible in the setting due to time pressures to deliver GCSE Maths content and the impact on this. However, all students had achieved a grade 3 in their GCSE Maths at the end of KS4.

Low engagement in the current study may be due to the population of students who were re-sitting their GCSE Maths as a mandatory requirement. Students re-sitting exams may lack motivation, confidence and self-efficacy (Smith, 2017; Johnston-Wilder et al., 2015). All of the students in the current study had not achieved a pass grade in their previous GCSE Maths exams and had therefore experienced failure. Some students had failed these exams multiple times. This may have led to feelings of incompetence and lack of autonomy, which are thought to impact upon motivation and behaviour (Deci & Ryan, 2012).

We considered student reporting of T1 measures in the current study, as compared to participant reporting in other studies. Data are means. In the current study, students descriptively (no analysis was undertaken) reported more negative attitudes towards mathematics, as measured by the ATMI (110.40), than has been reported in studies with general secondary-school student populations (Oppong Asante, 2012, 154.39 for males, 137.01 for females; Mirza & Hussain, 2018, 124.2). This is in line with research that has identified that students who are working at lower grades or below their target grade reported more negative attitudes towards maths (Mirza & Hussain, 2018; Tapia & Marsh, 2004). The current population also reported an academic efficacy score, as measured using the PALS (15.52), below participants in other studies (Midgley et al., 2000, 21.44; Tyler et al., 2016, 18.51; Reed et al., 2015, 18.55 for 6th-graders, 17.4 for 9th-graders). Moreover, in the current study, reported general anxiety, as measured using the YAM-5 (9.55), was higher than reported in previous community samples (Garcia-Lopez et al., 2017, 6.55; Simon et al., 2017, 3.61). However, unexpectedly, maths anxiety, as measured by the AMAS (20.93), was similar to that reported in other studies (Cipora et al., 2018, 20.96; Hopko et al., 2003, 21.1).

Many students who re-sit GCSE Maths do not achieve a pass during further education. This is a barrier to accessing future courses and jobs. More research is needed to explore learning techniques that may benefit this population, and particularly how to engage all students, including those with lower academic-efficacy, poorer attitudes towards maths, and those who have experienced more unsuccessful re-sits. Conducting qualitative research to gather these students' views relating to GCSE Maths and learning techniques is likely to be helpful to explore barriers to engaging in revision interventions and to consider how students could be further supported.

2.4.2 Conclusions and Implications for Practice

The findings from this study must be interpreted with caution due to high attrition. However, students who engaged in both RL or RS demonstrated better learning (maths test score, when outliers were removed), as compared to a no-intervention control group. Furthermore, considering only students who completed at least one RL or RS session, the correlation between number of RL sessions completed and test score was significant, however the correlation between number of RS sessions completed and test score was not. This provides a stronger indication of the potential value of successive relearning. There was no evidence that RL or RS led to improvements in attitudes towards maths, academic efficacy, attentional control, general anxiety or maths anxiety. However, students who engaged in RL or RS experienced improvements in metacognition as compared to a control group. There was also evidence that RL and RS led to a within-week decrease in anxiety and increase in confidence (as measured by predicted test score), and RL (vs. RS) led to greater improvements in reported mastery of session content and attention.

Engagement in sessions was low. Students with low academic efficacy were significantly less likely to engage, as were, descriptively but not significantly, students with more negative attitudes towards maths and those who had taken more GCSE Maths re-sits.

Students reported several reasons for low engagement, including insufficient time due to other commitments and responsibilities, difficulty accessing sessions online, unawareness of sessions and low motivation. These barriers suggest that students did not view completing sessions as a priority.

Although RL and RS were associated with a higher test score (vs. C), due to low engagement, the interventions' usefulness for the target population as a whole is questionable in the format as used in this study. It may, therefore, be useful to explore how these techniques may be used in a more supported and prescriptive way during lessons. However, although this may raise engagement in these techniques, in Further Education, there is often a move towards adult models of learning whereby students are expected to be increasingly internally motivated and self-directed (Knowles, 1968). Therefore, it may also be useful to consider modifications to the intervention itself, e.g., using gamification which may increase engagement (e.g., Buckley & Doyle, 2016). It may also be helpful to address reported barriers to engagement, such as by increasing the flexibility of session days, e.g., so that students can complete the first session of the week on Monday or Tuesday rather than Monday only. It may also be helpful for students to be persuaded to view sessions as worthwhile, e.g., by presenting the evidence that it can be helpful. Other reported barriers indicate that it may be useful to support students to use offline RL methods, e.g., flashcards, or allowing students greater control over the session pace and length. However, further research would be needed to establish whether these adjustments raise engagement and lead to learning benefits. Furthermore, as higher academic efficacy was associated with engagement, raising this may support students to engage in independent revision, especially in effortful techniques that involve testing oneself- and therefore risk perceived failure.

Appendix A Example Search Strategy – PsycINFO and PsycARTICLES

S29. S16 AND S21 AND S28

S28. S22 OR S23 OR S24 OR S25 OR S26 OR S27

S27. Mark

S26. Score

S25. Grade

S24. Achievement

S23. Attainment

S22. Performance

S21. S17 OR S18 OR S19 OR S20

S20. Statistics

S19. Numeracy

S18. Arithmetic

S17. Math*

S16. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15

S15. "Self-assessment"

S14. "Under-confiden*"

S13. "Over-confiden*"

S12. Overconfiden*

S11. "Meta-cognitive monitoring"

S10. “Metacognitive monitoring”

S9. “Meta-knowledge”

S8. Metaknowledge

S7. “Meta-comprehen*”

S6. Metacomprehen*

S5. “Meta-memor*”

S4. Metamemor*

S3. “Judgment* learn*”

S2. “Meta-cogniti*”

S1. Metacogniti*

Appendix B Borderline Cases at Secondary Screening

Table A1

The Studies Which Were Disputed For Inclusion by the Two Researchers at the Stage of Secondary Screening. The Inclusion of These Studies was Resolved After Discussion With a Third Researcher

Study	Researcher 1	Researcher 2	Final decision following discussion with researcher 3 (reason)
Dover & Shore (1991)	Exclude	Include	Exclude (group design which investigates the differences between gifted and average ability students and not the association between metacognition and maths performance).
Rosenzweig et al., (2011)	Exclude	Include	Exclude (group design which investigates the differences between learning disabled, low achieving and average achieving students and not the association between metacognition and maths performance).

Appendix C Records Excluded at Secondary Screening

Following screening of titles and abstracts, 115 papers were identified as relevant. After reading the full text of these 115 papers during the secondary screening stage, a further 84 papers were excluded for the following reasons:

- Does not report the strength of association between metacognition and maths performance (n = 32).
- Participant ages outside range for inclusion (n = 17).
- No measure of metacognition (n = 14).
- Does not separate metacognition from self-regulation (n = 3).
- Metacognition not reported pre-intervention or in a control group (in an intervention study) (n = 4)
- No measure of maths performance (n = 7).
- Unable to access the paper (after searching available databases and online, contacting the author and requesting the paper through the University of Southampton Inter-Library Loan Service) (n = 4).
- Uses the same data as another included study (n = 3).

Please see next page for a list of all excluded studies.

Table A2

Studies Excluded at the Secondary Screening Stage and Reasons for Exclusions

Study	Reason for exclusion
Abdollah, M. A, Kadivar, P., & Abdollahi, M. H. (2006). Relationships between cognitive styles, cognitive and meta-cognitive strategies with academic achievement, *Psychological Research*, *8*(3-4), 30-44 (in Persian).	Does not report the strength of association between metacognition and maths performance.
Areepattamannil, S. (2014). International note: What factors are associated with reading, mathematics, and science literacy of Indian adolescents? A multilevel examination. *Journal of Adolescence*, *37*(4), 367-372. https://doi.org/10.1016/j.adolescence.2014.02.007	Does not report the strength of association between metacognition and maths performance.
Areepattamannil, S., & Caleon, I. S. (2013). Relationships of cognitive and metacognitive learning strategies to mathematics achievement in four high-performing East Asian education systems. *The Journal of Genetic Psychology*, *174*(6), 696-702. https://doi.org/10.1080/00221325.2013.799057	Does not report the strength of association between metacognition and maths performance.
Aydın, U., & Ubuz, B. (2010). Structural model of metacognition and knowledge of geometry. *Learning and Individual Differences*, *20*(5), 436-445. https://doi.org/10.1016/j.lindif.2010.06.002	Participant ages outside range for inclusion (17-18 years).
Bae, Y. (2014). *The relationships among motivation, self-regulated learning, and academic achievement* [Doctoral dissertation, Texas A & M University]. Texas A & M Digital Collections. https://oaktrust.library.tamu.edu/handle/1969.1/153862	Participant ages outside range for inclusion (43.2% participants are outside age range- 16-18 years, and unable to contact author for middle-school data only as no correspondence address).

Study	Reason for exclusion
Baliram, N. S. (2016). *Reflective assessment, feedback and academic achievement in high school mathematics* [Doctoral dissertation, Seattle Pacific University].	No measure of metacognition.
Baliram, N., & Ellis, A. K. (2019). The impact of metacognitive practice and teacher feedback on academic achievement in mathematics. *School Science and Mathematics*, *119*(2), 94-104. **https://doi.org/10.1111/ssm.12317**	No measure of metacognition.
Bond, J. B., & Ellis, A. K. (2013). The effects of metacognitive reflective assessment on fifth and sixth graders' mathematics achievement. *School Science and Mathematics*, *113*(5), 227-234. **https://doi.org/10.1111/ssm.12021**	No measure of metacognition.
Bongiovani, M. B. (1985). *The effect of grade level, achievement and type of task on metacognitive awareness in elementary mathematics* [Doctoral dissertation, University of Maryland]. University of Maryland Digital Collections. https://drum.lib.umd.edu/handle/1903/25282	Does not report the strength of association between metacognition and maths performance.
Callan, G. L., Marchant, G. J., Finch, W. H., & Flegge, L. (2017). Student and school SES, gender, strategy use, and achievement. *Psychology in the Schools*, *54*(9), 1106-1122. **https://doi.org/10.1002/pits.22049**	Uses the same data as another included study (Callan, Marchant, Finch & German, 2016)
Cheng, H., Wu, G., Gao, L., Zhu, J., & Liu, D. (2014). Research on status and characteristics of the mastery of students' mathematical learning strategies in Chinese junior high schools. *Procedia-Social and Behavioral Sciences*, *116*, 3218-3225. https://doi.org/10.1016/j.sbspro.2014.01.738	No measure of maths performance
Chin, E., Lin, Y., Chuang, C., & Tuan, H. (2007, July 8-13). *The influence of inquiry-based mathematics teaching on 11th grade*	Participant ages outside range for inclusion (age

Study	Reason for exclusion
high achievers: Focusing on metacognition. Psychology of Mathematics Education (PME) conference, Seoul.	not reported but due to stage of schooling expected to be 17 years).
Cooper, F. (2008). *An examination of the impact of multiple intelligences and metacognition on the achievement of mathematics students*. [Doctoral dissertation, Capella University].	Does not report the strength of association between metacognition and maths performance.
Creighton-Lacroix, W. D. (2002). *The self-regulation of test anxiety using metacognitive strategy instruction.* [Doctoral dissertation, The University of Alberta].	No measure of metacognition.
Dennis, J. A. (1987). *The influence of cognitive monitoring instruction on academic task performance.* [Doctoral dissertation, University of Southern California].	Does not report the strength of association between metacognition and maths performance.
Deotto, A., Westmacott, R., Fuentes, A., de Veber, G., & Desrocher, M. (2019). Does stroke impair academic achievement in children? The role of metacognition in math and spelling outcomes following pediatric stroke. *Journal of Clinical and Experimental Neuropsychology*, *41*(3), 257-269. https://doi.org/10.1080/13803395.2018.1533528	Participant ages outside range for inclusion (mean age = 9.5 years).
Dermitzaki, I., & Efklides, A. (2002). The structure of cognitive and affective factors related to students' cognitive performance in language and maths. *Psychology: The Journal of the Hellenic Psychological Society*, *9*, 58-74.	Unable to access paper (authors emailed via correspondence address twice and no reply and

Study	Reason for exclusion
	paper not available via the University of Southampton Inter-Library Loan Service).
Digiacomo, G., & Chen, P. P. (2016). Enhancing self-regulatory skills through an intervention embedded in a middle school mathematics curriculum. *Psychology in the Schools*, *53*(6), 601-616. https://doi.org/10.1002/pits.21929	Does not report the strength of association between metacognition and maths performance.
Dina, F., & Efklides, A. (2009). Student profiles of achievement goals, goal instructions and external feedback: Their effect on mathematical task performance and affect. *European Journal of Education and Psychology*, *2*(3), 235-262. https://www.redalyc.org/pdf/1293/129312574006.pdf	Does not report the strength of association between metacognition and maths performance.
Dover, A., & Shore, B. M. (1991). Giftedness and flexibility on a mathematical set-breaking task. *Gifted Child Quarterly*, *35*(2), 99-105. https://doi.org/10.1177%2F001698629103500209	Does not report the strength of association between metacognition and maths performance (excluded following discussion with third researcher).
Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research*, *50*(4), 241-250. https://doi.org/10.1016/j.ijer.2011.08.005	Does not report the strength of association between metacognition and maths performance.
Edwards, T. G. (2008). *Reflective assessment and mathematics achievement by secondary at-risk students in an alternative*	No measure of metacognition.

Study	Reason for exclusion
secondary school setting. [Doctoral dissertation, Seattle Pacific University]. ProQuest Digital Collections https://search.proquest.com/docview/304801373?accountid=13963	
Efklides, A., & Tsiora, A. (2002). Metacognitive experiences, self-concept, and self-regulation. *Psychologia*, *45*(4), 222-236. https://doi.org/10.2117/psysoc.2002.222	Participant ages outside range for inclusion.
Erbas, A. K., & Bas, S. (2015). The contribution of personality traits, motivation, academic risk-taking and metacognition to the creative ability in mathematics. *Creativity Research Journal*, *27*(4), 299-307. https://doi.org/10.1080/10400419.2015.1087235	No measure of maths performance (measure mathematical creativity but report in the paper that this is not predicted by, or predictive of, mathematical academic ability).
Fernández, T. G., Kroesbergen, E., Pérez, C. R., González-Castro, P., & Gonzalez-Pienda, J. A. (2015). Factors involved in making post-performance judgments in mathematics problem-solving. *Psicothema*, *27*(4), 374-380.	Does not report the strength of association between metacognition and maths performance.
Franz (2006). *Knowing about not knowing: A cognitive view of mathematics anxiety.* [Doctoral dissertation, McGill University].	No measure of metacognition.
García Fernández, T., Fernández Cueli, M. S., Rodríguez Pérez, C., Krawec, J., & González Castro, M. P. (2015). Metacognitive knowledge and skills in students with deep approach to learning. Evidence from mathematical problem solving. *Revista de Psicodidáctica*, *20*(2), 209-226. http://digibuo.uniovi.es/dspace/bitstream/10651/34288/1/13060-52168-2-PB.pdf	No measure of maths performance.

Study	Reason for exclusion
Grant, G. (2014). *A metacognitive-based tutoring program to improve mathematical abilities of rural high school students: An action research study* [Doctoral dissertation, Capella University]. ProQuest Digital Collections https://search.proquest.com/docview/1626379373?accountid=13963	No measure of metacognition.
Hejazi, E., Rastgar, A., Gholamai, L. M., & Ghorban, J. R. (2009). Intelligence beliefs and academic achievement: Mediating role of academic goals and academic engagement. *Psychological Research, 12*(1-2). (in Arabic)	No measure of maths performance.
Hong, E. (1995). A structural comparison between state and trait self-regulation models. *Applied Cognitive Psychology*, *9*(4), 333-349. https://doi.org/10.1002/acp.2350090406	Does not separate metacognition from self-regulation (also includes test anxiety in measure).
Hong, E., Peng, Y., & Rowell, L. L. (2009). Homework self-regulation: Grade, gender, and achievement-level differences. *Learning and Individual Differences*, *19*(2), 269-276. https://doi.org/10.1016/j.lindif.2008.11.009	Does not separate metacognition from self-regulation.
Huang (2011). A survey and research on the relationship between number sense and metacognition of 6th and 8th students.	Unable to access the paper (no correspondence email address, requested via University of Southampton Inter-Library Loan Service and advised that the thesis is unavailable to purchase, and the Service staff cannot identify any libraries which hold it).

