UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

School of Psychology

Understanding the role of metacognition and working memory in maths achievement

by

Emma Walker

Thesis for the degree of Doctorate in Educational Psychology

June 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

Doctorate in Educational Psychology

UNDERSTANDING THE ROLE OF METACOGNITION AND WORKING MEMORY IN MATHS ACHIEVEMENT

By Emma Jayne Walker

Recent increases in youth unemployment have driven reforms to educational provision. This has created an increased focus on gaining a formal qualification in maths because it affects access to jobs/course places. These governmental reforms have placed a heightened responsibility on schools to provide appropriate interventions to meet pupils' needs. Theoretical models developed to understand what affects maths achievement have highlighted the importance of providing appropriate learning opportunities to develop requisite cognitive skills of Working Memory (WM) and Metacognition (MC). Support has been found for these models through correlational research demonstrating the link between WM, MC and maths. Studies have also shown that anxiety, both general and maths specific can interact with WM and MC to affect performance in maths. A review of research indicated that WM training and MC interventions delivered in a maths context can contribute to improved maths achievement, however none of these studies considered the effect of anxiety. The current empirical study examined the effect on 13 and 14 year olds' maths achievement of WM training (CogMed) and a MC intervention (in the form of one-to-one tutoring). Maths achievement, WM, MC and anxiety (maths and general) were measured at pre, post and follow-up (8 weeks) to explore their contribution to maths achievement. Significant improvements for both intervention groups from pre to post and from pre to follow-up were found for WM and maths achievement. A significant reduction in maths anxiety was found in both groups over the same timescale, but there were no significant changes in MC or general anxiety. Evidence was found for a link between changes in WM and general anxiety and a link between changes in MC and maths anxiety. Results are discussed in relation to understanding what factors are important in the observed changes, and implications for educational interventions are considered.

ROLE OF METACOGNITION AND WORKING MEMORY IN MATHS

List of Contents

List of tables

List of figures

Declaration of Authorship

I, EMMA JAYNE WALKER, declare that the thesis entitled '**Understanding the role of metacognition and working memory in maths achievement'** and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- 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;
- where I have consulted the published work of others, this is always clearly attributed;
- 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;
- I have acknowledged all main sources of help;
- 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;
- none of this work has been published before submission.

Signed: ………………………………………………………………………..

Date:……………………………………………………………………………….

Acknowledgements

I would like to thank my supervisors Dr. Julie Hadwin and Dr. Helen Richards for their support with this thesis from the initial development stages to the completed research study. I am indebted to them for their patience, dedication and hard work throughout this process and I am eternally grateful for the time and support they have given me.

I would also like to thank the school staff who made this research possible, and to the pupils who showed such dedication to the interventions they undertook. Their commitment and time is greatly appreciated.

My thanks go to my family, friends and fellow trainees who have helped me to stay smiling. I would like to say a special thank you to my friend John Jakeman for his helpful advice. To my dear friend Katie Smith I wish to say that I am forever grateful for her invaluable support with this research. I would like to thank my wonderful parents for always being there for me and for their unshakeable confidence that I would succeed. Last, but not least I would like to thank my fiancé Andy Burt for living with me through all the stress and putting up with me working all hours, but most of all I want to thank him for his unquestionable love which has always got me through anything.

Abbreviations

AMAS = Abbreviated Maths Anxiety Scale

ANOVA = Analysis of Variance

ASD = Autistic Spectrum Disorder

AWMA = Automated Working Memory Assessment

CIT-34 = Collective Verbal Intelligence Test for grades 3-4

DCSF = Department for Children, Schools and Families

DFE = Department For Education

ECAC = Every Child A Chance

ES = Effect Size

fMRI = Functional Magnetic Resonance Imaging

GCSE = General Certificate of Secondary Education

JMAI = Junior Metacognitive Awareness Inventory

KRT = Kortrijk Arithmetic Test

KS = Key Stage

LTM = Long Term Memory

MC = Metacognition

MD = Maths Difficulties

NC = National Curriculum

NFER = National Foundation for Educational Research

OFSTED = Office for Standards in Education, Children's Services and Skills

RCT = Randomly Controlled Trial

RT = Reaction Time

SCAS = Spence Children's Anxiety Scale

SE = Standard Error

SEN = Special Educational Needs

STM = Short Term Memory

T1 = Pre-test

T2 = Post-test

T3 = Follow-up

VWM = Verbal Working Memory

VSWM = Visuospatial Working Memory

WASI = Wechsler Abbreviated Scales of Intelligence

WM = Working Memory

WRAT = Wide Ranging Achievement Test

ROLE OF METACOGNITION AND WORKING MEMORY IN MATHS

What cognitive factors are associated with achievement in mathematics? A systematic review of interventions.

The global financial crisis has contributed to a rise in unemployment and in many countries the youth labour markets have been particularly affected resulting in high unemployment figures for 19 – 24 year olds (Wolf, 2011). A maths qualification is an important factor in gaining employment or a place in further/higher education because employers and educational establishments typically use a basic level of mathematics as a 'sifting criteria' to award jobs or course places (Wolf, 2011). A qualification in maths affects earning potential, for example adults with a level 1 qualification in maths (equivalent to General Certificate of Education, GCSE, grades D-G) earned on average 26% more than adults with lower level qualifications (Every Child a Chance, 2009). The lifetime cost of unemployment due to numeracy difficulties per annual cohort (based on a yearly school intake of 35, 843 pupils) was estimated by the Every Child A Chance (2009) report to be approximately £1872.7million through lost national insurance, income tax and payment of benefits.

Studies have shown that poor achievement in maths at the start of secondary school affects future performance. For example at 11 years of age the gap between maths attainment of children with Special Educational Needs (SEN), where SEN is defined as difficulty in a subject area that necessitates provision additional to classroom teaching, and typically developing children widens at a faster rate than in any other year (Department For Education, 2011a). Longitudinal research with adolescents suggests that pupils with SEN who experience Maths Difficulties (MD) (where maths difficulties are defined as performance on numeracy based tasks that is substantially below age related expectations) typically make below average progress in maths from primary to secondary school (Shalev, Auerbach, Manor & Gross-Tsur, 2000; Shalev, Manor & Gross-Tsur, 2005). Poor levels of progress in this population continue until the end of seconda ry school because approximately 30% of children with SEN make nationally expected progress in maths from age 11 to 16 years compared to approximately 67% of children without SEN (Department For Education, 2010).