Study	Reason for exclusion
Human, P. W. (1992). *The effects of process journal writing on learning in mathematics: A study of metacognitive processes* [Doctoral dissertation, East Texas State University].	Metacognition not reported pre-intervention or in a control group (in an intervention study).
Humphrey, M. (2008). *Mathematical word problem solving performance of students in an alternative setting: Investigating contributing factors and intervention* [Doctoral dissertation, University of Northern Colorado].	Participant ages outside range for inclusion (15-18 years).
Hutajulu, M. & Wahyudin, W. (2018, November). The effectiveness of metacognitive learning in enhancing student's mathematical analysis. In *Journal of Physics: Conference Series* (Vol. 1132, No. 1, p. 012041). IOP Publishing.	No measure of metacognition.
Kaur, B., & Areepattamannil, S. (2012, July 2-6). *Influences of metacognitive and self-regulated learning strategies for reading on mathematical literacy of adolescents in Australia and Singapore.* The 35th Annual Conference of the Mathematics Education Research Group of Australasia (MERGA 2012), Singapore.	Uses the same data as an included study (PISA, 2009 database)
Kaya, S. (2007). *The influences of student views related to mathematics and self-regulated learning on achievement of algebra I students* [Doctoral dissertation, The Ohio State University]. The Ohio State University Library Digital Collections http://rave.ohiolink.edu/etdc/view?acc_num=osu1185905759	Does not report the strength of association between metacognition and maths performance (reports association between metacognition with resource management strategies and algebra score, but not metacognition and algebra score separate of resource management strategies).

Study	Reason for exclusion
Kazemi, F., Yektayar, M., & Abad, A. M. B. (2012). Investigation the impact of chess play on developing meta-cognitive ability and math problem-solving power of students at different levels of education. *Procedia-Social and Behavioral Sciences*, *32*, 372-379. https://doi.org/10.1016/j.sbspro.2012.01.056	Metacognition not reported pre-intervention or in a control group (in an intervention study).
Kendall, W. W. (1991). *Imbedding metacognition in the math problem-solving curriculum: An intervention leading students to analyze their own errors and its impact* [Doctoral dissertation, Harvard University].	Metacognition not reported pre-intervention or in a control group (in an intervention study).
Kesici, S., Erdogan, A., & Özteke, H. I. (2011). Are the dimensions of metacognitive awareness differing in prediction of mathematics and geometry achievement? *Procedia-Social and Behavioral Sciences*, *15*, 2658-2662. https://doi.org/10.1016/j.sbspro.2011.04.165	Participant ages outside range for inclusion (students expected to be 15-17 years but unsure how many are over 16. Author emailed twice to clarify and no response).
Khabiri, P. S. (1993). *The role of metacognition, effort and worry in math problem solving requiring problem translation* [Doctoral dissertation, University of Southern California].	Participant ages outside range for inclusion (some participants are 12th graders and so are expected to be over 16. It is unclear how many participants this applies to. Author emailed twice to ask but no response).
Kikas, E., Mädamürk, K., & Palu, A. (2019). What role do comprehension-oriented learning strategies have in solving math calculation and word problems at the end of middle school? *British*	No measure of metacognition

Study	Reason for exclusion
Journal of Educational Psychology. [Epub ahead of print]. https://doi.org/10.1111/bjep.12308	
Kramarksi, B. (2000). Metacognition and the ability to solve mathematical problems presented in concrete and abstract contexts. *Megamwt*, *40*(4), 660-685.	Does not report the strength of association between metacognition and maths performance.
Kramarski, B., & Friedman, S. (2014). Solicited versus unsolicited metacognitive prompts for fostering mathematical problem solving using multimedia. *Journal of Educational Computing Research*, *50*(3), 285-314. https://doi.org/10.2190%2FEC.50.3.a	Metacognition not reported pre-intervention or in a control group (in an intervention study).
Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*(1), 281-310. https://doi.org/10.3102%2F00028312040001281	Does not report the strength of association between metacognition and maths performance (reported that both variables are measured pre-intervention, author emailed via correspondence address twice but no response).
Kramarski, B., Mevarech, Z. R., & Lieberman, A. (2001). Effects of multilevel versus unilevel metacognitive training on mathematical reasoning. *The Journal of Educational Research*, *94*(5), 292-300. https://doi.org/10.1080/00220670109598765	Does not report the strength of association between metacognition and maths performance (reported that both variables are measured pre-intervention, author emailed via correspondence address twice but no response).

Study	Reason for exclusion
Kramarski, B., & Ritkof, R. (2002). The effects of metacognition and email interactions on learning graphing. *Journal of Computer Assisted Learning*, *18*(1), 33-43. https://doi.org/10.1046/j.0266-4909.2001.00205.x	Does not report the strength of association between metacognition and maths performance.
Kramarski, B., & Zoldan, S. (2008). Using errors as springboards for enhancing mathematical reasoning with three metacognitive approaches. *The Journal of Educational Research*, *102*(2), 137-151. https://doi.org/10.3200/JOER.102.2.137-151	Does not report the strength of association between metacognition and maths performance (reported that both variables are measured pre-intervention, author emailed via correspondence address twice but no response).
Lai, Y., Zhu, X., Chen, Y., & Li, Y. (2015). Effects of mathematics anxiety and mathematical metacognition on word problem solving in children with and without mathematical learning difficulties. *PloS one*, *10*(6), e0130570. https://dx.doi.org/10.1371%2Fjournal.pone.0130570	Participant ages outside range for inclusion (10-year olds).
Legg, A. M., & Locker Jr, L. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology*, *11*(3), 471-485.	Participant ages outside range for inclusion (university undergraduate students).
LeMay, J. (2016). *The effects of using selected metacognitive strategies on ACT mathematics sub-test scores* [Doctoral dissertation, Liberty University].	No measure of metacognition.

Study	Reason for exclusion
Liggins, L. (2006). *Metacognition and the minority achievement gap: Differential effect of instructional methods as a function of ethnicity* [Doctoral dissertation, Illinois Institute of Technology].	Participant ages outside range for inclusion (6-8 years old).
Liu, X. (2008). The junior middle school students' mathematical estimation performance and its relationship with their metacognitive ability. *International Journal of Psychology*, *43*(3-4), 285.	No measure of maths performance.
Lucangeli, D., Cornoldi, C., & Tellarini, M. (1998). Metacognition and learning disabilities in mathematics. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities, Vol. 12* (pp. 219–244). Elsevier Science/JAI Press	Does not report the strength of association between metacognition and maths performance
Maqsud, M. (1998). Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers. *Educational Research*, *40*(2), 237-243. https://doi.org/10.1080/0013188980400210	Participant ages outside range for inclusion (15-17 years old).
McCallie, A. (2016). *The impact of self-reflection on sixth grade students in a mathematics course* [Doctoral dissertation, Trevecca Nazarene University]. ProQuest Digital Collections. https://search.proquest.com/docview/1821368845?accountid=13963	No measure of metacognition.
Mevarech, Z. R., & Amrany, C. (2008). Immediate and delayed effects of meta-cognitive instruction on regulation of cognition and mathematics achievement. *Metacognition and Learning*, *3*(2), 147-157.	Does not report the strength of association between metacognition and maths performance (reported that this was measured pre-intervention but not reported, authors emailed twice but no response).

Study	Reason for exclusion
Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology*, *73*(4), 449-471. https://doi.org/10.1348/000709903322591181	Does not report the strength of association between metacognition and maths performance.
Miranda, A., & Fortes, M. C. (1989). Aplicación de las técnicas cognitivo comportamentales a la resolución de problemas de matemáticas. *Revista de Psicología de la Educación*, *1*(2), 57-72.	No measure of metacognition
Montague, M., & Applegate, B. (1993). Mathematical problem-solving characteristics of middle school students with learning disabilities. *The Journal of Special Education*, *27*(2), 175-201. https://doi.org/10.1177%2F002246699302700203	Does not report the strength of association between metacognition and maths performance.
Montague, M., & Bos, C. S. (1990). Cognitive and metacognitive characteristics of eighth grade students' mathematical problem solving. *Learning and Individual Differences*, *2*(3), 371-388. https://doi.org/10.1016/1041-6080(90)90012-6	Does not report the strength of association between metacognition and maths performance.
Morosanova, V. I., Fomina, T. G., Kovas, Y., & Bogdanova, O. Y. (2016). Cognitive and regulatory characteristics and mathematical performance in high school students. *Personality and Individual Differences*, *90*, 177-186. https://doi.org/10.1016/j.paid.2015.10.034	Does not report metacognition separate from self-regulation.
Nelson, L. L. (2012). *The effectiveness of metacognitive strategies on 8th grade students in mathematical achievements and problem solving skills* [Doctoral dissertation, Southern University and Agricultural and Mechanical College]. ProQuest Digital Collections. https://search.proquest.com/docview/1015126655	Does not report the strength of association between metacognition and maths performance (reported that both

Study	Reason for exclusion
	variables are measured pre-intervention and there is a control group but unable to contact the authors as no correspondence address listed).
Okamoto, M., & Kitao, N. (1992). The role of metacognitive knowledge and aptitude in arithmetic problem solving. *Psychologia: An International Journal of Psychology in the Orient*, *35*(3), 164–172.	Unable to access the paper (no correspondence address, unavailable via University of Southampton Inter Library Loan Service).
O'Neal, L. M. (2015). *The effects of metacognitive writing on student achievement in advanced placement calculus* [Doctoral dissertation, Seattle Pacific University].	Participant ages outside range for inclusion (16-18-years old).
Onu, V. C., Eskay, M., Igbo, J. N., Obiyo, N., & Agbo, O. (2012). Effect of training in math metacognitive strategy on fractional achievement of Nigerian schoolchildren, *US-China Education Review*, *3*, 316-325. https://files.eric.ed.gov/fulltext/ED532899.pdf	No measure of metacognition.
Panaoura, A. (2007). The interplay of processing efficiency and working memory with the development of metacognitive performance in mathematics. *The Mathematics Enthusiast*, *4*(1), 31-52.	Participant ages outside range for inclusion.

Study	Reason for exclusion
Panaoura, A. & Philippou, G. (2007). The developmental change of young pupils' metacognitive ability in mathematics in relation to their cognitive abilities. *Cognitive Development*, *22*(2), 149-164. https://doi.org/10.1016/j.cogdev.2006.08.004	Participant ages outside range for inclusion (3rd to 5th grade).
Peklaj, C., & Pečjak, S. (2011). Emotions, motivation and self-regulation in boys' and girls' learning mathematics. *Horizons of Psychology*, *20*(3), 33-58.	No measure of maths performance.
Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics*, *20*(1), 18-41. https://doi.org/10.4219%2Fjaa-2008-869	Does not report the strength of association between metacognition and maths performance (maths performance is only reported as a covariate).
Ratner, B. (1987). *An integration of confidence assessment and the Rasch model in the area of aptitude* [Doctoral dissertation, The State University of New Jersey].	Unable to access the paper (no correspondence email address and not available via the University of Southampton Inter-Library Loan Service).
Riggs, R. M. (2012). *Can practice calibrating by test topic improve public school students' calibration accuracy and performance on tests?* [Doctoral dissertation, Old Dominion University]. https://doi.org/10.25777/pj9w-th02	Does not report the strength of association between metacognition and maths performance.
Rinne, L. F., & Mazzocco, M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PloS one*, *9*(7), e98663.	Does not report the strength of association between metacognition and maths performance.

Study	Reason for exclusion
Rosenzweig, C., Krawec, J., & Montague, M. (2011). Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: A think-aloud analysis. *Journal of Learning Disabilities*, *44*(6), 508-520. https://doi.org/10.1177%2F0022219410378445	Does not report the strength of association between metacognition and maths performance (following discussion with third researcher).
Rutherford, T. (2017). Within and between person associations of calibration and achievement. *Contemporary Educational Psychology*, *49*, 226-237. https://doi.org/10.1016/j.cedpsych.2017.03.001	Participant ages outside range for inclusion (2^{nd} – 5^{th} graders).
Saricam, H., & Ogurlu, Ü. (2015). Metacognitive awareness and math anxiety in gifted students. *Cypriot Journal of Educational Sciences*, *10*(4), 338-348. http://dx.doi.org/10.18844/cjes.v10i4.228	No measure of maths performance.
Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Developmental Psychology*, *49*(10), 1994-2004. https://psycnet.apa.org/doi/10.1037/a0031200	No measure of metacognition.
Slife, B. D., Weiss, J., & Bell, T. (1985). Separability of metacognition and cognition: Problem solving in learning disabled and regular students. *Journal of Educational Psychology*, *77*(4), 437-445. https://psycnet.apa.org/doi/10.1037/0022-0663.77.4.437	Does not report the strength of association between metacognition and maths performance (measured maths performance for purpose of

Study	Reason for exclusion
	matching participants only).
Sun-Lin, H. Z., & Chiou, G. F. (2017). Effects of self-explanation and game-reward on sixth graders' algebra variable learning. *Journal of Educational Technology & Society*, *20*(4), 126-137.	Does not report the strength of association between metacognition and maths performance.
Sweeney, C. M. (2010). *The metacognitive functioning of middle school students with and without learning disabilities during mathematical problem solving* [Doctoral dissertation, University of Miami].	Does not report the strength of association between metacognition and maths performance.
Yang, C. K. (2011, September 8-10). *How do students' problem solving strategies and preferences in learning environments relate to their mathematical performance? A comparative study between South Korea and the United States.* Society for Research on Educational Effectiveness conference, Washington D.C.	Does not report the strength of association between metacognition and maths performance.
Young, A. E. (2010). *Explorations of metacognition among academically talented middle and high school mathematics students* [Doctoral dissertation, UC Berkeley].	Uses the same data as another included study (Doctoral dissertation later published as Young & Worrell, 2018).
Zampieri, M., & Schelini, P. W. (2013). O uso de medidas intelectuais na análise do monitoramento metacognitivo de crianças.	Participant ages outside range for inclusion (mean = 10.4 years).

Study	Reason for exclusion
Psicologia: Teoria e Pesquisa, *29*(2), 177-183. https://doi.org/10.1590/S0102-37722013000200007	
Zhou, S., Han, C., & Chen, Y. (2012). The mediating role of metacognitive ability in mathematics between mathematics anxiety and mathematic academic achievement. *International Journal of Psychology*, *47*, 559-564.	Does not report the strength of association between metacognition and maths performance

Appendix D Quality Assessment Checklist (Adapted From the CASP Cohort Study Checklist, 2018)

Table A3

The Items Taken From the CASP Cohort Study Checklist (2018), Alongside how the Items Were Interpreted for the Current Review and What Evidence was Required in Order to Assign an Affirmative Response

Checklist item	Criteria to be rated as 'yes'
1) Did the study address a clearly focused issue?	There are clear stated research aims and a clear rationale for the study.
(2) Was the cohort recruited in an acceptable way?	Participants are acceptably representative of the cohort from which they are taken from, given the importance of informed consent e.g. a self-selecting sample from one school was acceptable.
(3) Was the exposure accurately measured to minimise bias?	The measure(s) of metacognition are valid and reliable.
(4) Was the outcome accurately measured to minimise bias?	The measure(s) of maths performance are valid and reliable.
(5a) Have the authors identified all important confounding factors?	The author(s) have listed/ described confounding factors of the relationship between metacognition and maths performance e.g. intelligence.
(5b) Have they taken account of the confounding factors in the design and/or analysis?	The author(s) have considered the impact of confounding factors e.g. by reporting semi-partial correlations between metacognition and maths performance, accounting for a confounding factor such as intelligence.