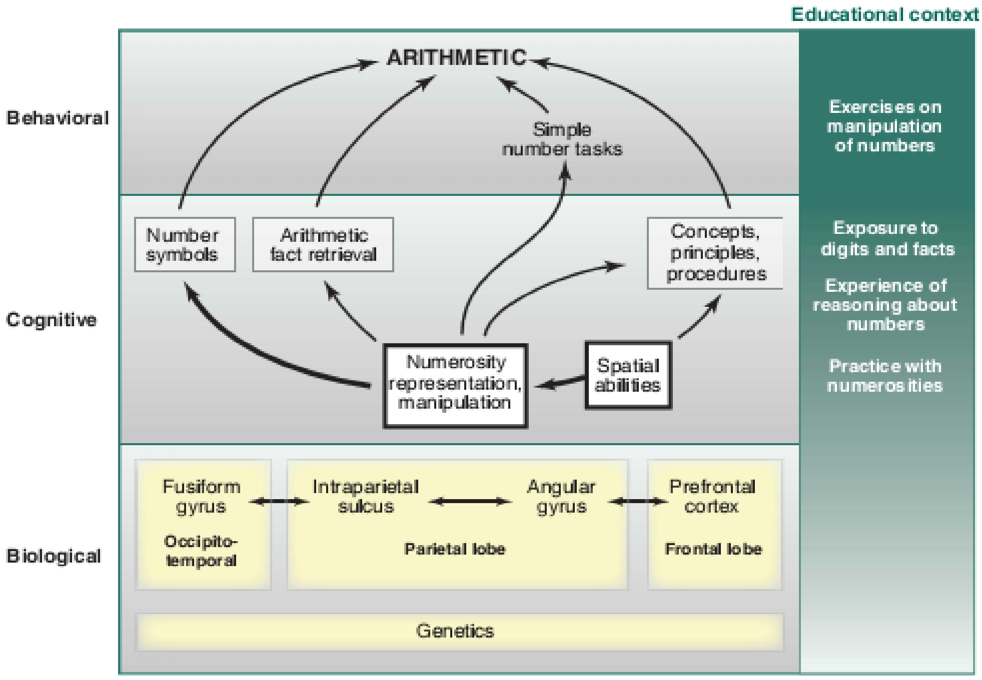
Maths achievement has clear implications both for the economy and young peoples' futures. Therefore several theoretical frameworks have aimed to develop a better understanding of what affects progress in numeracy. This review will draw on these theoretical frameworks and related empirical research to highlight the cognitive factors that affect maths achievement and consider how these factors have been targeted in intervention studies aimed at improving maths attainment. Analysis of contemporary intervention studies and their limitations provides a basis to develop a conceptual understanding of what factors are critical to the development of maths skills in children and young people.

Factors affecting maths achievement

Theoretical frameworks related to maths achievement have highlighted the role of knowledge and skill development (Kaufman & Neurk, 2005). These frameworks have typically suggested that knowledge and skills gained from prior learning are important for future development of maths skills. For example, the accumulator model (Gallistel & Gelman, 2000) focuses on the development of counting skills. It proposed that numbers are represented in the brain via a 'non-verbal counting process' where the brain tracks the overall magnitude of each new item or amount added to a group. It is suggested that the brain learns to link magnitudes to number names and symbols which subsequently develops counting skills. Cippolotti and Butterworth (1995) further proposed that 'abstract internal representations' (similar to knowledge of magnitude) can be used in a variety of ways to complete maths tasks. It is suggested that these abstract internal representations facilitate understanding and production of a verbal or written number through providing a context that makes the numbers meaningful (e.g. magnitude), but in an abstract form so they can be linked to a number of symbolic forms (e.g. numerals).

Researchers have also explored maths achievement and cognitive factors. For example, Butterworth, Varma and Laurillard (2011) proposed a model (see Figure 1) which highlighted the role of cognitive abilities including working memory (WM) and metacognition (MC). This model suggests that if correct educational experiences are not provided there will be insufficient opportunities to develop brain areas linked to maths, which affects the growth of maths knowledge and cognitive functions associated with its development (WM & MC). This model was based on neuropsychological research which has shown that specific brain areas are related to different maths skills. For example, Ischebeck, Zamarian, Schocke & Delazer (2009) through fMRI scanning observed angular gyrus activation in a study requiring retrieval of information relating to previously learned maths strategies. Butterworth et al. (2011) state that the development of brain areas changes routine numerical activity by "shifting from frontal areas (which are associated with executive function and working memory) and medial temporal areas (which are associated with declarative memory) to parietal areas (which are associated with magnitude processing and arithmetic fact retrieval) (p.1050). These changes occur as mathematical information becomes 'well learned' which allows the brain to process mathematical information more efficiently and automatically and WM capacity is 'freed-up' because information can be automatically retrieved and no longer requires conscious processing. Although MC is not specifically mentioned Butterworth et al. (2011) highlighted the importance of being able to organise information as they state that "dyscalculic learners have not sufficiently developed the structures needed to coordinate the components needed for calculation" (p.1051). .

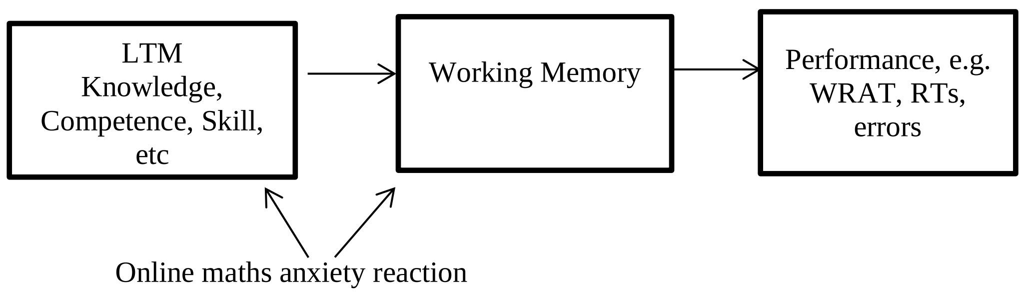
Figure 1: Causal Model of Possible Inter-Relations Between Biological, Cognitive, and Simple Behavioural Levels, Butterworth, Varma and Laurillard (2011). Reprinted with permission from AAAS.



A problem solving framework developed by Demetriou, Christou, Spandoudis and Platsidou, (2002) more explicitly highlights the role of MC and complements that of Butterworth et al. (2011) because it also proposes that maths knowledge develops through appropriate learning opportunities. This framework posits that MC skills organise, monitor and evaluate strategy use, but WM capacity provides the 'working space' for their execution and determines how many MC tasks can be performed at once (Demetriou et al., 2002). Panaoura (2007) empirically applied this model to maths by testing at three time points over three/four months the MC, WM and maths achievement of participants aged 8 – 11 years. Metacognition was measured using a five point Likert scale, WM was assessed using word recall and a matching task and maths achievement was measured using numerical, analogical and spatial tasks from the Standard Progressive Matrices. Results indicated a significant link between WM and MC, where improvements in WM were associated with improvements in MC. At follow-up it was also found that an increase in WM was linked to a significant increase in maths achievement. No direct link between MC and maths achievement was found, which highlights the importance of providing explicit opportunities such as interventions in order for MC to be developed to a point where it may influence maths skills (Desoete, Roeyers & De-Clerq, 2003).

Additional research has investigated further factors associated with achievement in school. Factors such as motivation (Luo, Paris, Hogan & Luo, 2011) and self-efficacy (Ayotola & Adedeji, 2009) have been found to impact on levels of maths attainment. Theoretical frameworks linked to maths achievement have highlighted the negative impact of heightened anxiety (Ashcraft & Kirk, 2001). The proposed mechanism through which this occurs is the presence of anxious thoughts (worries) which can occupy cognitive ‘space’, thus lowering WM capacity (Eysenck & Calvo, 1992) leaving less ‘space’ to process maths information which can impair effectiveness of performance. In the model below the effect of anxiety on MC occurs in the link between LTM and WM because MC, specifically procedural knowledge, allows an individual to be aware of relevant strategies in LTM that may be acted on using WM. This link is supported by research by Veenman, Kerseboom & Imthorn (2000) because they showed that if an individual has a high level of maths anxiety they are unable to use their MC knowledge when completing a maths task and thus perform poorly.

Figure 2: A Representation of the Online Math Anxiety Proposal. WRAT= Wide Range Achievement Test; LTM= Long-Term Memory. RT= Reaction Time (Ashcraft & Kirk, 2001)



Working Memory

Working Memory (WM) involves both the storage and processing of information (LeFevre, DeStefano, Coleman & Shanahan, 2005), and is proposed to consist of four components. The central executive is suggested to co-ordinate four processes: updating, inhibition, shifting and organisation of information from the phonological loop and visuospatial sketchpad. The phonological loop is described as a phonological/acoustic store and the visuospatial sketchpad is suggested to process visual and spatial information (Baddeley & Hitch, 1994). The episodic buffer is proposed to feed information in and out of Long Term

Memory (LTM) (Baddeley, 2000). The role of the episodic buffer is less well understood than other WM components and its links to maths achievement are unclear.

The link of maths achievement to WM has been demonstrated in theoretical frameworks (Butterworth et al., 2011; Demetriou et al., 2002) and empirical research (Alloway, Gathercole, Kirkwood & Elliott, 2009; Bull & Scerif, 2001; Gathercole & Alloway, 2008; Gathercole, Pickering, Knight & Stegman, 2004; Henry & MacClean, 2003). When assessed via auditory and visuospatial tasks Geary et al. (2007) observed that WM scores were positively associated with performance on number-set tasks and with fewer errors on number-line tasks and addition problems. Children with MD were found to have WM scores on average one standard deviation below a control group, make more errors on number-line tasks/addition problems than typically achieving peers and used more simplistic strategies such as finger counting. These findings suggest that children with MD lack mental representations of numerical concepts (similar to younger children), and are more likely to rely on basic strategies such as counting. It is therefore possible that when individuals with MD encounter a new maths concept they utilise the visuospatial sketchpad to create mental representations (Bull, Andrews-Espy & Wiebe, 2008). As age increases children are more frequently required to use spoken cues (task instructions), however this may be difficult for children with poor WM as they struggle to process verbal information (Gathercole, Durling, Evans, Jeffcock & Stone, 2008). Individuals with MD have also been found to have a slower speed of processing than peers (Geary et al., 2007), and more difficulties inhibiting irrelevant information (Bull et al., 2008).

Recent reviews designed to explore methodology and near/far transfer effects of WM training have raised concerns about the effect of WM training on non-trained skills such as arithmetic, citing small Cohen’s d Effect Sizes (ES) and methodological issues (Melby- Lervåg & Hulme, 2013; Shipstead, Redick & Engle, 2010). However the ES quoted in Melby- Lervåg & Hulme (2013) for arithmetic (0.18) was the largest of all studied areas and close to Cohen’s classification of a small ES (0.2) suggesting that maths is perhaps the most promising area of research for WM training. This suggests that WM training may be a useful starting point to develop maths skills but to have a greater effect it may usefully be considered in conjunction with variables such as MC (Autin & Croizet, 2012) and anxiety (Ashcraft & Kirk, 2001). Methodological concerns of WM training also included use of a passive/no control group, a lack of randomisation and failure to use a pre-test-post-test design. Gathercole, Dunning and Holmes (2012) argue that even with these methodological issues the fact that gains are reported following training suggests a greater possibility that the training was effective than if no change had been found. In a new area of research such as WM training where there is currently a limited evidence base it is helpful to adopt the position taken in a review by Morrison and Chein

(2011), that methodological inconsistencies do not invalidate results and instead caution is needed when interpreting findings.

Metacognition

Metacognition (MC) is defined as the ability to appraise, monitor and control thoughts (Flavell, 1979), and is broadly thought to contain two components of metacognitive knowledge and regulation of cognition. Metacognitive knowledge is itself split into three components: procedural knowledge about use of relevant strategies, conditional knowledge about when and why to use a particular strategy and declarative knowledge about what affects performance (Schraw & Moshman, 1995). Regulation of cognition is linked to planning, monitoring and evaluation of performance (Legg & Locker, 2009; Schraw, Crippen & Hartley, 2006).

The link of maths achievement to MC is demonstrated through associations between metacognitive skills and performance on maths tasks (Desoete, Roeyers, & Buysse, 2001; Desoete, Roeyers & Huylebrock, 2006; Özsoy, 2011; Schoenfeld, 1992; Van der Walt, Maree & Ellis, 2008). Associations have been found between maths and both components of MC skills – metacognitive knowledge and regulation of cognition. For example, Özsoy (2011) measured 11 year olds' metacognitive knowledge (procedural, conditional and declarative) and found that it positively correlated with achievement on maths tests of number, geometry, measurement and statistics. Regression analysis of these results found that 42% of the total variance in maths scores was explained by MC. Van der Walt et al. (2008) found that the MC regulation skills of participants aged 12 – 17 years (monitoring and evaluation) were positively correlated with achievement on geometry word problems.