Checklist item	Criteria to be rated as 'yes'
(7) What are the results of this study?	N/A
(8) How precise are the results?	The study was considered to be sufficiently precise where results were reported with 95% confidence ($p < 0.5$) or greater.
(9) Do you believe the results?	The researcher believes that the method is robust and replicable.
(10) Can the results be applied to the local population?	The study participants and setting are appropriate to the current review population (11-16-year olds).
(11) Do the results of this study fit with other available evidence?	The results fit with other available evidence (e.g. of a positive correlation between metacognition and maths performance).
(12) What are the implications of this study for practice?	N/A

Appendix E Quality Assessment Checklist (adapted from the CASP Cohort study checklist 2018)

Table A4

The Final Responses to the Quality Assessment CASP Checklist for Each Included Study, as Agreed Upon by the two Researchers

Study	1. Addresses a focused issue	2. Participants acceptably recruited	3. Measure of metacognition	4. Measure of maths performance	5a. Confounding factors identified	5b. Confounding factors accounted for.	8. Precise.	9. Results believable.	10. Applicable to local population.	11. Fit with other evidence	
Ahmed et al., (2013)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9
Aşık & Erktin (2019)	Yes	Can’t tell	Yes	Yes	No	No	Sufficient	Yes	Yes	Yes	6
Bishara & Kaplan (2018)	Yes	Yes	Yes	Yes	No	No	Sufficient	Yes	Can’t tell .	Yes	6

Study	1. Addresses a focused issue	2. Participants acceptably recruited	3. Measure of metacognition	4. Measure of maths performance	5a. Confounding factors identified	5b. Confounding factors accounted for.	8. Precise.	9. Results believable.	10. Applicable to local population.	11. Fit with other evidence	
Callan & Cleary (2019)	Yes	Can’t tell	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	8
Callan et al., (2016)	Yes	Yes	Yes	Yes	Yes	No	Sufficient	Yes	Yes	Yes	8
Chiu et al., (2007)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9
Erktin (2004)	Yes	Can’t tell	Yes	Can’t tell	Yes	Yes	Sufficient	Yes	Yes	Yes	7
Fadlelmula et al., (2015)	Yes	Yes	Yes	Yes	No	No	Sufficient	Yes	Yes	Yes	7
Fitzpatrick (1994)	Yes	Yes	Can’t tell	Can’t tell	Yes	No	Sufficient	Yes	Yes	No	5

Study	1. Addresses a focused issue	2. Participants acceptably recruited	3. Measure of metacognition	4. Measure of maths performance	5a. Confounding factors identified	5b. Confounding factors accounted for.	8. Precise.	9. Results believable.	10. Applicable to local population.	11. Fit with other evidence	
Fusco (1995)	Yes	Yes	Yes	Yes	Can't tell	Can't tell	Sufficient	Yes	Yes	Yes	7
Harris (2015)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Can't tell	No	7
Hassan & Rahman (2017)	Yes	Can't tell	Yes	Can't tell	No	No	Sufficient	Can't tell	Can't tell	Unclear	2
Ichihara & Arai (2006)	Yes	Yes	Yes	Yes	No	No	Sufficient	Yes	Yes	Yes	7
Maras et al., (2019)	Yes	Yes	Yes	Yes	Yes	Can't tell	Sufficient	Yes	Yes	Yes	8
Martín et al., (2008)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9

Study	1. Addresses a focused issue	2. Participants acceptably recruited	3. Measure of metacognition	4. Measure of maths performance	5a. Confounding factors identified	5b. Confounding factors accounted for.	8. Precise.	9. Results believable.	10. Applicable to local population.	11. Fit with other evidence	
Ning (2016)	Yes	Yes	Yes	Yes	No	No	Sufficient	Yes	Yes	No	6
Özcan (2016)	Yes	Yes	Yes	Yes	No	No	Sufficient	Yes	Yes	Yes	7
Özcan & Eren Gümüş (2019)	Yes	Can't tell	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	8
Özsoy, G. (2011)	Yes	Can't tell	Yes	Yes	No	No	Sufficient	Yes	Yes	Yes	6
Peng et al., (2014)	Yes	Yes	No	Can't tell	Yes	Yes	Sufficient	Yes	Yes	Yes	6
Sink et al., (1991)	Yes	Yes	Yes	Yes	Yes	Can't tell	Sufficient	Yes	Yes	Yes	8
Tian et al., (2018)	Yes	Yes	Yes	Can't tell	Yes	Yes	Sufficient	Yes	Yes	Yes	8
van der Stel & Veenman (2014)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9

Study	1. Addresses a focused issue	2. Participants acceptably recruited	3. Measure of metacognition	4. Measure of maths performance	5a. Confounding factors identified	5b. Confounding factors accounted for.	8. Precise.	9. Results believable.	10. Applicable to local population.	11. Fit with other evidence	
Van der Stel et al., (2010)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9
Van der Walt et al., (2008)	Yes	Yes	Yes	Yes	No	No	Sufficient	Yes	Yes	Yes	7
Veenman et al., (2000)	Yes	Can't tell	Yes	Yes	No	No	Sufficient	Yes	Can't tell	Yes	5
Veenman et al., (2005)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9
Veenman & Spaans (2005)	Yes	No	Yes	Can't tell	Yes	Yes	Sufficient	Yes	Yes	Yes	7
Walker (2013)	Yes	Yes	Yes	Yes	Yes	Can't tell	Sufficient	Yes	Can't tell	Yes	7

Study	1. Addresses a focused issue	2. Participants acceptably recruited	3. Measure of metacognition	4. Measure of maths performance	5a. Confounding factors identified	5b. Confounding factors accounted for.	8. Precise.	9. Results believable.	10. Applicable to local population.	11. Fit with other evidence	
Yap (1993)	Yes	Yes	Yes	Yes	Yes	Yes	Sufficient	Yes	Yes	Yes	9
Young & Worrell (2018)	Yes	Yes	Yes	Yes	Yes	No	Sufficient	Yes	Can’t tell	Yes	7

Appendix F Study Aims and Other Findings

Table A5

The Stated Aims and Other Key Findings (in Addition to Findings Relating Directly to the Association Between Metacognition and Maths Performance) of Included Studies

	Study	Aim	Other Key Findings
1	Ahmed et al., (2013)	To explore how emotions influence information processing strategies	Over time, students' enjoyment and pride in mathematics declined and boredom increased. Anxiety remained relatively stable. Changes in positive emotions were positively associated with changes in self-regulated learning and achievement.
2	Aşık & Erktin (2019)	To investigate the mediating role of metacognitive experiences in the relationship between metacognitive knowledge and maths performance	Students' task-related metacognitive experiences have a significant mediating effect on the relationship between metacognitive knowledge and problem-solving performance

	Study	Aim	Other Key Findings
3	Bishara, & Kaplan (2018)	To investigate the correlation between locus of control and metacognitive knowledge in maths in students with learning disabilities	Internal locus of control was associated with greater use of metacognitive knowledge. Higher levels of internal locus of control and higher use of metacognitive knowledge resulted in increased math achievement. Students with learning disabilities were more likely to have an external locus of control than students without learning disabilities. As compared to students with learning disabilities, students without learning disabilities scored significantly higher on the maths test and the measure of metacognition.
4	Callan & Cleary (2019)	To investigate sequential phase relations between forethought, performance and self-reflection and the impact on maths performance	There was a significant positive correlation between students' strategic planning, strategy use, and metacognitive monitoring with mathematics performance.
5	Callan et al., (2016)	To compare, across countries, the relationship between metacognition, learning strategies, and achievement	There was a significant positive correlation between socioeconomic status and metacognition (r= .25). Males were significantly less likely to use metacognitive strategies than females (r= -.13). Metacognitive strategies remained a significant

	Study	Aim	Other Key Findings
		and the influence of socioeconomic status and gender	predictor of achievement when controlling for socioeconomic status and gender.
6	Chiu et al., (2007)	To examine the links between academic achievement and learning strategies across countries	This relationship varied depending on the country. There was a positive correlation between metacognition and maths performance in Albania, Hong Kong, Portugal and Thailand. There was a significant negative correlation in Austria, Belgium, Germany, Hungary, Iceland, Israel, Denmark, Finland, Norway and the USA. There was no statistically significant correlation in Australia, Brazil, Bulgaria, Chile, Czech Republic, Italy, Korea, Latvia, Liechtenstein, Luxembourg, Mexico, Holland, New Zealand, Romania, Russia, Sweden, Switzerland, Macedonia and the UK. Students who reported using memorisation strategies often scored lower in all subjects. The association between metacognition and achievement was stronger in more individualistic countries with a less equal distribution of income. Compared to students in individualistic societies, to achievement scores of students in collective cultures were linked

	Study	Aim	Other Key Findings
			more strongly to school mates' use of metacognitive strategies and less strongly to their own use of metacognitive strategies.
7	Erktin (2004)	To examine the role of metacognitive skills in student achievement	Self-efficacy, cognitive capacity and motivation positively predicted maths achievement. Maths anxiety negatively predicted maths achievement.
8	Fadlelmula et al., (2015)	To examine the relationships between students motivational beliefs, the use of self-regulated learning strategies and maths achievement.	Students' mastery goal orientation was significantly related to the use of self-regulated strategies and maths achievement. Elaboration was positively associated with maths achievement. Self-efficacy was associated with achievement goals, the use of self-regulated learning strategies and maths achievement.
9	Fitzpatrick, (1994)	To investigate the relationship of cognitive factors, attributions and gender with mathematical problem-solving.	There were no gender differences in metacognition There was a significant positive correlation between maths achievement and score on the AIMM (cognitive and metacognitive skills combined, r= .86).
10	Fusco (1995)	To explore the relationships between attributions, metacognition and maths problem-solving performance.	There was not a significant relation between attribution group (strategy, effort or unknown causes) and metacognition. Although, students who attributed their performance to strategy were more likely to regulate

	Study	Aim	Other Key Findings
			their exploration and were more likely to solve the problem than students in the other two groups.
11	Harris (2015)	To explore the association between metacognition and academic achievement in a Montessori setting, and how demographics (including disability and socio-economic status) impact on these relationships	The association remained non-significant when students with disabilities were removed from the data. For students in grades 3-8, there was not a significant correlation between metacognition and socioeconomic status or having a disability. There was a strong correlation between maths performance and having a disability and also between maths score and being economically disadvantaged.
12	Hassan & Rahman (2017)	To investigate the role of the learning environment in enhancing metacognitive awareness	The learning environment contributed to 31% of the variance in maths achievement.
13	Ichihara & Arai (2006) (*Japanese*)	To explore the relationship between motivational beliefs and learning outcomes, and how metacognitive activities influences this relationship	Increased metacognition was associated with greater motivation.
14	Maras et al., (2019)	To explore the outcomes of a metacognitive maths intervention for	There was no difference in detecting errors between children with autism and children without autism. However, children with autism showed reduced cohesion between their pre- and post-test intentions.

	Study	Aim	Other Key Findings
		children who have a diagnosis of autism	The feedback intervention significantly increased achievement in children with and without autism.
15	Martín et al., (2008)	To explore individual and school variables associated with academic achievement	Socioeconomic status predicted student learning in social science but not in maths and language. Females obtained better results in language and males obtained better results in social science. There were no differences in math achievement based on gender. Students with the highest abilities at the start of ESO (secondary school) made the least amount of progress as compared to peers.
16	Ning (2016)	To explore different factor structures of the Jr. MAI	For students with lower levels of metacognition the Jr. MAI seemed to represent a unidimensional measure of metacognition, whilst for those with higher levels of metacognition the Jr. MAI appeared to reflect a two-dimensional measure of regulation and knowledge of cognition. Those with lower levels of metacognition reported using significantly less deep learning strategies and had significantly lower levels of mathematics performance than students with higher levels of metacognition.

	Study	Aim	Other Key Findings
17	Özcan (2016)	To investigate the relationship between mathematical problem-solving skills and self-regulated learning	A model was formed in which 24% of the total variance in mathematical problem-solving skills is explained by the three sub-dimensions of the self-regulated learning model: internal motivation (13%), willingness to do homework (7%), and post-problem retrospective metacognitive experience (4%).
18	Özcan & Eren Gümüş (2019)	To investigate the effects of self-efficacy, maths anxiety and metacognitive experience on maths problem-solving performance	There was a significant correlation between metacognition and maths self-efficacy and maths anxiety. Metacognitive experience was the only non-cognitive construct which had a direct effect on problem-solving performance. Metacognitive experience mediated the effects of self-efficacy, motivation, and mathematics anxiety on performance. Motivation and mathematics anxiety had an indirect effect on problem- solving performance through self-efficacy.

	Study	Aim	Other Key Findings
19	Özsoy (2011)	To explore the relationship between metacognitive skills and achievement in maths	Maths achievement was significantly correlated with metacognition. The strongest correlations were with procedural knowledge and prediction, and the lowest was with declarative knowledge and planning.
20	Peng et al., (2014)	To examine the relationships between testing-related motivation (test value, effort, self-efficacy and test anxiety), test-taking strategies (test tactics and metacognitive strategies), gender and maths test performance	Motivation variables (test value, effort, self-efficacy and test anxiety) influenced the use of test-taking strategies. There were gender differences in self-efficacy and test anxiety. The use of metacognitive strategies was significantly associated with test value, effort, self-efficacy, test tactics and test anxiety. Students' reported use of metacognitive strategies had a direct effect on test tactics. However, metacognitive strategies and test tactics did not demonstrate significant effects on math performance.
21	Sink et al., (1991)	To investigate the relationship between self-regulation and academic performance	Metacognitive variables such as planning and self-assessment were significantly related to achievement in mathematics, reading, and science. However, student and teacher perceptions of scholastic ability were more salient factors in predicting academic performance. There

	Study	Aim	Other Key Findings
			was no relationship between locus of control and classroom grades/performance on standardised achievement measures.
22	Tian et al., (2018)	To investigate whether the relationship between metacognitive knowledge and maths performance is mediated by self-efficacy and motivation	The use of metacognitive strategies was significantly associated with intrinsic motivation and self-efficacy in maths. Metacognitive knowledge, self-efficacy and intrinsic motivation predicted mathematics performance. There were some sex differences with males scoring higher on some metacognitive scales than females, and lower on others. The association between metacognitive knowledge and mathematics performance was mediated by self-efficacy and intrinsic motivation.
23	van der Stel & Veenman (2014)	To explore the development of metacognitive skills	Metacognitive skills contribute to learning performance, partly independent of intellectual ability. Metacognitive skills appear to be predominantly general by nature over the years. Although a smaller domain-specific component was found in the first 2 years, this disintegrated in the third year. Metacognitive skills do not develop linearly at a consistent pace. Between 13 and 14 years there was an increase in metacognitive skills frequency and quality, but most scales stayed stable or regressed in students' 3rd year.

	Study	Aim	Other Key Findings
24	van der Stel et al., (2010)	To compare metacognitive skills in second- and third-year secondary-school students	The frequency of metacognitive activity, especially planning and evaluation, increased with age. Intelligence was a strong predictor of math performance in 13- to 14-year-olds, but it was less prominent in 14- to 15-year-olds. The quality of metacognitive skills appeared to predict math performance in both age groups but was a stronger predictor in 14 to 15-year olds, even when intelligence is taken into account.
25	van der Walt et al., (2008)	To investigate the value of metacognition in the learning of mathematics	Metacognitive strategies in respect of the prediction, evaluation, monitoring and reflection of the learners in the research group were inadequate to facilitate critical thinking and the notion about thinking about one's own thinking.
26	Veenman et al., (2000)	To investigate the role of metacognition in test anxiety	Low test-anxious participants exhibited superior metacognitive skilfulness during math performance relative to high test-anxious subjects. Differences in metacognitive skilfulness were performance-related.

	Study	Aim	Other Key Findings
27	Veenman et al., (2005)	To explore the relationship between intelligence and metacognitive skill and to explore the impact of giving “hints” on metacognitive skills	Without hints (cues), metacognitive skilfulness is the main predictor of initial learning, while intelligence additionally enters the regression equation after the presentation of metacognitive hints. GPA also appears to be predicted by a combination of intellectual and metacognitive skills.
28	Veenman & Spaans (2005)	To explore to what extent the development of metacognitive skills is associated with intellectual growth, and to explore the generality versus domain specificity of metacognitive skills across age groups.	Metacognitive skilfulness develops alongside, but not fully dependent on intellectual ability. Metacognitive skilfulness outweighs intelligence as a predictor of learning performance. Metacognitive skills appear to be general for third year (older) students, but rather domain-specific for first-year (younger) students.
29	Walker (2013)	To explore and compare the effects of a working memory intervention and metacognitive intervention on metacognition, working memory and maths achievement.	Participants who engaged in working memory or one-to-one tutoring intervention made significant improvements from pre to post and from pre to follow-up in terms of working memory and maths achievement. There was also a significant reduction in maths anxiety in these students