Maths anxiety as a moderator of performance on maths tasks

Maths anxiety refers to "feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of maths problems" (Richardson & Suinn, 1972, p.551). Theoretical models (Eysenck & Calvo, 1992) and related research suggest that individuals high in maths anxiety perform poorly on maths tasks due to reduced WM resources (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007; Kellog, Hopko & Ashcraft, 1999; Maloney, Ansari & Fugelsang, 2010; Prevatt, Wellies, Li & Proctor, 2010). For example, Ashcraft and Kirk (2001) presented undergraduate students with a dual task maths problem involving the presentation of a set of letters to be remembered followed by an addition problem that required the carrying operation (thus taxing WM). The addition problem had to be solved before the letters were recalled therefore the letter recall task also required WM due to the delayed recall. Participants with higher levels of maths anxiety made more errors on the addition problems compared with non-maths-anxious individuals. There was no between group difference on the letter recall task suggesting WM was not generally poor, thus indicating a maths specific difficulty.

Research evidence indicates that maths anxiety also interacts with MC to negatively affect performance in maths tasks, but this is an under-researched area (Legg & Locker, 2009). Veenman et al., (2000) measured metacognitive skilfulness through 'think aloud protocols' where participants had to verbalise their thinking whilst solving a maths problem during which researchers recorded the presence of MC skills, such as prediction and evaluation. This study found that highly anxious participants aged 12 -13 years were rated as lower in metacognitive skilfulness on a maths task compared with less anxious peers. However when metacognitive cues were given during the maths task, such as prompts to review strategy use, the effects of anxiety were ameliorated so that everyone made significant improvements. Similarly, Legg and Locker (2009) found that lower levels of self-rated MC were associated with poorer performance on a timed maths test in participants who rated themselves as highly anxious on a maths anxiety questionnaire. This study therefore showed that if individuals were anxious they were more likely to have poor MC and perform poorly on the maths test, however if they were low in anxiety then MC scores were not found to affect maths achievement.

Aims and objectives of the current review

This review aims to understand how cognitive factors identified as important for maths achievement have been targeted in intervention studies for children and young people at risk of underachieving in maths. Researchers have recognised the links between WM/MC and maths performance and have therefore considered how interventions that focus on these factors can help us to comprehend their role in maths achievement. Therefore this review will consider how effective WM and MC intervention studies have been in increasing maths achievement in children and young people. A critical assessment of the reviewed studies' methodological quality will also be conducted.

Method

Data Sources

Searches were conducted in four electronic databases: PsycInfo via EBSCO (1887 - 2013) (n=15); Web of Science via Web of Knowledge (1950 - 2013) (n=7); EMBASE via OVID (1980 – 2013) (n=0); Medline via EBSCO (1973 – 2013) (n=1). The search terms used were: metacognition, mathematics, mathematics achievement, word problem solving, intervention, training, working memory training, executive function (see Appendix A for further details). The search terms included a list of keywords generated by the author and from those identified in key papers found during the literature search. Additional articles were obtained from a search of WM training websites, a manual search of meta-analyses found as key papers during the literature search and scanning of reference lists of key papers eligible for inclusion.

Participants

Studies were included if participants were 18 years or younger and attended a mainstream or special school. Recruitment to this intervention could be either universal, such as being chosen due to membership of a class group, or selected, such as for having low maths scores. To be eligible for inclusion, participants were not required to have a pre-existing difficulty in maths, and were not excluded based on diagnoses of other difficulties, for example, Autistic Spectrum Disorder (ASD).

Study Design

Studies were eligible for inclusion if they used a quantitative methodology. This included: Randomised Controlled Trials (RCT), controlled trials and quasi-experimental design. Studies were also included whether they had an active, passive or no control group. Case studies were excluded.

Type of intervention

The intervention was eligible for inclusion if it targeted WM and/or MC and measured the impact of training on maths achievement. Interventions were excluded if they did not target WM or MC. Several WM intervention protocols were included in this review: CogMed (Klingberg, Fernell, Oleson & Johnson, 2005), Memory Booster (Leedale, Singleton & Thomas, 2004), Jungle Memory (Alloway & Alloway, 2008), 'Old Yellow'(Van der Molen, Van Luit, Van der Molen, Klugkist & Jongemans, 2010) and a memory strategy training programme devised by Witt (2011). MC interventions were typically delivered during maths problem solving exercises via a computer or in a classroom, with MC support given by teachers. Studies were excluded if interventions were not delivered with a consistent frequency and content for all participants. For example, studies of 'ad-hoc' one-to-one tutoring that covered different content with pupils dependent on need were excluded.

Outcome variables and analysis.

Only studies that included maths achievement as an outcome measure were eligible for inclusion in the review. Studies were excluded if there was no evidence of quantitative analysis. Quantitative analysis could be group based (experimental vs. control) or within group analysis (pre vs. post).

Publication requirements.

Papers were included if they were published in peer reviewed journals. Unpublished work such as dissertations, presentations at conferences and review articles were not included.

Data Extraction and Synthesis.

Data extracted from the eligible papers included: 1) Descriptive information about participants (age, gender, characteristics such as maths or WM difficulty), 2) Study design, 3) Descriptive information about the intervention (content, duration, frequency, method of delivery) and the control groups (active, passive or no-control), 4) Outcome measures (validity and reliability) and 5) Treatment effects (for measured variables). Treatment effects included differences between and within experimental and control groups over time.

Quality Assessment

The quality of eligible studies was assessed using a checklist devised by Downs and Black (1998). This checklist provides a framework through which to consider methodological quality of interventions. It consists of 27 items split into five subscales:

1) Reporting (10 items): to assess clarity of reporting: hypotheses, outcome measures, sample, interventions and results.

2) External validity (3 items): to assess the extent to which findings may be considered representative of and generalizable to the target population

3) Internal validity (7 items): To assess the presence/control of factors in research methodology that may have biased results, such as intervention non-compliance or the absence of single/double blind procedures.

4) Internal validity (6 items): To assess the presence/control of factors within participant groups that may have biased results (e.g. the absence of statistical adjustment for confounding variables or the absence of a randomised design).

5) Power (1 item): To assess the ability of statistical tests to detect an experimental effect of which it is 95% certain that the effect is not due to chance, and is evidenced by a power calculation.

Results

The database search included a total of 29 records for all types of intervention (WM: 7 & MC: 22).

Working memory

Sample characteristics. The WM studies in this review (see Table 1) included young people who had: no identified difficulties, SEN, low WM and Attention Deficit Hyperactivity Disorder (ADHD). The gender split was generally equal, except in two studies where respectively 86% and 87% of the sample were male (Alloway, 2012; Gray et al., 2012). The age range of participants varied from 5–17 years. Studies were conducted in the United Kingdom, the Netherlands, and Canada.

Study Design. Studies included RCTs, quasi-experimental and non-randomised designs (where allocation was typically based on class membership or maths/WM scores) with active, passive or both types of control group (see Table 1 for details). There were five single blind studies (where either the school, research assistants or child was blind to group allocation). Four studies (Alloway, 2012; Gray et al., 2012; Van der Molen, 2010; Witt, 2011) used two time points (pre and post testing) to assess intervention effects, whereas two studies assessed follow-up: after six months for the intervention group only (Holmes, Gathercole & Dunning, 2009) and after eight months (Packiam-Alloway et al., 2013). One study (St-Clair-Thompson, Stevens, Hunt & Bolder, 2010) conducted pre and post testing on sub-groups of participants.

Intervention. Four types of adaptive WM training were used in the reviewed studies. Adaptive WM training involves the trial-by-trial adjustment of difficulty levels based on the participant's performance and included CogMed (Klingberg et al., 2005), Jungle Memory (Alloway & Alloway, 2008) and 'Old Yellow' (Van der Molen et al., 2010). Two studies used CogMed which trains both visuospatial and verbal Short Term Memory (STM) and WM. It employs 12 tasks, with eight tasks presented in each training session. Tasks involved temporary storage and/or manipulation of sequential verbal or visuospatial information, or a combination of both (Klingberg et al., 2005). Two studies used 'Jungle Memory' (Alloway & Alloway, 2008) which used three tasks each session to train visuospatial STM & WM and verbal WM. The first game involved scanning a four by four square grid to locate letters or word endings and then subsequently recalling their location. The second game involved processing letter rotations and remembering the location of a red dot and the final game required participants to solve maths problems then remember their solutions in the same order. The final intervention based on the plasticity principle only trained visuospatial WM. In this study the authors (Van der Molen et al., 2010), designed the computerised 'Old Yellow' WM programme based on the 'Odd One Out' test by Henry (2001). On every trial, participants were presented three shapes, and were asked to identify the 'odd one out', following which the shapes disappeared and the learner indicated its previous location.

A novel variation of WM training was 'Memory Booster' (Leedale et al., 2004), used by St-Clair-Thompson et al., (2010). Retrieval of items to enable more efficient processing in WM was trained through a computer programme that taught three memory strategies to remember information presented: rehearsal, visual imagery i.e. linking displayed objects to a mental image to increase memorability and visual chaining where displayed objects were linked in a narrative. After learning each strategy, participants were presented with items and encouraged to remember them using the taught strategy. Participants recalled items after a short delay and correct answers earned positive verbal feedback and an increase in the number of presented items. The final intervention was designed and delivered by Witt (2011) with the aim of improving retrieval of items from LTM. The training consisted of practical activities to learn the same strategies as taught during ‘Memory Booster’.

All but one intervention (Witt, 2011) was computerised and all were delivered over 5 to 8 weeks, with an average of 18 sessions and a range from 6 – 32 sessions. Session frequency ranged from daily to once a week, for approximately 35 – 45 minutes. However, sessions in ‘Old Yellow’ (Van der Molen et al., 2010) were six minutes and in Witt (2011) they were 15 minutes.

All of the studies that used interventions to increase WM capacity used an active control group, although control group activities were varied. For example, two studies used a non-adaptive version of the training programme (Holmes et al, 2009; Van der Molen, et al., 2010), one offered a lower frequency of the WM training programme (once per week) (Packiam-Alloway et al., 2013) and one offered a programme that trained only STM (Van der Molen et al., 2010). Two studies offered targeted educational support, one of which taught curriculum related content (Alloway, 2012), and the other provided a maths training programme (Gray et al., 2012). St-Clair-Thompson et al. (2010) and Witt (2011) used a passive control group.

Measures. Tests of WM typically assessed components associated with the Baddeley and Hitch (1974) model: visuospatial and verbal STM and WM. Measurements commonly included: block recall, listening recall, visual patterns test and digit span (forwards and backwards). In the block recall task participants were asked to recall, in the same order, sequences tapped on blocks. In the listening recall task participants heard sentences and classified them as true or false. When all sentences had been heard participants recalled in order the last words of the sentences. In the visual patterns test participants were shown a matrix of squares, some of which were coloured. After a short delay a blank matrix was presented and participants had to recall which squares had been coloured. A variation on this was the shape recall test which displayed a coloured shape on a four by four grid which then disappeared. Participants had to verify if two shapes shown on screen were the same colour and shape as the previous one, and then recalled the original shape’s location on the four by four grid. The digit span tasks involved remembering a series of numbers and recalling them forwards or backwards. One study (Alloway, 2012) assessed a single rather than multi-component WM model using a letter recall task, where participants were shown a letter on a computer screen immediately followed by another letter. Participants had to verify if letters were the same and then remember them in the correct order.

Two studies included a practical assessment of WM (Holmes et al., 2009; St-Clair-Thompson et al., 2010) using protocols (based on work by Engle, Carullo & Collins, 1991;

Gathercole et al., 2008), where participants followed an increasing number of instructions until they could no longer perform them correctly. This involved sitting in front of an array of props (e.g. rulers, pencils, folders) in a range of colours and being given an increasing number of verbal instructions, e.g. "Touch the yellow pencil and then put the blue ruler in the red folder" (Holmes et al., 2009).

Measures of achievement typically included a standardised test of reading and arithmetic. Arithmetic tests included the National Foundation for Educational Research (1999a,b) test which posed questions on number, algebra, measures, shape/space, and the arithmetic subtest from the Wechsler Abbreviated Scales of Intelligence (WASI) (1999). Also used were the numerical operations subtest from the Wechsler Objective Numerical Dimensions (1996) and the Wide Ranging Achievement Test (Wilkinson & Robertson, 2006) which both involved solving math problems including addition, subtraction, multiplication, division and algebra. Witt (2011) assessed maths via a series of researcher generated addition problems that required the carrying operation. The reading test used was the Group Reading Test II (National Foundation for Educational Research, 1992) which presented a number of pictures about which participants had to complete sentences using an option from a multiple choice selection. Spelling was assessed using the Wechsler Objective Reading Dimensions (1993).

The ability to apply knowledge and skills was assessed using the WASI (1999) vocabulary subtest where participants named a picture shown to them and the matrix test which assessed non-verbal ability by showing incomplete patterns from which participants had to select a shape to complete. Additional outcome measures included response inhibition (using a Stroop test) and motivation (Likert scale rating statements about interest in maths) (Van der Molen et al., 2010). Attention and hyperactivity were measured using questionnaires, including the Strengths and Weaknesses of ADHD symptoms and Normal behaviour (Swanson et al., 2001) and the IOWA Conners rating scale (Pelham, Milich, Murphy & Murphy, 1989).