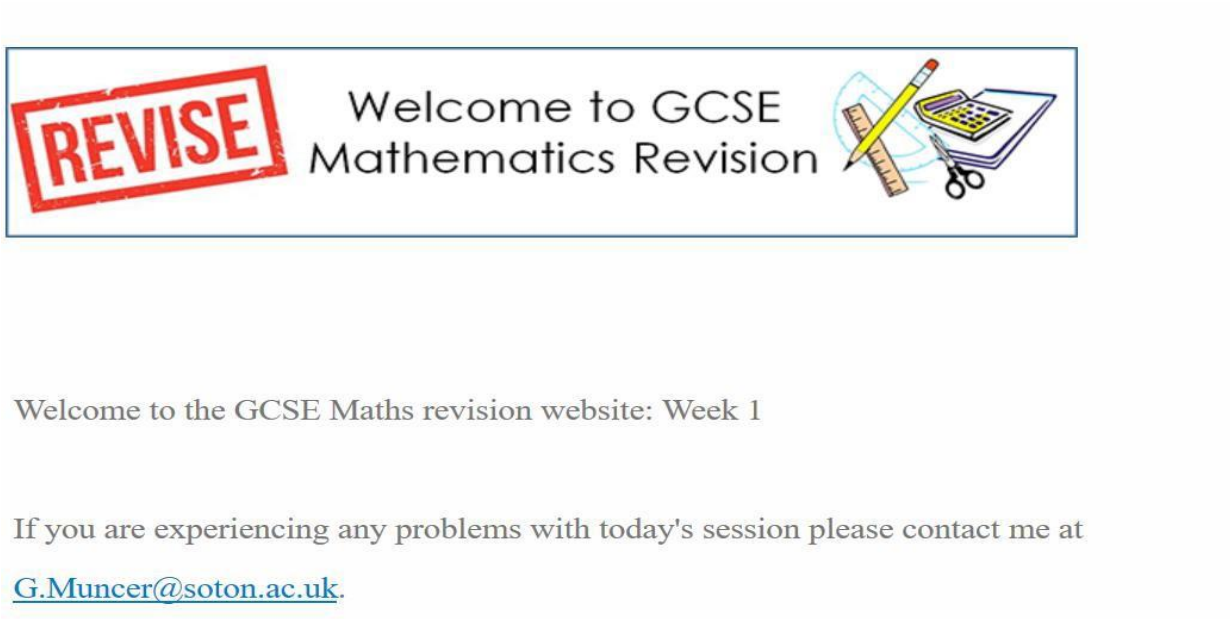
	Study	Aim	Other Key Findings
			over the same timescale, but there were no significant changes in metacognition or general anxiety.
30	Yap (1993)	To identify whether O'Neil et al.'s (1992) model of self-regulated learning and math achievement is supported by the National Assessment of Educational Progress (NAEP) data	State worry affects students' test performance and not vice versa.
31	Young & Worrell (2018)	To explore the relationships between metacognition, effort, worry and maths achievement	Study 1 results indicated that Jr. MAI scores were internally consistent and yielded an interpretable two-factor structure after the elimination of several items; however, the scores were not significantly or meaningfully related to GPA or current and future mathematics achievement. In Study 2 (n = 30), Jr. MAI scores did not predict students' metacognitive behaviours during mathematics problem-solving tasks. In contrast, students' metacognitive behaviours observed during problem-solving were meaningfully related to mathematics

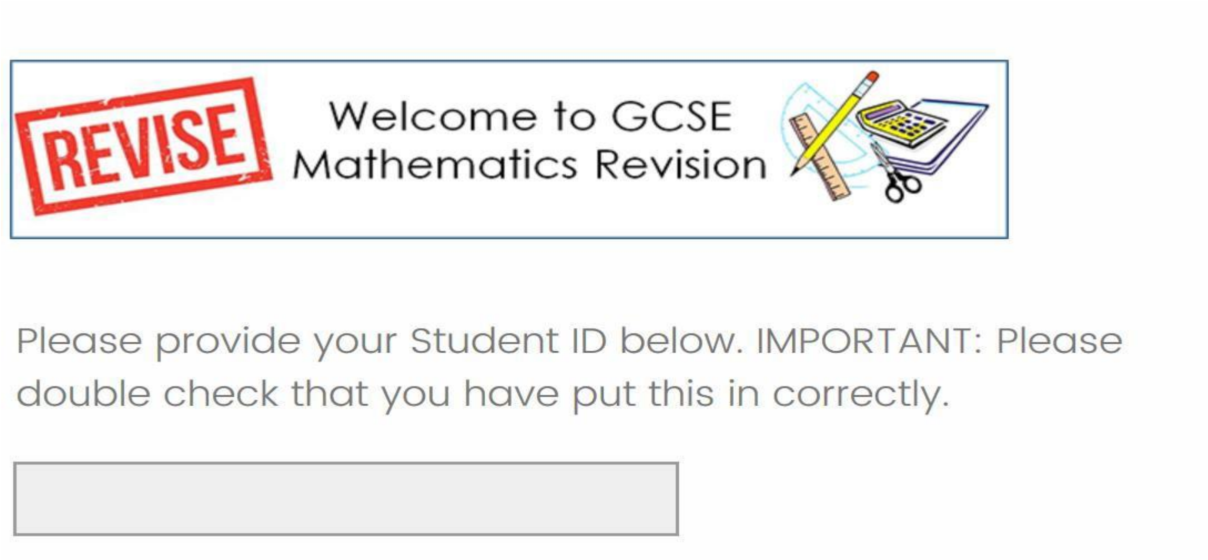
Study	Aim	Other Key Findings
		achievement with medium to high effect sizes. Findings support the predictive validity of metacognition with regard to academic achievement when operationalized with problem-solving interviews but call into question the criterion-related validity of Jr. MAI scores.

Appendix G Screenshots of the Intervention

Figure A1. *Screenshots of the Intervention as Seen by Participants in the Relearn and Restudy Conditions*

The following two screenshots were seen by all participants when they clicked on the session link





Following this, participants in the relearn condition only saw the following instructions.

REVISE

Welcome to GCSE Mathematics Revision

You will be asked to look at 20 statements about maths.

Your task is to **fill in the missing word in the statement.**

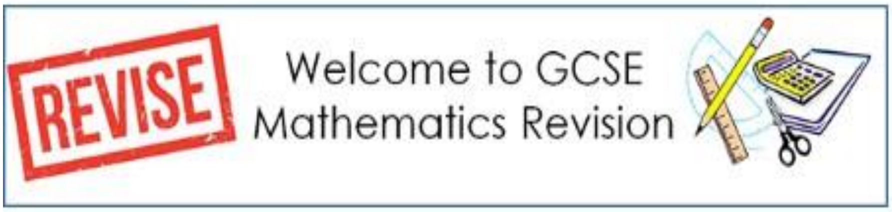
After answering, you will receive feedback. If you get a question wrong, you will see it for a second time later on.

The questions are timed. When each question is shown for the first time you have 20 seconds to read the question and type in the answer. If you finish quicker than this, you will still NOT be able to continue to the next question until the time runs out so please be patient :)

After each question and feedback you will be asked to click on the arrow button to proceed.

The feedback for each question is shown for 5 seconds. If you answer any questions incorrectly, and they are shown for a second time, you will only have 15 seconds to read the question and type your answer.

Participants in the restudy condition only saw the following instructions



Welcome to your GCSE maths revision task

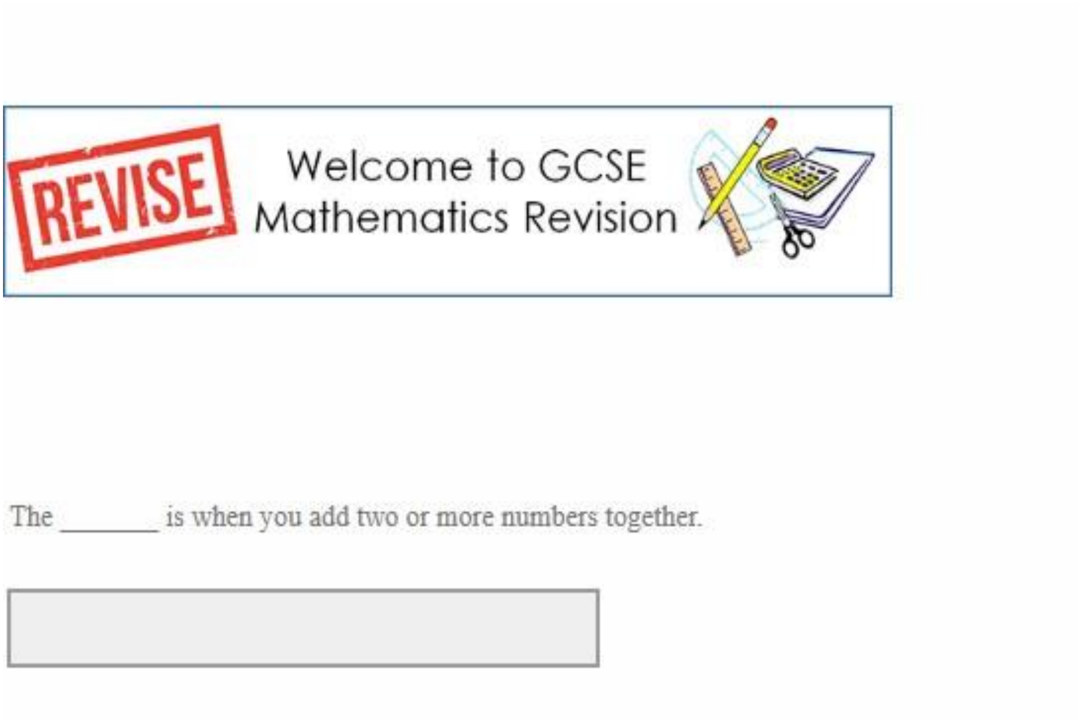
You will be asked to look at 20 statements about maths. These will be shown to you one at a time and each statement will be shown twice (please do read it both times)

Your task is to **read and remember the statement.**

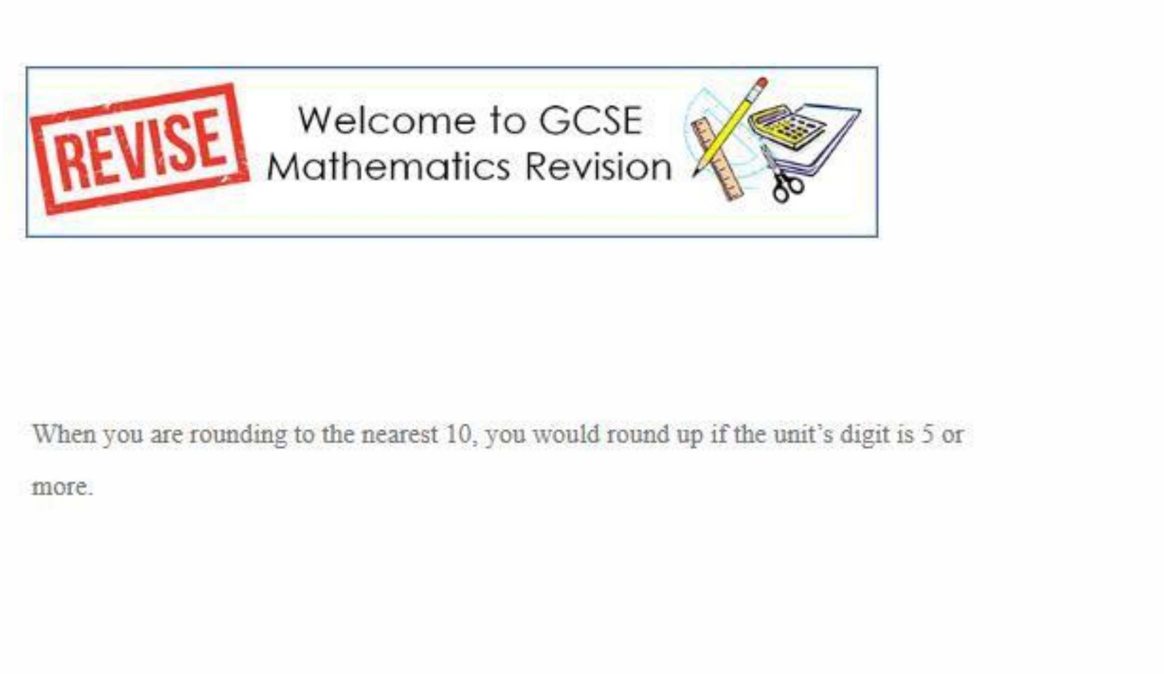
The display of the statements are timed. When each statement is shown for the first time you have 25 seconds to read and remember the statement. On the second time, this is 20 seconds. If you read the sentence quicker than this, you will still NOT be able to continue to the next question until the time runs out so please be patient :)

After the time is up for each sentence, you will be asked to click on the arrow button to proceed.

Participants then saw 20 items. An example of a relearn item is shown below.



An example restudy item is shown below.



Appendix H Maths Test

Student sheet ***Instructions**: Fill in the gap to complete the sentences.*

You have 15 minutes.

IMPORTANT: What is your student ID number? _____________

1. When you arrange a set of numbers in order from smallest to largest and find the number in the middle. This is called the_________.
2. A number which can only be divided by______and itself is called a prime number.
3. In the expression y>x the symbol shows that the value of y is_________than X.
4. The___________is the difference between the largest and smallest number in a set (largest number minus smallest number).
5. To calculate the mean you would_________the total value of all numbers in a set and then divide them by how many numbers there are within the set.
6. When rounding to the nearest 10, you would round down if the last number is ________or less.
7. The___________is the most frequently occurring number in a set.
8. When working out an equation you follow this order: Brackets,_________Division, Multiplication, Addition, Subtraction (BIDMAS).
9. 5183 to three significant figures is_________.
10. When a decimal number is repeated forever, this is called a___________decimal.
11. The reciprocal of a number is_____divided by the number.
12. There is a_________proportion between two values when one is a multiple of the other.
13. An __________is a whole number with no decimal points.
14. Fractions which do not show whole numbers separately, and whose numerators are larger than their ________are sometimes called top-heavy fractions. Another word for them is Improper fractions.
15. If the ratio of two weights is 1:5. The second weight is_______times heavier than the first weight.
16. Fractions which are worth the same amount but are written differently, for example 2/4 and 4/8 are called___________fractions.
17. The_________of a straight-line graph shows the rate of change.

18. The percentage of a whole is________%
19. A number which divides exactly into another number is called a___________.
20. The_________number which will divide into two or more numbers is called the highest common factor.
21. The_________of a line is called the gradient.
22. There are five times as many students as teachers, therefore the ratio of teachers to students is_________.
23. How many times larger or smaller a shape will be if it is enlarged or reduced is called it's___________factor.
24. A number which is in the_________of another number is called a multiple.
25. The number 25.743 to_________significant figures is 25.7.
26. If one variable increases, and a second variable also increases, these two variables have a___________correlation.
27. There are_________centimetres in a metre.
28. 4f+8_________is 4(f+2).
29. In an equation, if two minus signs are together, for example 3- - 2, you replace both of the minus signs with a___________sign.
30. A straight line on a graph which goes as centrally through the plotted coordinate points as possible is called a line of best_________.
31. The subject of the equation Y= 45-3x is________.
32. 40 000 in___________form is 4 x 10^4.
33. To calculate the perimeter of a rectangle, you would add up the length of all the ________.
34. Here is a sequence: 12,15,18,21. The nth term of this sequence is n+_____.
35. The___________of a circle is the distance from one edge of the circle, through the centre, to the other side of the circle.
36. 20 pence in pounds is £__________.

Appendix I Attitudes Towards Mathematics Inventory

Directions: This inventory consists of statements about your attitude toward mathematics. There are no correct or incorrect responses. Read each item carefully. Please think about how you feel about each item. Darken the circle that most closely corresponds to how the statements best describes your feelings. Use the following response scale to respond to each item.

PLEASE USE THESE RESPONSE CODES:

A – Strongly Disagree

B – Disagree

C – Neutral

D – Agree

E – Strongly Agree

1. Mathematics is a very worthwhile and necessary subject.
2. I want to develop my mathematical skills.
3. I get a great deal of satisfaction out of solving a mathematics problem.
4. Mathematics helps develop the mind and teaches a person to think.
5. Mathematics is important in everyday life.
6. Mathematics is one of the most important subjects for people to study.
7. High school math courses would be very helpful no matter what I decide to study.
8. I can think of many ways that I use math outside of school.
9. Mathematics is one of my most dreaded subjects.
10. My mind goes blank and I am unable to think clearly when working with mathematics.
11. Studying mathematics makes me feel nervous.
12. Mathematics makes me feel uncomfortable.
13. I am always under a terrible strain in a math class.
14. When I hear the word mathematics, I have a feeling of dislike.
15. It makes me nervous to even think about having to do a mathematics problem.
16. Mathematics does not scare me at all.

17. I have a lot of self-confidence when it comes to mathematics.
18. I am able to solve mathematics problems without too much difficulty.
19. I expect to do fairly well in any math class I take.
20. I am always confused in my mathematics class.
21. I feel a sense of insecurity when attempting mathematics.
22. I learn mathematics easily.
23. I am confident that I could learn advanced mathematics.
24. I have usually enjoyed studying mathematics in school.
25. Mathematics is dull and boring.
26. I like to solve new problems in mathematics.
27. I would prefer to do an assignment in math than to write an essay.
28. I would like to avoid using mathematics in college.
29. I really like mathematics.
30. I am happier in a math class than in any other class.
31. Mathematics is a very interesting subject.
32. I am willing to take more than the required amount of mathematics.
33. I plan to take as much mathematics as I can during my education.
34. The challenge of math appeals to me.
35. I think studying advanced mathematics is useful.
36. I believe studying math helps me with problem solving in other areas.
37. I am comfortable expressing my own ideas on how to look for solutions to a difficult problem in math.
38. I am comfortable answering questions in math class.
39. A strong math background could help me in my professional life.
40. I believe I am good at solving math problems.

Appendix J Academic Efficacy Scale of the Patterns of Adapting Learning Survey

The first question is an example.

1. I like strawberry ice cream.

1	2	3	4	5
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

HERE ARE SOME QUESTIONS ABOUT YOURSELF AS A STUDENT IN YOUR MATHS CLASS. **PLEASE CIRCLE THE NUMBER THAT BEST DESCRIBES WHAT YOU THINK.** Think specifically about how you feel about maths when you answer the questions.