Outcomes. All WM training studies showed improved scores on tests of WM (N=7). This finding also extended to more practical, instruction based WM assessments (Holmes et al., 2009; St-Clair-Thompson, 2010). Van der Molen et al., (2010) found that only STM scores showed significant improvement, whereas Gray et al., (2012) found that participants significantly improved on WM and not STM measures. There is some evidence to suggest that the observed effects were sustained. For example, Packiam-Alloway et al., (2013), Holmes et al. (2009) and Van der Molen et al., (2010) completed follow-up assessments respectively at 10 months, 6 months and 10 weeks after training completion and found WM improvements were still present. Studies that utilised an active control group found significant effects of WM training on WM measures (Alloway, 2012; Holmes et al., 2009; Van der Molen et al., 2010), with the exception of Gray et al. (2012).

Further findings suggested that WM training was associated with improvements in untrained tasks, i.e. maths (Alloway, 2012; Witt, 2011) and spelling (Packiam-Alloway et al., 2013), however most studies reported this effect at follow-up and not post-test (Holmes et al., 2009; Van der Molen et al., 2010). Four (/7) studies found that the WM training group showed significantly more improvement in maths scores following the intervention than a control, and one study (/7) found that the WM training group showed significantly more improvement in maths achievement over time compared to a control. Two studies found no pre-post significant difference in maths achievement between the experimental and control group (Gray et al., 2012; St-Clair-Thompson et al., 2010). The impact of WM training on untrained areas such as response inhibition, motivation, attention and hyperactivity was found to be non-significant (Gray et al., 2012; Van der Molen et al., 2010).

Quality assessment

Reporting subscale. All of the studies partially or fully reported: study hypotheses/aims; outcomes; participant characteristics; intervention details, principal findings, variability and research specific confounds (e.g. fatigue and motivation levels whilst completing assessments). The main limitations in reporting were the lack of information provided about attrition and participants lost to follow-up (with the exception of Van der Molen, 2010 and Gray et al., 2012). A consistent problem was the lack of reporting on potential adverse events (e.g. if participants had to miss lessons to take part in an intervention).

External validity. In all but one study (Alloway, 2012), insufficient information was provided about whether participants who agreed to take part in the study were representative of the population from which they were recruited. No studies gave information as to whether staff, places, and facilities used in the study were representative of that which would be available in schools. This was an area of weakness for all reviewed studies.

Internal validity (bias). Some studies used single blind procedures (5/7 studies) and most studies used appropriate statistical tests and reliable measures (6/7 studies). However, no studies used a double blind methodology which is a limitation in management of bias. The reviewed studies typically monitored intervention compliance and maintained a consistent time period between post-testing and follow up for all participants.

Internal validity (confounding/selection bias). The risk of confounding or selection biases varied considerably between studies. Three studies were RCTs but it was not clear if condition allocation was concealed from participants and staff until it was ‘irrevocable’, therefore randomisation may have been compromised. Authors typically reported adjustments made for confounds present in participant data (e.g. differences between the groups in scores on the pre-test measures or for loss of data at follow-up). Confounding and selection bias were potentially present in a number of the reviewed studies.

Power. None of the studies reported a power calculation and therefore it was unclear if they had sufficient power to detect an effect if one was present.

Summary

Working memory and maths achievement were measured across WM training studies. In general results reported improvements in these areas. However, it is unclear if the improvements in WM persisted beyond the training intervention because only three studies conducted follow-up measures. The reviewed studies only found evidence of a broader effect of WM training on maths achievement but not for cognition (response inhibition), motivation or symptoms of attention and hyperactivity, which is contrary to findings from other WM studies (e.g. Klingberg et al., 2005).

The quality of studies was generally acceptable although there were methodological limitations such as studies without randomised samples and an absence of double blind measurements. Reporting was typically good and bias was reduced in some studies through high internal validity. Limitations included a lack of information about external validity, for example it was unclear how representative the participants were of the wider population from which they were drawn. There was also an absence of power calculations thus rendering it unclear as to whether studies had sufficient statistical power.

Table 1. Review of Working Memory Study Characteristics and Findings

Key for abbreviated test names:
Working Memory: AWMA = Automated Working Memory Assessment WMIT = Working Memory Instructions Task LRT = Letter Recall Test WMTB = Working Memory Test Battery DR = Digit Recall NWR = Non-Word Recall VPT = Visual Patterns Test BDR = Backwards Digit Recall FDR = Forwards Digit Recall LR= Listening Recall SS= Spatial Span BR = Block Recall CANTAB = Cambridge Neuropsychological Testing Automated Battery WMRS = Working Memory Rating Scale SR = Shape Recall
Maths: WOND = Wechsler Objective Number Dimensions MA = Mental Arithmetic AP = Addition Problems
Achievement: GRTII = Group Reading Test II WISC IV = Wechsler Intelligence Scale for Children IV WASI = Wechsler Abbreviated Scales of Intelligence WORD = Wechsler Objective Reading Dimensions

Abbreviations:
EG = Experimental Group MC = Metacognition CT = Controlled Trial WM = Working Memory CG = Control Group RCT = Randomised Controlled Trial *(NB: Difference in favour of experimental group unless stated otherwise)* 1 = Significant difference between groups at post –test2 = Significant difference between groups at follow-up 3 = Significant difference between pre and post in experimental group 4 = Significant difference between pre and follow-up in experimental group 5 = EG improved significantly more from pre to post than CG 6 = EG improved significantly more from pre to follow-up than CG

Study (Quality assessment score)	Design	Target Sample	Intervention	Outcome Measures	Key Results
Packiam-Alloway, Bibile & Lau (2013)	CT Active and passive control group Pre, post and follow-up testing (8 months)	**Characteristics:** Learning difficulties and receiving special educational provision **N (% male):** 94 (43%) **Age range:** Not given **Mean age:** EG: 11.2years (no s.d) Passive CG:10.11 years (no	**EG (n=23)** Jungle Memory 4 times a week **Passive CG (n=39)** Regular classroom activities **Active CG (n=32)** Jungle Memory once a week	**Working Memory:** BDR, LRT, SR **Maths** Numerical Operations test from WOND **Achievement:** Spelling test from WORD Vocabulary and Matrix subtests from WASI	1.BDR, LRT, WORD 2.LRT, BDR, WORD 3. LRT, BDR

	Single blind (researcher)	s.d.) Active CG: 10.06 years (no s.d.)			
Alloway (2012)	RCT Active control group Pre & post testing	**Characteristics:** 'Learning disabilities' **N (% male):** 13 (87%) **Age range:** Not given **Mean age:** EG: 12.9 years (s.d. 1 year) CG: 13yrs (s.d. 0.4)	**EG (n = 8):** Jungle Memory 3 times a week, for 30 minutes a time, over 8 weeks. **CG(n = 7):** Educational support on 'skills for attainment' as outlined on their Individual Education Plan (IEP): Three times a week, for 30 minutes a time, over an 8 week period.	**Working Memory:** LRT **Maths** Numerical Operations test from WOND **Achievement:** Crystallized ability: vocabulary subtest from WASI	5. All measures

Gray et al. (2012)	RCT Single blind (researchers) Active control group Pre & post measures	**Characteristics:** Full time at residential school IQ > 80 Coexisting Learning Difficulties(LD)/ADHD/behaviour difficulties WM < 25^{th} percentile WRAT 4 < 25^{th} percentile Excluded: conduct disorder, aggression, depression, anxiety **N (% male):** 60 (87% male) **Age range:** 12-17 years **Mean age:** Experimental group: 14yrs 4mths (s.d. 1.3yrs) Control group: 14yrs 2mths (s.d. 1.1yrs)	**EG(n = 32):** CogMed for 45 minutes a day, for 4/5 days a week for 5 weeks. **CG(n = 20):** Academy of Math for 45 minutes a day, for 4/5 days a week for 5 weeks.	**Working Memory:** BDR & FDR (WISC-IV) SS (CANTAB, Fray, Robbins, & Sahakian, 1996) WMRS (Alloway, Gathercole, Kirkwood, & Elliott, 2009) **Maths:** WRAT (reading, spelling, maths & comprehension) **Attention & Hyperactivity** SWofADHD&NB Scale (SWAN, Swanson et al., 2001) IOWA Conners scale (Pelham, Milich, Murphy, & Murphy, 1989)	No stats: 36% greater improvement at post-test compared to CG (BDR) & 28% for SS

Witt (2011)	Quasi-experimental (matched pairs) Pre & post testing Passive control	**Characteristics:** Attended 4 primary schools in the same region in England. **N (% male):** 38 (39%) **Age range:** 9 – 10 years **Mean age:** Overall: 9y 7 m (SD 3.43 m)	**EG(n =19):** 6 weeks –once a week for 15 minutes given 1:1 support on researcher designed WM tasks. **CG(n =19):** Normal classroom teaching	**Working Memory** DR VPT **Achievement:** AP	1. Maths, Visuospatial WM and Verbal WM 5. Maths, Visuospatial WM and Verbal WM

St-Clair-Thompson et al. (2010)	Quasi-experimental Passive control Single blind (researchers)	**Characteristics:** Mixed SES, SEN **N (% male):** 118 (46%) **Age range:** 5 – 8 years **Mean age:** EG: 6y 10m (no s.d. given) CG: 6y 11m, no s.d. given)	**EG (n = 117):** Memory Booster over 6 – 8 weeks, with 2 sessions of 30 minutes. **Control group (n = 137):** Passive control	**Working Memory:** WMTB: Pre and post WM digit recall; block recall and listening recall (n=254) listening recall (n=141) WMIT (n=77); MA (n=34). **Maths** NfER mental maths (pre& post n=81; follow-up n=69) WISC-IV arithmetic (pre& post n=77; follow-up 70); **Achievement:** GRTII (pre & post n=81; follow-up n= 72)	1.Verbal STM/WM, WMIT 3. Visuospatial STM (block recall), Verbal STM/WM, WMIT, Maths & Reading
Van der Molen et al. (2010)	RCT Single blind (researchers) Active control Pre & post: twice in a four week period (every 2 weeks) to find an average. Follow-up (10 weeks)	**Characteristics:** IQ score 55 -85 **Excluded:** ADHD, ASD or previous serious head injuries **N (% male)** 54 (57%) **Age range:** 13 –16 years **Mean age:** EG: 15.32 years (SD = 0.68) CG: (B) 15 years (SD = 0.70) CG: (C): 15.43 years (SD = 0.66)	**EG (n = 41):** 'Old Yellow' WM programme, 3 times a week, for 6 minutes over a 5 week period. **CG (B) (n = 27):** Non-adaptive version of EG programme. **CG (C) (n = 27):** Identify the odd one out on a grid of shapes.	**Working Memory:** DR; NWR; BR; VPT; BDR LR; SS **Maths** Arithmetic test (De Vos, 1992) **Achievement:** Reading test (Brus & Voeten, 1973) Story recall (Van der Molen, 2007) **Response Inhibition** The Stroop (Hammes 1978) **Motivation** Likert scales 1-10	5. DR & NWR (verbal STM). 6. BR & VPT (visual STM); Maths; Story recall

Holmes, Gathercole & Dunning (2009)	CT Active control group Single blind (participants) Pre to post testing 6 month follow-up (intervention only)	**Characteristics:** WM scores at or below the 15th percentile**.** **N (% male):** 27 (64%) **Age range:** 8 **-** 11 years **Mean age:** EG: 10y 1m (no s.d.given) CG: 9y 9m (no s.d.given)	**EG(n = 22):** CogMed: 35 minutes a day, for at least 20 days in a period of 5 – 7 weeks. **CG(n = 20):** Non-adaptive version of CogMed**.**	**Working Memory** AWMA: WMIT **Maths** WOND **Achievement:** WASI: verbal and performance IQ. WORD	1. Verbal & visuospatial WM & WMIT 3.All aspects of WM & in WMIT 4. All aspects of WM (incl. WMIT) & maths scores (WORD)

Metacognition

Sample characteristics. The MC studies in this review (see Table 2) were all conducted on young people who attended mainstream schools (n=22). Some studies also specified that participants had been identified as having MD (n=6), or general learning difficulties (n=1) and the remainder of the studies used populations with a full range of maths ability. A number of studies did not report the number of male and female participants but those that did recorded a roughly equal split (see Table 2). Participants were from: Israel, Belgium, France, South Africa, USA, Singapore, the Netherlands and Finland and were all aged between 8-17 years (see Table 2 for further details).

Study Design. There were 11 out of 22 studies that were RCTs using a passive or active (i.e. alternative training such as study skills) control group. Ten studies were quasi-experimental where participants were assigned to experimental or active/passive control groups in a non-random manner, and one study had no control group. Only one study used double blind procedures (Desoete et al., 2003) and three used single blind procedures (Fuchs et al., 2003; Kajamies, Vauras & Kinnunen, 2010; Kapa, 2007), therefore the majority of studies (n=18) made no use of blinding. Most studies (n=15) used two times points (pre and post measures), although some (n=7) used three (pre, post and follow-up). Different follow-up periods were used in these studies and these ranged from 2 weeks to 1 year (see Table 2 for further details).