1. I'm certain I can master the skills taught in class this year.

1	2	3	4	5
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

2. I'm certain I can figure out how to do the most difficult class work.

1	2	3	4	5
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

3. I can do almost all the work in class if I don't give up.

1	2	3	4	5
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

4. Even if the work is hard, I can learn it.

1	2	3	4	5
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

5. I can do even the hardest work in this class if I try.

1	2	3	4	5
NOT AT ALL TRUE		SOMEWHAT TRUE		VERY TRUE

Appendix K Attentional Control Scale

These questions are about how well you feel you concentrate on your work. Please read each statement and indicate how often it is true for you on the scale beside each question.

Item	Almost Never	Sometimes	Often	Always
1. It’s very hard for me to concentrate on a difficult task when there are noises around.	Almost Never	Sometimes	Often	Always
2. When I need to concentrate and solve a problem, I have trouble focusing my attention.	Almost Never	Sometimes	Often	Always
3. When I am working hard on something, I still get distracted by events around me.	Almost Never	Sometimes	Often	Always
4. My concentration is good even if there is music in the room around me.	Almost Never	Sometimes	Often	Always
5. When concentrating, I can focus my attention so that I become unaware of what’s going on in the room around me.	Almost Never	Sometimes	Often	Always
6. When I am reading or studying, I am easily distracted if there are people talking in the same room.	Almost Never	Sometimes	Often	Always
7. When trying to focus my attention on something, I have difficulty blocking out distracting thoughts.	Almost Never	Sometimes	Often	Always

8. I have a hard time concentrating when I am excited about something.	Almost Never	Sometimes	Often	Always
9. When concentrating I ignore feelings of hunger or thirst.	Almost Never	Sometimes	Often	Always
10. I can quickly switch from one task to another.	Almost Never	Sometimes	Often	Always
11. It takes me a while to get really involved in a new task.	Almost Never	Sometimes	Often	Always
12. It is difficult to coordinate my attention between the listening and writing required when taking notes during lessons.	Almost Never	Sometimes	Often	Always
13. I can become interested in a new topic very quickly when I need to.	Almost Never	Sometimes	Often	Always
14. It is easy for me to read or write while I am also talking on the phone.	Almost Never	Sometimes	Often	Always
15. I have trouble carrying out two conversations at once.	Almost Never	Sometimes	Often	Always
16. I have a hard time coming up with new ideas quickly.	Almost Never	Sometimes	Often	Always
17. After being interrupted or distracted, I can easily switch my attention back to what I was doing before.	Almost Never	Sometimes	Often	Always
18. When a distracting thought comes to mind, it is easy for me to shift my attention away from it.	Almost Never	Sometimes	Often	Always

19. It is easy for me to alternate between two different tasks.	Almost Never	Sometimes	Often	Always
20. It is hard for me to break from one way of thinking about something and look at it from another point of view.	Almost Never	Sometimes	Often	Always

Appendix L Generalised Anxiety Disorder Scale of the Youth Anxiety Measure for DSM-5

***Circle the number** which best describes how you feel for each question.*

1. I worry about a lot of things

0	1	2	3
Never	Sometimes	Often	Always

2. I think a lot about what can go wrong.

0	1	2	3
Never	Sometimes	Often	Always

3. I find it hard to stop worrying.

0	1	2	3
Never	Sometimes	Often	Always

4. I worry a lot about not doing well at school.

0	1	2	3
Never	Sometimes	Often	Always

5. I worry a lot about all the bad things that happen in the world.

0	1	2	3
Never	Sometimes	Often	Always

6. I don’t feel well because I worry so much.

0	1	2	3
Never	Sometimes	Often	Always

Appendix M Abbreviated Maths Anxiety Scale (AMAS)

Please indicate the level of your anxiety in the following situations. Please choose ONE box on each line.

		Not at all	A little	A fair amount	Much	Very much
1.	Studying for a math test.	☐	☐	☐	☐	☐
2.	Taking math section of the college entrance exam.	☐	☐	☐	☐	☐
3.	Taking an exam (quiz) in a math course.	☐	☐	☐	☐	☐
4.	Taking an exam (final) in a math course.	☐	☐	☐	☐	☐
5.	Picking up math textbook to begin working on a homework assignment.	☐	☐	☐	☐	☐
6.	Being given homework assignments of many difficult problems that are due the next class meeting.	☐	☐	☐	☐	☐
7.	Thinking about an upcoming math test 1 week before.	☐	☐	☐	☐	☐
8.	Thinking about an upcoming math test 1 day before.	☐	☐	☐	☐	☐
9.	Thinking about an upcoming math test 1 hour before.	☐	☐	☐	☐	☐
10.	Realizing you have to take a certain number of math classes to fulfill requirements.	☐	☐	☐	☐	☐
11.	Picking up math textbook to begin a difficult	☐	☐	☐	☐	☐

	reading assignment.					
12.	Receiving your final math grade in the mail.	☐	☐	☐	☐	☐
13.	Opening a math or stat book and seeing a page full of problems.	☐	☐	☐	☐	☐
14.	Getting ready to study for a math test.	☐	☐	☐	☐	☐
15.	Being given a "pop" quiz in a math class.	☐	☐	☐	☐	☐
16.	Reading a cash register receipt after your purchase.	☐	☐	☐	☐	☐
17.	Being given a set of numerical problems involving addition to solve on paper.	☐	☐	☐	☐	☐
18.	Being given a set of subtraction problems to solve.	☐	☐	☐	☐	☐
19.	Being given a set of multiplication problems to solve.	☐	☐	☐	☐	☐
20.	Being given a set of division problems to solve.	☐	☐	☐	☐	☐
21.	Buying a math textbook.	☐	☐	☐	☐	☐
22.	Watching a teacher work on an algebraic equation on the blackboard.	☐	☐	☐	☐	☐
23.	Signing up for a math course.	☐	☐	☐	☐	☐
24.	Listening to another student explain a math formula.	☐	☐	☐	☐	☐
25.	Walking into a math class.	☐	☐	☐	☐	☐

Appendix N Visual Analogue Scale (Anxious Affect, Mastery, Attention)

The scale below consists of words describing different feelings and emotions.

Rate each word by drawing a vertical line on the scale below to indicate the extent you feel that way RIGHT NOW

I FEEL………………………………………

ANXIOUS

Not at all A little Moderately Quite a bit Extremely

0--100

I UNDERSTAND THIS TOPIC

Not at all A little Moderately Quite a bit Extremely

0--100

FOCUSED

Not at all A little Moderately Quite a bit Extremely

0--100

CONFUSED ABOUT THIS TOPIC

Not at all A little Moderately Quite a bit Extremely

0--100

DISORGANISED

Not at all A little Moderately Quite a bit Extremely

0--100

WORRIED

Not at all A little Moderately Quite a bit Extremely

0---100

I CONCENTRATED

Not at all A little Moderately Quite a bit Extremely

0---100

RELAXED

Not at all A little Moderately Quite a bit Extremely

0---100

I HAVE MASTERED THIS TOPIC

Not at all A little Moderately Quite a bit Extremely

0---100

Appendix O Email to Invite Colleges to Participate in the Study

Good afternoon,

I am a year two trainee Educational Psychologist, studying at the University of Southampton.

As part of my doctorate, I am running a thesis project on a revision method which is called successive relearning. I am interested in whether successive relearning supports learners who are re-sitting their GCSE mathematics by improving their learning and broader outcomes such as attitudes to mathematics. I am looking to do this project with 16-18-year olds who are re-sitting their GCSE in mathematics. I was hoping to chat to someone at X College to see whether you may be interested in participating in this project. Please may this email be forwarded on to an appropriate member of staff- possibly a member of Senior Leadership Team/ Head of Maths.

This project will involve identifying an appropriate sample of learners who are re-sitting their GCSE in mathematics. All learners will be asked whether they would like to participate and will have the option not to. Learners who wish to participate will be asked to fill in several questionnaires online (this will take approximately 20 minutes) at the start of this study. They will be sent a link to access these questionnaires and the questionnaires will measure attitudes to mathematics, maths anxiety, and self-confidence in maths. Following the completion of these questions, some learners will be asked to fill in an online revision intervention at home three times a week on set days (Monday, Wednesday and Friday) for four weeks. The online intervention will take approximately 25 minutes per session. Learners will be sent an email link to do this, and this involves answering 20 revision questions which are relevant to the topic they have been studying or reading 20 relevant revision statements. Learners should carry on with other revision methods during this time and this will be made clear. The intervention can be completed on a computer, tablet or smart phone. Some learners will not be asked to take part in this online intervention. After four weeks, learners will be asked to sit a short test in class and fill in the questionnaires again.

I will create the online revision intervention and will write the questions for the intervention, and the end of topic test, and will mark this. As part of the project, I would be asking the college for support in the following:

1. Identifying appropriate participants.
2. I would be asking teachers to introduce the study to appropriate participants in a maths lesson, using resources provided by myself, and allowing students the time to fill in the initial questionnaires in class (approx. 20 minutes).
3. Four weeks later, I would be asking teachers to allow students time to complete the final questionnaires in class (approx. 20 minutes), allowing students time to complete a 20-minute test in class under exam conditions, and handing out debriefing sheets (provided by myself).

I have attached learner and teacher information sheets for further information but please do not hesitate to contact me if you wanted to discuss this further.

Many thanks,

Gemma

Appendix P Participant Information Sheet

Study Title: Exploring the impact of successive relearning on learning and motivation in a post-16 GCSE Mathematics qualification.

Researcher: Gemma Muncer

ERGO number: 47186

You are being invited to take part in the above research study. To help you decide whether you would like to take part or not, it is important that you understand why the research is being done and what it will involve. Please read the information below carefully and ask questions if anything is not clear or you would like more information before you decide to take part in this research. You may like to discuss it with others, but it is up to you to decide whether or not to take part.

What is the research about?

I am a trainee Educational Psychologist who is running this project as part of a doctorate in Educational Psychology qualification.

Previous research has identified that practice testing (testing yourself on information you need to learn) and distributed practice (when you space out your study periods rather than cramming them into a short space of time) improve learning and test performance (Bjork, Dunlosky, & Kornell, 2013), and reduce how nervous students feel for their examinations (Agarwal, D'Antonio, Roediger, McDermott, & McDaniel, 2014). A strategy which combines these two revision techniques is called ***successive relearning***. Previous research with adults has concluded that successive relearning is a useful revision strategy (Rawson, Dunlosky, & Sciartelli, 2013). We are interested in finding out whether it is also useful for GCSE maths students.

To investigate this, we have designed an online revision intervention. We are interested in finding out whether students who complete a revision exercise online three times each week will do better in a test than students who do not complete the intervention. I am also interested if students who complete the revision intervention will feel less anxious and more positive and confident about their maths studies.

Why have I been asked to participate?

You have been asked whether you would like to participate as you are aged 16-18 years old and you will be re-sitting your maths GCSE in 2019 or 2020.

What will happen to me if I take part?

If you choose to take part in this study, you complete some questionnaires on the computer or on your phone in your maths lesson. At the start of the questionnaires, please put in your student ID number when you are asked. The questionnaires will be about attitudes to mathematics, attentional control, anxiety and self-efficacy. Your answers will be kept anonymous and confidential, this means that your answers will not be shared with anyone in a way which identifies you.

You will be put in one of three groups (A, B or C). **Everyone in your class will <u>not</u> be in the same group,** it is likely that some students in your class will be in group A, some will be in group B and some will be in group C. This is decided randomly. You will be told which group you are in and will receive instructions for your group via your student email.

Group A: You will complete an online revision intervention three times each week (on Monday, Wednesday and Friday) over four weeks. This will mean answering twenty maths revision questions on an online programme. You can do this on a smartphone, computer or tablet.

Group B: You will complete an online revision intervention three times each week (on Monday, Wednesday and Friday) over four weeks. This will mean reading twenty revision statements about maths on an online programme. You can do this on a smartphone, computer or tablet.

Group C: You will complete revision at home as usual.

For Group A and B, the online revision sessions will take a **maximum** of 25 minutes.

Four weeks later, you will be asked to complete the same questionnaires which you completed at the start of the study. You will do this on a computer or on your phone in your maths lesson. You will also have a maths test. Before the test, you will be asked to estimate your score. Your score will be recorded and used as part of the study but will be kept anonymous.

Are there any benefits in my taking part?

If you choose to participate, you may experience personal learning and motivational benefits. Your participation may also benefit the learning and motivational outcomes of future students.

At the end of the study, students in groups A and B will be given 50p credit in the X Café per online revision session completed.

All students who complete questionnaires (in groups A, B or C) will be entered into a prize draw for three £50 Amazon vouchers.

Are there any risks involved?

Some of the questionnaires will be exploring feelings of anxiety which make you feel upset or worried.

If this happens, you can speak to your teacher or tutor who will signpost you to sources of support within the school.

You might also find the following websites about how to manage anxiety useful:

- https://kooth.com/
- https://www.samaritans.org/education/deal/coping-strategies/exam-stress
- https://www.childline.org.uk/info-advice/school-college-and-work/

What data will be collected?

If you give informed consent, I will collect the following information from the College:

- Your College email address
- The course you are enrolled on
- Your attendance
- The grade you achieved in GCSE maths if you sat it in Year 11.
- Your predicted grade.
- How many times you have re-sat GCSE maths.
- Your results in the test at the end of the study.
- What GCSE grade you get in maths when you re-sit in 2019/2020.

I will also collect the following additional information directly from you:

- Your age.
- Your sex (male or female).
- The questionnaires will be about attitudes to mathematics, attentional control, self-efficacy (confidence) and anxiety.

Your data will be recorded next to your student ID number so that you are not identifiable. No data you provide will be stored next to your name, and all information will be stored securely on a password protected computer.

Your email address will be stored to enable you to receive reminders and the link to complete the intervention online if you are in group A or B. This will be deleted following the study.

Will my participation be confidential?

Your participation and the information we collect about you during the course of the research will be kept strictly confidential.

Only members of the research team and responsible members of the University of Southampton may be given access to data about you for monitoring purposes and/or to carry out an audit of the study to ensure that the research is complying with applicable regulations. Individuals from regulatory authorities (people who check that we are carrying out the study correctly) may require access to your data. All of these people have a duty to keep your information, as a research participant, strictly confidential.

Data kept electronically will be password-protected and will only be used for the purpose of this study and accessed by the researcher and her two supervisors at The University of Southampton.

Do I have to take part?

No, it is entirely up to you to decide whether or not to take part. If you decide you want to take part, you will need to indicate that you are deciding to take part when you complete the questionnaires and intervention sessions online.

What happens if I change my mind?

You have the right to change your mind and withdraw at any time during the study without giving a reason and without your participant rights being affected.

You can withdraw by emailing me directly at X

You can withdraw following the study (up until 01.11.19). After this date data analysis will begin.

What will happen to the results of the research?

Your personal details will remain strictly confidential. Research findings made available in any reports or publications will not include information that can directly identify you.

The results of this study will be written up as part of my doctorate. It may be published. X College will receive a copy of the results following the study. You can obtain a copy of the results and a one-page summary of the study yourself by emailing g.muncer@soton.ac.uk. This will be available from 01.09.2020.

Where can I get more information?

If you have any questions or would like more information about the study, please email one of the research team at:

Gemma Muncer at g.muncer@soton.ac.uk

J. Hadwin at X

P. Higham at x

What happens if there is a problem?

If you have a concern about any aspect of this study, you should speak to the researchers who will do their best to answer your questions.

If you remain unhappy or have a complaint about any aspect of this study, please contact the University of Southampton Research Integrity and Governance Manager (023 8059 5058, rgoinfo@soton.ac.uk).

Data Protection Privacy Notice

The University of Southampton conducts research to the highest standards of research integrity. As a publicly-funded organisation, the University has to ensure that it is in the public interest when we use personally-identifiable information about people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use information about you in the ways needed, and for the purposes specified, to conduct and complete the research project. Under data protection law, 'Personal data' means any information that relates to and is capable of identifying a living individual. The University's data protection policy governing the use of personal data by the University can be found on its website (https://www.southampton.ac.uk/legalservices/what-we-do/data- protection-and-foi.page).

This Participant Information Sheet tells you what data will be collected for this project and whether this includes any personal data. Please ask the research team if you have any questions or are unclear what data is being collected about you.

Our privacy notice for research participants provides more information on how the University of Southampton collects and uses your personal data when you take part in one of our research projects and can be found at http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Inte grity%20Privacy%20Notice/Privacy%20Notice%20for%20Research%20Participants.pdf

Any personal data we collect in this study will be used only for the purposes of carrying out our research and will be handled according to the University's policies in line with data protection law. If any personal data is used from which you can be identified directly, it will not be disclosed to anyone else without your consent unless the University of Southampton is required by law to disclose it.

Data protection law requires us to have a valid legal reason ('lawful basis') to process and use your Personal data. The lawful basis for processing personal information in this research study is for the performance of a task carried out in the public interest. Personal data collected for research will not be used for any other purpose.

For the purposes of data protection law, the University of Southampton is the 'Data Controller' for this study, which means that we are responsible for looking after your information and using it properly.

Appendix Q Teacher Information Sheet

Study Title: Exploring the impact of successive relearning on learning and motivation in a post-16 GCSE Mathematics qualification.

Researcher: Gemma Muncer

ERGO number:

47186 Dear Teachers,

We are working with XXX College to conduct research to look at revision methods for improving outcomes for students who are re-sitting their GCSE in mathematics.

What is the research about?

I am a trainee Educational Psychologist who is running this project as part of a doctorate in Educational Psychology qualification.

Previous research has identified that practice testing (testing oneself on information to be learnt) and distributed practice (when study periods are spaced out over time) improve learning and test performance (Bjork, Dunlosky, & Kornell, 2013), and reduce how nervous students feel for their examinations (Agarwal, D'Antonio, Roediger, McDermott, & McDaniel, 2014). A strategy which combines these two revision techniques is called ***successive relearning***. Previous research with adults has concluded that successive relearning is a useful revision strategy (Rawson, Dunlosky, & Sciartelli, 2013). We are interested in finding out whether it is also useful for GCSE maths students.

To investigate this, we have designed an online revision intervention. I am interested in finding out whether students who complete a revision exercise online three times each week will do better in a test than students who do not complete the intervention. I am also interested if students who complete the revision intervention will feel less anxious and more positive and confident about their maths studies and will have more accurate judgements of their own learning.

It is possible that that students who complete the online revision intervention three times each week will perform better on an end of topic test, will feel less anxious and more positive and confident about their maths GCSE studies.

Why have my students been asked to participate?

Your students have been asked to participate as they are aged 16-18 years old and will be re- sitting their maths GCSE in 2019 or 2020. It is hoped that this study will support pupils' revision and learning for their GCSE in maths.

What is my role and what will happen if my students choose to take part?

Individual students will be randomly assigned to either group A, B or C. All members of your class will **not** be in the same group. Some students will be in group A, some will be in group B and some will be in group C.

We would appreciate the following support from you:

- In the week beginning Monday 9th September 2019, you will be provided with instruction sheets and information sheets for your students. We would appreciate it if you could read the student information sheet with students and allow them time to read their individual student instruction sheets during their maths lesson. Students will also have been emailed a link to a questionnaire to their College email address. If students consent to take part, they are asked to complete questionnaires in the same maths lesson, either on a computer or on a Smart phone. (N.B. Only if students do not have access to either of these devices in the lesson and still want to take part, please inform them to complete the questionnaire at home, but where possible we would appreciate it if students were able to complete the questionnaires in class).
- From the week beginning Monday, 16th September:
    - Students in groups A and B will be sent email links to complete a short online revision session (maximum 25 minutes) **outside** of lesson time three times each week on Monday, Wednesday and Friday for four weeks.
    - Group C will not need to complete online revision sessions and do not need to do anything additional during this period.
    - During this period, we do not require you to do anything differently in lesson.
- In the week beginning Monday, 14th October, students will be emailed a link to a questionnaire to their College email address. These questionnaires are the same as

those they completed a few weeks ago, however it is important that these are filled in to see whether their answers have changed in the past few weeks. After this, students are asked to complete a short maths test in class. This will be provided by the researcher, and students should complete this under exam conditions. This test will cover vocabulary and key points from the topic(s) you have been studying during the term so far. Students should not discuss answers with you or with each other. It is not a requirement of you to mark the test, but if you wish to do so you can go through the answers with your students after the test has finished to provide them with feedback. The researcher will provide all the materials you need for this. Following the test, please collect all student test papers and pass these to the Head of Maths. Students will be emailed a debriefing statement. You will be provided with a paper copy of this. We would appreciate it if you could read this to students following the test.

Are there any benefits in my students taking part?

If your students choose to participate, they may experience personal learning and motivational benefits. Their participation may also benefit the learning and motivational outcomes of future students.

Students who are in group A or B and choose to participate will receive 50p X Café credit per completed intervention session. All students (in groups A, B and C) who complete questionnaires will be entered into a prize draw for three £50 Amazon vouchers.

Are there any risks involved?

Some of the questionnaires will be exploring feelings of anxiety which make students feel upset or worried.

If this happens, they have been signposted to speak to their teacher or tutor to identify sources of support within the school.

They might also find the following websites about how to manage anxiety useful:

- https://kooth.com/
- https://www.samaritans.org/education/deal/coping-strategies/exam-stress
- https://www.childline.org.uk/info-advice/school-college-and-work/

What data will be collected?

If students give informed consent to take part in the study, we will collect the following information from the College for each student who takes part:

- Their College email address
- The course they are enrolled on
- Their attendance
- The grade they achieved in GCSE maths in Year 11.
- Their predicted grade for their maths GCSE
- How many times they have re-sat GCSE maths.
- Their results in the test at the end of the study.
- Their obtained GCSE maths re-sit grade in 2019/2020.

I will also collect the following additional information directly from students:

- Their age.
- Their sex (male or female).
- Questionnaire results about attitudes to mathematics, attentional control, self-efficacy (confidence) and anxiety.
- Their prediction of how many questions they will get correct in the test.

Students' data will be recorded next to their student ID number so that they are not identifiable. No data they provide will be stored next to their name, and all information will be stored securely on a password protected computer.

Students' email addresses will be stored to enable them to receive email reminders to complete sessions, and the questionnaire at the end of the study. Their email address will be deleted by the researchers following the study.

Will participation be confidential?

Your students' participation and the information we collect about students during the course of the research will be kept strictly confidential.

Only members of the research team and responsible members of the University of Southampton may be given access to data about your students for monitoring purposes and/or to carry out an audit of the study to ensure that the research is complying with applicable regulations. Individuals from regulatory authorities (people who check that we are carrying out the study correctly) may require access to this data. All of these people have a duty to keep your information, as a research participant, strictly confidential.

Data kept electronically will be password-protected and will only be used for the purpose of this study and accessed by the researcher and her two supervisors at The University of Southampton.

Do my students have to take part?

No, it is entirely up to your students to decide whether or not to take part. If they decide they want to take part, they will need to indicate that they are deciding to provide consent when they complete the questionnaires and intervention sessions online.

What happens if my students' change my mind?

Your students have the right to change their mind and withdraw at any time during the study without giving a reason and without their participant rights being affected.

Your students can withdraw by emailing me directly at g.muncer@soton.ac.uk

Your students can withdraw following the study (up until 01.11.2019). After this date data analysis will begin.

What will happen to the results of the research?

Your student's personal details will remain strictly confidential. Research findings made available in any reports or publications will not include information that can directly identify any of your students.

The results of this study will be written up as part of my doctorate. It may be published. XXX College will receive a copy of the results following the study. You can obtain a copy of the results and a one-page summary of the study yourself by emailing g.muncer@soton.ac.uk.
This will be available from 01.09.2020.

Where can I get more information?

If you have any questions or would like more information about the study, please email one of the research team at:

Gemma Muncer at g.muncer@soton.ac.uk

J. Hadwin at X

P. Higham at X

What happens if there is a problem?

If you have a concern about any aspect of this study, you should speak to the researchers who will do their best to answer your questions.

If you remain unhappy or have a complaint about any aspect of this study, please contact the University of Southampton Research Integrity and Governance Manager (023 8059 5058, rgoinfo@soton.ac.uk).

Data Protection Privacy Notice

The University of Southampton conducts research to the highest standards of research integrity. As a publicly-funded organisation, the University has to ensure that it is in the public interest when we use personally-identifiable information about people who have agreed to take part in research. This means that when you agree to take part in a research study, we will use information about you in the ways needed, and for the purposes specified, to conduct and complete the research project. Under data protection law, 'Personal data' means any information that relates to and is capable of identifying a living individual. The University's data protection policy governing the use of personal data by the University can be found on its website (https://www.southampton.ac.uk/legalservices/what-we-do/data- protection-and-foi.page).

This Information Sheet tells you what data will be collected for this project and whether this includes any personal data. Please ask the research team if you have any questions or are unclear what data is being collected about you.

Our privacy notice for research participants provides more information on how the University of Southampton collects and uses your personal data when you take part in one of our research projects and can be found at http://www.southampton.ac.uk/assets/sharepoint/intranet/ls/Public/Research%20and%20Inte grity%20Privacy%20Notice/Privacy%20Notice%20for%20Research%20Participants.pdf

Any personal data we collect in this study will be used only for the purposes of carrying out our research and will be handled according to the University's policies in line with data protection law. If any personal data is used from which you can be identified directly, it will not be disclosed to anyone else without your consent unless the University of Southampton is required by law to disclose it.

Data protection law requires us to have a valid legal reason ('lawful basis') to process and use your Personal data. The lawful basis for processing personal information in this research study is for the performance of a task carried out in the public interest. Personal data collected for research will not be used for any other purpose.

For the purposes of data protection law, the University of Southampton is the 'Data Controller' for this study, which means that we are responsible for looking after your information and using it properly.

Please do email us if you have any questions, something is unclear, or you need more information at:

Ms Gemma Muncer g.muncer@soton.ac.uk

Dr Julie Hadwin X

Dr Phil Higham X

Many thanks,

Gemma, Julie and Phil.

Appendix R Ethical Confirmation

ERGO II – Ethics and Research Governance Online https://www.ergo2.soton.ac.uk

Submission ID: 47186.A1
Submission Title: Exploring the impact of successive relearning on learning and motivation in a post-16 GCSE Mathematics qualification. (Amendment 1)
Submitter Name: Gemma Muncer

Your submission has now been approved by the Faculty Ethics Committee. You can begin your research unless you are still awaiting any other reviews or conditions of your approval.

Comments:

Click here to view the submission

TId:
23011_Email_to_submitter___Approval_from_Faculty_Ethics_committee__cat_B___C_Id: 118715 G.Muncer@soton.ac.uk coordinator

Appendix S Participant Consent Statement

I agree to participate in the research study. I understand the purpose and nature of this study and I am participating voluntarily. I understand that I can withdraw from the study at any time, without any penalty or consequences during the study, and after the study I can withdraw my data up until 01.11.19. Please select your answer and press the blue arrow at the bottom to continue.

Appendix T Participant Debrief Sheet

Study Title: *Exploring the impact of restudy versus retesting on learning and motivation in a post-16 GCSE Mathematics qualification.*

Researcher names Ms Gemma Muncer, Dr Julie Hadwin and Dr Phil Higham. **Ethics number**: 47186

Dear Student,

Thank you very much for taking part in this research; we hope that it was helpful for your learning and in preparing you for your GCSE in maths. We would like you to understand why you were asked to take part. We are working with your College to understand how we can adjust teaching methods in maths to help you and other students to achieve pass grades when you are re-sitting your Maths GCSE. We will look at whether different revision methods are linked to test results in Maths.

We also wanted to understand whether you felt that any change in revision also changed how you felt about your own ability to achieve in Maths. We asked you questions in relation to how well you think you can concentrate in class, how well you feel you can learn the material, and how much you worry in general and about exams.

Understanding how teaching is linked to the student experience is important, because it will help us to support you and other students to be confident about your achievements in College.

If there is anything that you are worried about after taking part in this research, please discuss this with an adult, at school or at home, that you feel comfortable with. You can also contact us on email, (Gemma Muncer, Julie Hadwin, and Phil Higham: g.muncer@soton.ac.uk X and Xand we can arrange to talk about it over the phone or in person.

Some of the questionnaires explored feelings of worry and anxiety. In addition to sources of support within the school, you might also find the following websites about how to manage anxiety useful:

- https://kooth.com/
- https://www.samaritans.org/education/deal/coping-strategies/exam-stress
- https://www.childline.org.uk/info-advice/school-college-and-work/

I would like to remind you that you have the right to withdraw your data, by emailing one of us up until 1st November, 2019. If you do so, this means that the information you have provided will not be used in my report. If you do not contact me, then your data will be used anonymously, which means that no names will be recorded, and it is not possible to identify you through the write-up of my research.

Thank you again for taking part in this project.

Best wishes,

Gemma Muncer, Julie Hadwin, and Phil Higham

If you have questions about your rights as a participant in this research, or if you feel that you have been placed at risk, you may contact the Chair of the Ethics Committee, Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: +44 (0)23 8059 3856, email fshs-rso@soton.ac.uk

Appendix U Descriptive Data of all Participant Responses at T1 and at T2

Table A6

The Means, Standard Deviations and Ranges for all Available Participant Responses at T1 and T2

	Relearn		Restudy		Control		All Participants	
Variable	T1	T2	T1	T2	T1	T2	T1	T2
	M (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]
Attitudes toward maths	110.83 (23.78) [55-170] *n = 59*	111.79 (21.01) [77-158] *n = 29*	109.82 (23.96) [61-170] *n = 61*	110.21 (29.48) [53-176] *n = 42*	110.03 (23.42) [62-168] *n = 35*	120.10 (24.26) [86-168] *n = 21*	110.40 (23.62) [55-170] *n = 156*	112.97 (25.93) [53-176] *n = 92*
Academic efficacy	16.03 (3.84) [7-25] *n = 59*	15.41 (3.48) [10-23] *n = 29*	15.16 (4.14) [6-25] *n = 61*	15.45 (4.20) [7-24] *n = 42*	15.21 (3.87) [7-24] *n = 33*	16.10 (3.95) [9-25] *n = 21*	15.52 (3.96) [6-25] *n = 154*	15.59 (3.90) [7-25] *n = 92*

	Relearn		Restudy		Control		All Participants	
Variable	T1	T2	T1	T2	T1	T2	T1	T2
	M (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]
Attentional control	49.47 (7.89) [32-68] *n = 60*	48.10 (7.94) [30-60] *n = 30*	48.05 (9.06) [26-67] *n = 62*	49.14 (8.26) [34-71] *n = 43*	49.06 (10.53) [25-78] *n = 33*	49.55 (10.36) [23-66] *n = 20*	48.80 (8.91) [25-78] *n = 156*	48.89 (8.57) [23-71] *n = 93*
Anxiety *General*	9.44 (4.85) [0-18] *n = 59*	9.31 (4.77) [2-17] *n = 29*	10.00 (5.07) [0-18] *n = 61*	10.74 (4.35) [2-17] *n = 42*	8.88 (4.81) [2-18] *n = 34*	7.73 (3.64) [3-15] *n = 22*	9.55 (4.90) [0-18] *n = 155*	9.58 (4.45) [2-17] *n = 93*
Maths	21.71 (8.16) [9-44] *n = 59*	19.93 (8.30) [9-39] *n = 30*	21. 36 (9.74) [9-45] *n = 61*	23.36 (9.65) [9-45] *n = 42*	18.88 (7.98) [9-36] *n = 34*	19.86 (7.46) [9-40] *n = 21*	20.93 (8.77) [9-45] *n = 155*	21.46 (8.85) [9-45] *n = 93*

	Relearn		Restudy		Control		All Participants	
Variable	T1	T2	T1	T2	T1	T2	T1	T2
	M (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]
Predicted test score	20.33 (5.53) [9-36] *n = 55*	19.61 (6.43) [1-30] *n = 28*	19.58 (5.30) [8-36] *n = 57*	19.71 (6.70) [4-31] *n = 41*	20.06 (6.83) [5-36] *n = 32*	18.60 (4.96) [10-26] *n = 20*	19.97 (5.72) [5-36] *n = 144*	19.43 (6.21) [1-31] *n = 89*
Actual test score	N/A	20.30 (6.37) [8-33] *n = 43*	N/A	21.04 (5.44) [9-33] *n = 48*	N/A	20.26 (6.47) [10-36] *n = 23*	N/A	20.61 (5.97) [8-36] *n = 114*
Difference between predicted and actual test Score	5.68 (5.56) [0-22] *n = 31*	4.93 (4.53) [0-19] *n = 28*	5.97 (3.79) [0-15] *n = 37*	5.37 (3.64 [0-16] *n = 35*	6.17 (6.38) [0-21] *n = 12*	6.39 (3.87) [0-15] *n = 18*	5.89 (4.90) [0-22] *n = 80*	5.44 (4.00) [0-19] *n = 81*

	Relearn		Restudy		Control		All Participants	
Variable	T1	T2	T1	T2	T1	T2	T1	T2
	M (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]	*M* (SD) [Range]
Attendance (%)	N/A	81.33 (20.97) [0-100] *n = 75*	N/A	83.47 (20.20) [5-100] *n = 73*	N/A	85.47 (15.43) [33.33-100] *n = 43*	N/A	83.09 (19.52) [0-100] *n = 191*

Note. T1 = Time 1 (before intervention), T2 = Time 2 (post-intervention), *M* = mean, SD = standard deviation, *n* = number of participants

Appendix V Coding Manual

Table A7

Descriptions of the Final Themes, Sub-Themes and Subordinate Sub-Themes, Alongside Example Participant Quotes

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
1. Engagement with revision sessions	1.1 Time constraints	Reference to not having engaged in session(s) due to a lack of time or due to engaging in other commitments.	Extra-curricular activities/ hobbies	Reference to not having completed session(s) due to being busy engaging in extra-curricular activities or hobbies.	"I was attending cadets" "football" "dance"
			Paid work	Reference to not having completed session(s) due to engaging in paid employment.	"work straight after college so was unable to find the time to complete them."
			Prioritising other college work	Reference to not having completed session(s) due to completing other college work including homework and	"Too busy with assignments"

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
				assignments (i.e. coursework), both in maths or related to other subjects.	“Didn't have time as well as other homeworks” “coursework” “my three already looming subjects stresses me out exponentially”
			Lack of flexibility of session days	Reference to not having completed session(s) due to being unavailable on the days which sessions were set to complete sessions e.g. due to a special event such as a birthday, by a regular event such as visiting a family member on certain days or finishing college later on certain days, or by being unwell.	“Away at dads and didn’t have time on some Fridays” “I did not have time on that evening” “My birthday” “Not set on great days”

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
	1.2 Issues accessing sessions online	Reference to not having engaged in session(s) due to difficulties in accessing online sessions. This could be related to hardware e.g. a difficulty with a mobile phone, or with software e.g. a difficulty with an internet connection.			“didnt have access because it wouldnt load up” “when i tried to log onto it my internet stoped working” “I often lose my phone”
	1.3 Monitoring and awareness	Reference to not having engaged in session(s) due to either not being aware of the sessions, or not having checked emails for the session links.	Not aware of the sessions	Reference to not having completed session(s) due to not being aware that the sessions were available.	“Missed the lesson telling us we needed to do them” “I didn’t know about them” “I was not informed of these revision sessions”

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
			Not checking emails	Reference to not having completed session(s) due to not checking emails for the session links.	"I don't check my emails and when I do the email got pushed down to the bottom."
	1.4 Motivation	Reference to not having engaged	Fell asleep/too tired	Reference to not having completed session(s) due to being too tired or falling asleep.	"I was too tired to concentrate, and I understood that I needed to sleep" "went to sleep"
			Did not want to engage	Any reference to not having completed session(s) due to not wanting to.	"I didn't want to do them"
			Forgot	Any reference to not having completed session(s) due to generally having forgotten.	"Just forgot"
2. Feelings about the sessions	2.1 Perceptions of the helpfulness	Any reference to how helpful the sessions were.	Positive and negative views of general helpfulness	Any reference to the sessions being generally helpful or unhelpful.	"I think they were useful." "very helpful"

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
	of the sessions				
					“kinda pointless”
			Opportunity to revisit lesson content	Reference to the sessions providing an opportunity to revise or re-visit basic concepts.	“I think they are useful, as you get to revise certain topics and practise them”
					“They were helpful in the fat they taught be some of the basics of the facts and figures I should know, alongside the names of things”
			Feeling indifferent/ unsure	Any reference to feeling generally unsure or indifferent about the sessions.	“I still don’t think I’m that much better at maths than I was before I started revising them”
					“meh, they were okay”
					“it was alright”

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
	2.2 Interest/ enjoyment of the sessions	Any reference to interest or enjoyment in the sessions, both positive and negative.	Positive and negative views of general enjoyment	Any positive or negative general comments regarding students' enjoyment of the sessions or whether or not they found the sessions fun.	"amazing" "boring" "They were really fun and interesting"
			Sessions too long	Reference to finding the sessions too long.	"way too long" "They took to long to do"
			Sessions too slow/ too quick	Reference to the pace of the sessions including finding the sessions too quick or finding the sessions too slow.	"The first one I took was very slow, and made it frustrating to complete" "some took too long to complete in time"

Theme	Sub-theme	Sub-theme description	Subordinate sub-theme	Subordinate sub-theme description	Example quote(s)
			Feeling indifferent/ unsure	Any reference to feeling generally unsure or indifferent about the sessions.	“meh, they were okay” “it was alright”

List of References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., D'Antonio, L., Roediger, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, *3*(3), 131–139. https://doi.org/10.1016/j.jarmac.2014.07.002

Ahmed, W., van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology*, *105*(1), 150–161. https://doi.org/10.1037/a0030160

Anderson, O., & Krathwhol, D. R. (2001). *A taxonomy for learning. teaching, and assessing (a revision of Bloom's taxonomy of educational objectives).* Addision Wesley Longman.

Aschbacher, P. R., Koency, G., & Schacter, J. (1995). *Los Angeles Learning Center alternative assessments guidebook (resource paper no. 12*). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Aşık, G., & Erktin, E. (2019). Metacognitive experiences: Mediating the relationship between metacognitive knowledge and problem solving. *Egitim ve Bilim*, *44*(197), 85–103. https://doi.org/10.15390/EB.2019.7199

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology*, *108*(3), 296–308. http://dx.doi.org/10.1037/0096-3445.108.3.296

Baker, L. (1989). Metacognition, comprehension monitoring, and the adult reader. *Educational Psychology Review, 1*(1), 3–38. https://doi.org/10.1007/BF01326548

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213-238. https://doi.org/10.3102/00346543061002213

Barzagar Nazari, K., & Ebersbach, M. (2018). Distributed practice: Rarely realized in self-regulated mathematical learning. *Frontiers in Psychology*, *9*, 2170. https://doi.org/10.3389/fpsyg.2018.02170

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088-1101. https://doi.org/10.2307/2533446

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, *61*(3), 228–247. https://doi.org/10.1016/j.cogpsych.2010.05.004

Binks, S. (2017). Testing enhances learning: A review of the literature. *Journal of Professional Nursing*, *34*(2018), 205-210. https://doi.org/10.1016/j.profnurs.2017.08.008

Bishara, S., & Kaplan, S. (2018). The relationship of locus of control and metacognitive knowledge of math with math achievements. *International Journal of Disability, Development and Education*, *65*(6), 631–648. https://doi.org/10.1080/1034912X.2018.1432033

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-207). MIT Press. https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/07/RBjork_1994a.pdf

Bjork, R. (2017). *Creating desirable difficulties to enhance learning*. Crown House Publishing.

Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society (2nd edition)* (pp. 59-68). Worth. https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/11/Making-Things-Hard-on-Yourself-but-in-a-Good-Way-20111.pdf

Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook 1: Cognitive domain.* David McKay Company.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive meta-analysis (version 3)*. Biostat.

Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). *Answers to frequently asked questions about thematic analysis*. Thematic Analysis: A Reflexive Approach. https://cdn.auckland.ac.nz/assets/psych/about/our-research/documents/Answers%20to%20frequently%20asked%20questions%20about%20thematic%20analysis%20April%202019.pdf

Buckley, P., & Doyle, E. (2016). Gamification and student motivation. *Interactive Learning Environments, 24*(6), 1162-1175. https://doi.org/10.1080/10494820.2014.964263

Callan, G. L., & Cleary, T. J. (2019). Examining cyclical phase relations and predictive influences of self-regulated learning processes on mathematics task performance. *Metacognition and Learning*, *14*(1), 43–63. https://doi.org/10.1007/s11409-019-09191-x

Callan, G. L., Marchant, G. J., Finch, W. H., & German, R. L. (2016). Metacognition, strategies, achievement, and demographics: Relationships across countries. *Kuram ve Uygulamada Egitim Bilimleri*, *16*(5), 1485–1502. https://doi.org/10.12738/estp.2016.5.0137

Campbell, J. I. (2005). *Handbook of mathematical cognition*. Psychology Press.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369-378. https://doi.org/10.1007/s10648-012-9205-z

Cepeda, N. J., Pashler, H., Vul, E., Rohrer, D., & Wixted, J. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380. https://doi.org/10.1037/0033-2909.132.3.354

Çetinkaya, P. (2000). *Metacognition: Its assessment and relationship with reading comprehension, achievement, and aptitude for sixth grade student* [Master's dissertation]. Boğaziçi University.

Çetinkaya, P., & Erktin, E. (2002). Assessment of metacognition and its relationship with reading comprehension, achievement, and aptitude. *Boğaziçi Üniversitesi Eğitim Dergisi*, *19*(1), 1-11.

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educational Psychology Review*, *30*(2), 483–501. https://doi.org/10.1007/s10648-017-9426-2

Chiu, M. M., Chow, B. W. Y., & Mcbride-Chang, C. (2007). Universals and specifics in learning strategies: Explaining adolescent mathematics, science, and reading achievement across 34 countries. *Learning and Individual Differences*, *17*(4), 344–365. https://doi.org/10.1016/j.lindif.2007.03.007

Cipora, K., Willmes, K., Szwarc, A., & Nuerk, H.-C. (2018). Norms and validation of the online and paper-and-pencil versions of the abbreviated math anxiety scale (AMAS) for Polish adolescents and adults. *Journal of Numerical Cognition*, *3*(3), 667–693. https://doi.org/10.5964/jnc.v3i3.121

Clarke, V., & Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research*, *18*(2), 107–110. https://doi.org/10.1002/capr.12165

Crawford, C., & Cribb, J. (2013). *Reading and maths skills at age 10 and earnings in later life: a brief analysis using the British cohort study.* Institute of Education, University of London.

Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the motivated strategies for learning questionnaire. *Learning and Individual Differences*, *21*(4), 337–346. https://doi.org/10.1016/j.lindif.2011.03.002

Critical Appraisal Skills Programme (2019). *CASP cohort study checklist*. CASP Checklists. https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (pp. 416–436). Sage Publications Ltd. https://doi.org/10.4135/9781446249215.n21

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of Learning and Motivation - Advances in Research and Theory* (pp. 63–147). Elsevier Inc. https://doi.org/10.1016/S0079-7421(10)53003-2

Dennison, R. S., Krawchuk, C. M., Howard, B. C., & Hill, L. (1996, April 8-11). *The development of a children's self-report measure of metacognition.* Annual meeting of the American Educational Research Association, New York.

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, *28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Department for Education (2020). *Revised A level and other 16 to 18 results in England, 2018/2019*. Key Stage 5 Exam Marking, Qualifications and Results. *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/859515/2019_revised_A-Level_and_other_16_to_18_results_in_England.pdf*

Dermitzaki, I. (2005). Preliminary investigation of relations between young students' self-regulatory strategies and their metacognitive experiences. *Psychological Reports*, *97*(3), 759-768. https://doi.org/10.2466%2Fpr0.97.3.759-768

Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*, *111*(2), 225–236. https://doi.org/10.1037/0021-843X.111.2.225

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction*, *16*(1), 12–25. https://doi.org/10.1016/j.learninstruc.2005.12.003

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, *34*(5), 435-447. https://doi.org/10.1177/002221940103400505

Devine, A., Fawcett, K., Szucs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*(33), 1-9. https://doi.org/10.1186/1744-9081-8-33

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. https://doi.org/10.1007/s11409-008-9029-x

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795-805. https://doi.org/10.1037/0021-9010.84.5.795

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions. *Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529

Education and Skills Funding Agency (2019). *16 to 19 funding: maths and English condition of funding.* Funding for 16 to 19 Year Olds in Schools https://www.gov.uk/guidance/16-to-19-funding-maths-and-english-condition-of-funding

Efklides, A. (2006). Metacognitive experiences: The missing link in the self-regulated learning process. *Educational Psychology Review*, *18*(3), 287–291. https://doi.org/10.1007/s10648-006-9021-4

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, *13*(4), 277-287. https://doi.org/10.1027/10169040.13.4.277.

Efklides, A., and Vlachopoulos, S. P. (2012). Measurement of metacognitive knowledge of self, task, and strategies in mathematics. *European Journal of Psychological Assessment, 28*(3), 227–239. https://doi.org/10.1027/1015-5759/a000145

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Erktin, E. (2004). Teaching thinking for mathematics through the enhancement of metacognitive skills. *Research in the Schools, 11*(1), 3–13.

Fadlelmula, F. K., Cakiroglu, E., & Sungur, S. (2015). Developing a structural model on the relationship among motivational beliefs, self-regulated learning strategies, and achievement in mathematics. *International Journal of Science and Mathematics Education*, *13*(6), 1355-1375. https://doi.org/10.1007/s10763-013-9499-4

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fitzpatrick, C. (1994). *Adolescent mathematical problem solving: The role of metacognition, strategies and beliefs* (ED374969). ERIC. https://eric.ed.gov/?id=ED374969

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Fusco, D. R. (1995). *The role of strategy, effort and unknown attributions in a metacognitive model of mathematical problem solving* [Doctoral dissertation, The City University of New York]. ProQuest Digital Collections. https://search.proquest.com/pqdtglobal/docview/304181779/19A59985BBA0439CPQ/1?accountid=13963

Garcia-Lopez, L. J., Saez-Castillo, A. J., & Fuentes-Rodriguez, G. (2017). Psychometric properties of the youth anxiety measure for DSM-5, part I (YAM-5-I) in a community sample of Spanish-speaking adolescents. *Journal of Affective Disorders*, *208*, 455–459. https://doi.org/10.1016/j.jad.2016.10.045

Garrett, A. J., Mazzocco, M. M. M., & Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learning Disabilities Research & Practice*, *21*(2), 77-88. https://doi.org/10.1111/j.1540-5826.2006.00208.x

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. https://doi.org/10.1002/rev3.3077

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*. Routledge.

Haddad Center. (2012). *Battery math tests*. Bar Ilan University.

Harris, M. M. (2015). *The role of metacognition in a montessori environment and the effects on academic achievement* [Doctoral dissertation, Union University]. ProQuest Digital Collections. https://search.proquest.com/docview/1733667442?accountid=13963

Hassan, N. M., & Rahman, S. (2017). Problem solving skills, metacognitive awareness, and mathematics achievement: A mediation model. *The New Educational Review*, *49*(3), 201-212.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Hemmings, B., & Kay, R. (2010). Prior achievement, effort, and mathematics attitude as predictors of current achievement. *The Australian Educational Researcher, 37*(2), 41-58. https://doi.org/10.1007/BF03216921

Henfi, J. (1990). *Redactiesommen.* Ajodakt.

Higgins, S., Hall, E., Baumfield, V., & Moseley, D. (2005). *A meta-analysis of the impact of the implementation of thinking skills approaches on pupils*. Thinking Skills Review Group. http://dro.dur.ac.uk/1890/

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Higham, P. A., Zengel, B., Bartlett, L., & Hadwin, J. A. (2019). *The benefits of successive relearning on multiple learning outcomes*. Manuscript in Preparation.

Hong, E., & Peng, Y. (2004). *Test-taking strategies questionnaire*. Unpublished document, University of Nevada.

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. *Educational Research Review*, *17*, 63-84. https://doi.org/10.1016/j.edurev.2015.11.002

Hopkins, R. F., Lyle, K. B., Hieb, J. L, & Ralston, P. A. S. (2016). Spaced retrieval practice increases students' short- and long-term retention of mathematics knowledge.

Educational Psychology Review, 28, 857-873. https://doi.org/10.1007/s10648-015-9349-8

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment, 10*(2), 178-182. https://doi.org/10.1177/1073191103252351

Hughes, G. I., Taylor, H. A., & Thomas, A. K. (2018). Study techniques differentially influence the delayed judgment-of-learning accuracy of adolescent children and college-aged adults. *Metacognition and Learning*, *13*(2), 109–126. https://doi.org/10.1007/s11409-018-9180-y

Ichihara, M., & Arai, K. (2006). Moderator effects of meta-cognition: A test in math of a motivational model. *Japanese Journal of Educational Psychology*, *54*(2), 199-210.

Janes, J. L., Dunlosky, J., Rawson, K. A., & Jasnow, A. (2020). Successive relearning improves performance on a high-stakes exam in a difficult biopsychology course. Applied Cognitive Psychology. Accepted Manuscript. https://doi.org/10.1002/acp.3699

Johnston-Wilder, S., Lee, C., Brindley, J., & Garton, E. (2015, November 16-18). *Developing mathematical resilience in school-students who have experienced repeated failure*. Annual International Conference of Education, Research and Innovation (ICERI2015), Seville, Spain. https://www.researchgate.net/profile/Sue_Johnston-Wilder/publication/315741077_DEVELOPING_MATHEMATICAL_RESILIENCE_IN_SCHOOL-STUDENTS_WHO_HAVE_EXPERIENCED_REPEATED_FAILURE/links/58e103cfaca272059ab06370/DEVELOPING-MATHEMATICAL-RESILIENCE-IN-SCHOOL-STUDENTS-WHO-HAVE-EXPERIENCED-REPEATED-FAILURE.pdf

Knowles, M. S. (1978). Andragogy: Adult learning theory in perspective. *Community College Review*, *5*(3), 9-20.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, *17*(5), 493–501. https://doi.org/10.1080/09658210902832915

Kramarski, B., & Mizrachi, N. (2006). Online discussion and self-regulated learning: Effects of instructional methods on mathematical literacy. *Journal of Educational Research, 99*(4), 218–231. https://doi.org/10.3200/JOER.99.4.218-231

Kramarski, B., Rich, I., Mevarech, Z., & Libereman, A. (2005). *The effect of metacognitive processes on achievement motivation and mathematical thinking among students in middle schools*. Bar Ilan University.

Lucangeli, D., & Cornoldi, D. (1997). Mathematics and metacognition: What is the nature of the relationship? *Mathematical Cognition*, *3*(2), 121-139. https://doi.org/10.1080/135467997387443

Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*(2), 94-97. https://doi.org/10.1177/0098628311401587

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence*, *27*(2), 165–179. https://doi.org/10.1016/j.adolescence.2003.11.003

Maras, K., Gamble, T., & Brosnan, M. (2019). Supporting metacognitive monitoring in mathematics learning for young people with autism spectrum disorder: A classroom-based study. *Autism*, *23*(1), 60–70. https://doi.org/10.1177/1362361317722028

Martin, R., Hodgson, H., Maloney, A. & Rayner, I. (2014). *Pro bono economics report for national numeracy: Cost of outcomes associated with low levels of adult numeracy in the UK*. National Numeracy. https://www.probonoeconomics.com/sites/default/files/files/PBE%20National%20Numeracy%20costs%20report%2011Mar.pdf

Martín, E., Martínez-Arias, R., Marchesi, A., & Pérez, E. M. (2008). Variables that predict academic achievement in the Spanish compulsory secondary educational system: A longitudinal, multi-level analysis. *The Spanish Journal of Psychology*, *11*(2), 400–413. https://doi.org/10.1017/S113874160000442X

Mathematics Diagnostic Testing Project (2006). *The MDTP assessment system*. The California State University/University of California Mathematics Diagnostic Testing Project. https://mdtp.ucsd.edu/assessments/index.html

Maxwell, S. E. & Delaney, H. D. (2004). *Designing experiments and analyzing data (2nd edn.).* Lawrence Erlbaum Associates.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, *26*(1-2), 49-63. https://doi.org/10.1023/A:1003088013286

Meltzer, L. J., Levine, M. D., Karniski, W., Palfrey, J. S., & Clarke, S. (1984). An analysis of the learning styles of adolescent delinquents. *Journal of Learning Disabilities*, *17*(10), 600-608. https://doi.org/10.1177/002221948401701006

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review*, *15*(1), 174–179. https://doi.org/10.3758/PBR.15.1.174

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., & Urdan, T. (2000). *Manual for the patterns of adaptive learning scales*. University of Michigan.

Mirza, A., & Hussain, N. (2018). Performing below the targeted level: An investigation into KS3 pupils' attitudes towards mathematics. *Journal of Education and Educational Development*, *5*(1), 8-24. http://dx.doi.org/10.22555/joeed.v5i1.1390

Missouri Department of Elementary and Secondary Education (1990). *Mastery and achievement tests: Guide to score, interpretation and Use*. Jefferson City.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., ... Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Chinese Integrative Medicine*, *7*(9), 889-896. https://doi.org/10.3736/jcim20090918

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Revista Espanola de Nutricion Humana y Dietetica*, *20*(2), 148–160. https://doi.org/10.1186/2046-4053-4-1

Montague, M., & Bos, C. S. (1987). Cognitive and metacognitive characteristics of eighth grade students' mathematical problem solving. *Learning and Individual Differences*, *2*(3), 371–388. https://doi.org/10.1016/1041-6080(90)90012-6

Moran, T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin*, *142*, 831–864. http://dx.doi.org/10.1037/bul0000051

Moreno, A. (2002). The assessment of metacognitive skills. In A. Marchesi & E. Martín (Eds.), *Secondary education assessment* (pp. 119-136). Editorial SM.

Muris, P., Simon, E., Lijphart, H., Bos, A., Hale, W., & Schmeitz, K. (2017). The youth anxiety measure for DSM-5 (YAM-5): development and first psychometric evidence of a new scale for assessing anxiety disorders symptoms of children and adolescents. *Child Psychiatry & Human Development*, *48*(1), 1-17. https://doi.org/10.1007/s10578-016-0648-1

National Numeracy. (2020). *What is numeracy?* National Numeracy. https://www.nationalnumeracy.org.uk/what-numeracy

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–141). Academic. https://doi.org/10.1016/S0079-7421(08)60053-5

Ning, H. K. (2016). Examining heterogeneity in student metacognition: A factor mixture analysis. *Learning and Individual Differences*, *49*, 373–377. https://doi.org/10.1016/j.lindif.2016.06.004

Ofqual (2019). *An infographic: GCSEs in 2019.* Key Stage 3 and 4 Exam Marking, Qualifications and Results. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/826795/GCSE_infographic_17_1_.pdf

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition and Learning*, *13*(2), 179–212. https://doi.org/10.1007/s11409-018-9183-8

Ólafsson, R. P., Smári, J., Gudmundsdóttir, F., Ólafsdóttir, G., Hardardóttir, H. L., & Einarsson, S. M. (2011). Self reported attentional control with the attentional control scale: Factor structure and relationship with symptoms of anxiety and depression. *Journal of Anxiety Disorders*, *25*(6), 777–782. https://doi.org/10.1016/j.janxdis.2011.03.013

O'Neil, H. F., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *The Journal of Educational Research*, *89*(4), 234-245. https://doi.org/10.1080/00220671.1996.9941208

Oppong Asante, K. (2012). Secondary students' attitudes towards mathematics. *IFE Psychologia: An International Journal*, *20*(1), 121-133.

Özcan, Z. Ç. (2016). The relationship between mathematical problem-solving skills and self-regulated learning through homework behaviours, motivation, and metacognition. *International Journal of Mathematical Education in Science and Technology*, *47*(3), 408–420. https://doi.org/10.1080/0020739X.2015.1080313

Özcan, Z. Ç., & Eren Gümüş, A. (2019). A modeling study to explain mathematical problem-solving performance through metacognition, self-efficacy, motivation, and anxiety. *Australian Journal of Education*, *63*(1), 116–134. https://doi.org/10.1177/0004944119840073

Özsoy, G. (2005). The relationship between problem solving skills and mathematical achievement. *Gazi University Journal of Education*, *25*(3), 179-190.

Özsoy, G. (2011). An investigation of the relationship between metacognition and mathematics achievement. *Asia Pacific Education Review*, *12*(2), 227–235. https://doi.org/10.1007/s12564-010-9129-6

Panaoura, A., & Philippou, G. (2003). *The construct validity of an inventory for the measurement of young pupils' metacognitive abilities in mathematics* (ED501054). ERIC. https://eric.ed.gov/?id=ED501054

Parsons, S (2002). *Basic skills and crime.* Basic Skills Agency. https://discovery.ucl.ac.uk/id/eprint/1566250/

Parsons, S. & Bynner, J. (2005). *Does numeracy matter more?* National Research and Development Centre for Adult Literacy and Numeracy. https://dera.ioe.ac.uk/22317/1/doc_2979.pdf

Pastötter, B., Weber, J., & Bäuml, K. H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, *27*(2), 280–285. https://doi.org/10.1037/a0031797

Pearson Education. (2008). *AIMSweb® benchmark and progress monitoring system for Grades K–8*. Bloomington.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, *88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Peng, Y., Hong, E., & Mason, E. (2014). Motivational and cognitive test-taking strategies and their influence on test performance in mathematics. *Educational Research and Evaluation*, *20*(5), 366–385. https://doi.org/10.1080/13803611.2014.966115

Perry, J., Lundie, D., & Golder, G. (2019). Metacognition in schools: What does the literature suggest about the effectiveness of teaching metacognition in schools? *Educational Review*, *7*(4), 483-500. https://doi.org/10.1080/00131911.2018.1441127

Pintrich, P. R. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ).* (ED338122). ERIC. https://eric.ed.gov/?id=ED338122

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–667. https://doi.org/10.1037/a0033194

Price, G., & Ansari, D. (2013). Dyscalculia: Characteristics, causes, and treatments. *Numeracy*, *6*(1), 1-16. https://doi.org/10.5038/1936-4660.6.1.2

Rawson, K. A., & Dunlosky, J. (2013). Relearning attenuates the benefits and costs of spacing. *Journal of Experimental Psychology: General*, *142*(4), 1113–1129. https://doi.org/10.1037/a0030498

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, *25*(4), 523–548. https://doi.org/10.1007/s10648-013-9240-4

Reed, H. C., Kirschner, P. A., & Jolles, J. (2015). Self-beliefs mediate math performance between primary and lower secondary school: A large-scale longitudinal cohort study. *Frontline Learning Research*, *3*(1), 36-54.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353.

Robbins, S. B., Le, H., Davis, D., Lauver, K., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288. https://doi.org/10.1037/0033-2909.130.2.261

Roebers, C. M., Krebs, S. S., & Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: Their interrelations and their role for test performance. *Learning and Individual Differences*, *29*, 141–149. https://doi.org/10.1016/j.lindif.2012.12.003

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181- 210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*(9), 1209–1224. https://doi.org/10.1002/acp.1266

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*(6), 481–498. https://doi.org/10.1007/s11251-007-9015-8

Saraç, S., & Karakelle, S. (2012). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education*, *4*(2), 301-315.

Sato, J., & Arai, K. (1998). The relation between the use of learning strategies, learning goals and causal attributions. *Tsukuba Psychological Research*, *20*, 115-124.

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology*, *19*(2), 143-154 https://doi.org/10.1006/ceps.1994.1013

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460-475.

Schraw G., Olafson L., Weibel M. & Sewing, D. (2012). Metacognitive knowledge and field-based science learning in an outdoor environmental education program. In A. Zohar & Y. Dori (Eds.), *Metacognition in science education: Trends in current research* (pp. 57-77). Springer. https://doi.org/10.1007/978-94-007-2132-6_4

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*, 747–770. https://doi.org/10.1146/annurev-psych-010418

Simon, E., Bos, A. E. R., Verboon, P., Smeekens, S., & Muris, P. (2017). Psychometric properties of the youth anxiety measure for DSM-5 (YAM-5) in a community sample. *Personality and Individual Differences*, *116*, 258–264. https://doi.org/10.1016/j.paid.2017.04.058

Sink, C. A., Barnett, J. E., & Hixon, J. E. (1991, April 3-7). *Self-regulated learning and academic performance in middle school children.* Annual meeting of the American Educational Research Association, Chicago.

Smith, A. (2017). *Report of Professor Sir Adrian Smith's review of post-16 mathematics.* Department for Education. http://hdl.voced.edu.au/10707/438828.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology*, *27*(1), 51–79. https://doi.org/10.1006/ceps.2001.1091

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *International Journal of Phytoremediation*, *21*(1), 117–139. https://doi.org/10.1076/edre.10.2.117.27905

Stahl, E., Pieschl, S., & Bromme, R. (2006). Task complexity, epistemological beliefs and metacognitive calibration: An exploratory study. *Journal of Educational Computing Research*, *35*(4), 319-338. https://doi.org/10.2190/1266-0413-387K-7J51

Steinmayr, R., Ziegler, M., & Träuble, B. (2010). Do intelligence and sustained attention interact in predicting academic achievement? *Learning and Individual Differences, 20*(1), 14–18. https://doi.org/10.1016/j.lindif.2009.10.009

The Sutton Trust (2018). *Best in class. Evidence, funding and school reform in England and the United States.* https://www.suttontrust.com/wpcontent/uploads/2019/12/BestinClass2018-Formatted.pdf

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tapia, M. (1996). *The attitudes toward mathematics instrument*. (ED404165). ERIC. https://eric.ed.gov/?id=ED404165

Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, *8*(2), 16-22.

Tian, Y., Fang, Y., & Li, J. (2018). The effect of metacognitive knowledge on mathematics performance in self-regulated learning framework-multiple mediation of self-efficacy and motivation. *Frontiers in Psychology, 9*, 2518. https://doi.org/10.3389/fpsyg.2018.02518

Tobias, S., Everson, H. T., & Laitusis, V. (1999). *Towards a performance-based measure of metacognitive knowledge monitoring: Relationships with self-reports and behaviour ratings.* (ED432590). ERIC. https://eric.ed.gov/?id=ED432590

Tosto, M. G., Asbury, K., Mazzocco, M. M. M., Petrill, S. A., & Kovas, Y. (2016). From classroom environment to mathematics achievement: The mediating role of self-perceived ability and subject interest. *Learning and Individual Differences, 50,* 260–269. https://doi.org/10.1016/j.lindif.2016.07.009

Tyler, K. M., Burris, J. L, & Coleman, S. T. (2016). Investigating the association between home-school dissonance and disruptive classroom behaviours for urban middle school students. *The Journal of Early Adolescence, 38(*4), 530-553. https://doi.org/10.1177/0272431616678987

van der Stel, M., & Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: A longitudinal study from a developmental perspective. *European Journal of Psychology of Education*, *29*(1), 117–137. https://doi.org/10.1007/s10212-013-0190-5

van der Stel, M., Veenman, M. V. J., Deelen, K., & Haenen, J. (2010). The increasing role of metacognitive skills in math: A cross-sectional study from a developmental perspective. *International Journal on Mathematics Education*, *42*(2), 219–229. https://doi.org/10.1007/s11858-009-0224-2

van der Walt, M. S., Maree, J. G., & Ellis, S. M. (2008). Metacognition in the learning of mathematics in the senior phase :Some implications for the curriculum. *International Journal of Adolescence and Youth*, *14*(3), 205–235. https://doi.org/10.1080/02673843.2008.9748004

Vaughn, K. E., Dunlosky, J., & Rawson, K. A. (2016). Effects of successive relearning on recall: Does relearning override the effects of initial learning criterion? *Memory and Cognition*, *44*(6), 897–909. https://doi.org/10.3758/s13421-016-0606-y

Veenman, M. V. J., Bavelaar, L., De Wolf, L., & Van Haaren, M. G. P. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences*, *29*, 123–130. https://doi.org/10.1016/j.lindif.2013.01.003

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress and Coping*, *13*(4), 391–412. https://doi.org/10.1080/10615800008248343

Veenman, M. V. J., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science*, *33*(3), 193–211. https://doi.org/10.1007/s11251-004-2274-8

Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods. *Mathematics Education*, *51*(4), 691–701. https://doi.org/10.1007/s11858-018-1006-5

Veenman M. V. J. & van Hout-Wolters, B. H. A. M. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In *Abstracts of the 10th biennial meeting of the European Association for Research on Learning and Instruction.* Cooperativa Libraria Editrice Università di Padova. https://hdl.handle.net/11245/1.222029

Veenman, M. V. J., van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*, 3-14. https://doi.org/10.1007/s11409-006-6893-0

Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences*, *15*(2), 159-176. https://doi.org/10.1016/j.lindif.2004.12.001

Vermunt, J. D. H. M., & van Rijswijk, F. A. W. M. (1987). *Inventory of learning styles for higher education.* Catholic University of Brabant.

Verschaffel, L., van Dooren, W., Greer, B., & Mukhopadhyay, S. (2010). Reconceptualising word problems as exercises in mathematical modelling. *Journal Fur Mathematik-Didaktik*, *31*(1), 9–29. https://doi.org/10.1007/s13138-010-0007-x

Walker, E. (2013). *Understanding the role of metacognition and working memory in maths achievement* [Doctoral dissertation, University of Southampton]. University of Southampton Digital Collections. https://eprints.soton.ac.uk/358501/

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test (WRAT4).* Psychological Assessment Resources.

Wolf, A. (2011). *Review of vocational education- The Wolf report*. Department for Education. http://www.educationengland.org.uk/documents/pdfs/2011-wolf-report-vocational.pdf

Wolters, C. A., Pintrich, P. R., & Karabenick, S. A. (2006). Assessing academic self-regulated learning. In K. A. Moore & L. H. Lippman (Eds.), *What do children need to flourish?* (pp. 251–270). Springer. https://doi.org/10.1007/0-387-23823-9_16

Yap, E. G. (1993). *A structural model of self-regulated learning in math achievement* [Doctoral dissertation, University of Southern California]. ProQuest Digital Collections. https://search.proquest.com/docview/1627936537?accountid=13963

Young, A. E., & Worrell, F. C. (2018). Comparing metacognition assessments of mathematics in academically talented students. *Gifted Child Quarterly*, *62*(3), 259–275. https://doi.org/10.1177/0016986218755915

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329