Intervention. All training programmes taught maths content using MC supportive teaching methods to enable participants to solve maths problems. Some interventions presented maths problems via a computer programme (n=6) or through classroom teaching (n=16). Six different computer based MC training programmes were used in the reviewed studies: ‘CRIME’ (Teong, 2003); ‘Maths Explorer’ (Seo & Bryant, 2012) and computer programmes designed and/or used by, the respective authors (Jacobse & Harskamp, 2009; Kajamies et al., 2010; Kapa, 2007; Kramarski & Mizrachi, 2006). These programmes utilised broadly the same format and MC components. This included the presentation of maths word problems in a computer game format or via an online forum (e.g. Kramarski & Mizrachi, 2006). Whilst participants were solving these problems teachers asked questions and gave prompts to encourage the use of the following MC skills: task understanding, linking the current problem to previous knowledge, generation and evaluation of possible strategies. Kramarski and Mizrachi (2006) included a second experimental group who received the same MC strategies but used them during whole class problem solving. Two studies also gave feedback on negative performance to inform future actions (Jacobse & Harskamp, 2009; Kajamies et al., 2010). Most computer based interventions (n=5) used a passive control group where pupils received normal classroom maths instruction, with the exception of Seo and Bryant (2012) who had no control group.

The time period over which computer-based interventions were administered ranged between 2 weeks to 8 weeks (see Table 2 for further details). Seo and Bryant (2012) was the only study that did not deliver the intervention for a consistent period of time for all participants as training was stopped when a minimum level of maths skills were reached (13/18 correctly answered maths questions on four consecutive occasions). Duration of sessions in the Seo and Bryant (2012) study were approximately 20-30 sessions over 4 to 7 weeks. Generally the frequency and duration of intervention sessions in reviewed studies were relatively consistent and were carried out once or twice a week for 30–60 minutes. The only exception was Seo and Bryant (2012) who reported attendance of up to five sessions a week for 20–30 minutes.

Classroom based interventions targeted similar MC skills to computer-based training. MC skills were taught by a teacher during maths lessons on problem solving tasks or in mixed ability peer groups of five participants (n=7). Some studies taught additional skills to aid maths problem solving, such as text comprehension (Seo & Bryant, 2012) and problem solving rules (Vauras, Kinnumen & Rauhanummi, 1999). Vauras et al. (1999) taught MC skills and text comprehension strategies outside of class time and then applied these as part of the intervention to a maths problem solving game called the 'Quest for the Golden Chalice' (designed by the researchers). Masqud (1998) and Verschaffel et al. (2009) also delivered the intervention outside of class time and in addition to normal maths lessons. Most studies used a passive control group where participants received instruction in class as normal. Three studies also used an active control which taught strategies such as: calculation (Desoete et al., 2003), co-operative problem solving in small groups (Kramarski, Mevarech & Amrani, 2002), and step by step instructions to solve maths questions (Fuchs et al., 2003; Mevarech, 1999).

The time period over which classroom-based interventions were administered typically ranged from 2 weeks to 5 months, but in one case was a year. The frequency and duration of the intervention sessions was more consistent and was carried out five times a week for between 40-60 minutes. The only exceptions to this were sessions run twice a week (Vauras et al., 1999), six or seven times a week (Fuchs et al., 2003), 28 sessions evenly distributed over the course of an academic year (Hoek, van den Eeden & Terwel, 1999), and 20 sessions over 4 months (Verschaffel et al., 2009).

The only study that did not train MC strategies in the context of teaching maths problem solving skills was Pennequin, Sorel, Nanty and Fontaine (2010). They devised a programme based on the Strategy Evaluation Matrix and Regulatory Checklist. Five 1 hour sessions were delivered over 7 weeks through direct instruction about and modelling of when to use a strategy and why, such as skim reading to identify salient points, planning, checking understanding and evaluating performance. Maths skills were not part of this training. This study used an active control group and they were taught generic study skills.

Measures. All studies measured maths achievement which was typically assessed using non-standardised tests devised by the researchers. Some researchers used a standardised test in addition to the non-standardised measure. For example, Kajamies et al. (2010) used the RMAT (Räsänen, 2004), which is a test of arithmetic skills (e.g. simple computations and algebra questions). Hoek et al. (1999) used the Mathematical Reasoning Ability Test (Horn, 1969), which measures mathematical and inductive reasoning. Desoete et al. (2003) was the only study to solely use standardised measures which were the KRT2 and 3 (translated as: Kortrijk Arithmetic Test, Cracco et al., 1995). These were pencil and paper tests containing a series of questions about number system knowledge and mental computation.

The non-standardised maths tests can be categorised into three types: structured tests containing questions assessing specific maths topics (e.g. algebra), word problems and 'open-ended' questions with more than one possible solution. Most studies (n=9) used more than one of these non-standardised measures. The first category (structured tests) often used multiple choice or single answer questions to assess the maths topic taught during the intervention. For example tests assessed: factorisation, algebra, linear equations, interest and using one or two step calculations (Kramarski & Mizrachi, 2006; Masqud, 1998; Mevarech, 1999; Mevarech & Kramarski, 1997), geometry, fractions and the metric system (Cardelle-Elawar, 1992, 1995) or graph interpretation, answering questions using information from a graph (Kramarski & Mevarech, 2003). Tests in the second category (word problems) included maths problems that were based on, or similar to, those trained in the intervention. These assessments differed from the first category because instead of phrasing questions as a calculation they were embedded in a narrative from which participants had to extract information to solve the problem. Tests in the final category (open-ended) can be further split into two types: firstly, those assessing the way in which the problem was solved and awarding points for complexity of the approach used. An example of this may be whether information was organised meaningfully to reach a conclusion, evidenced by diagrams and algebraic equations (Hoek et al., 1999; Kapa, 2007; Mevarech & Kramarski, 1997; Vauras et al., 1999). The second type assessed skills that were not explicitly taught during the intervention. For example, Kramarski and Mevarech (2003) used a graph construction test in which participants created graphs from written information, rather than interpreting pre-prepared graphs as was taught during the intervention.

Approximately half of the studies did not assess metacognition but those that did used a range of quantitative and qualitative measures. Four studies used questionnaires which were based on: Schraw and Dennison's (1994) Metacognitive Awareness Inventory; Swanson's

(1990) Metacognitive Questionnaire; and a questionnaire adapted from Montague and Bos (1990). The remaining five studies assessed metacognition whilst participants were solving maths problems. These measures included: think aloud protocols, where participants verbalised thought processes at each problem solving stage (Jacobse & Harskamp, 2009; Teong, 2003); scoring written or videotaped maths problem solving for evidence of metacognitive strategies (Kapa, 2007; Mevarech & Kramarski, 2003); asking participants to predict/evaluate performance on maths questions, then comparing these predictions/evaluations with actual outcomes (Desoete et al., 2003; Pennequin et al., 2010).

Additional outcome measures included: verbal and non-verbal intelligence, reading comprehension and attitude to maths (see Table 2 for full details). Verbal intelligence was assessed by the Collective Verbal Intelligence Test for grades 3-4 (CIT-34, Stinissen, Smolders, & Coppens-Declerck, 1974) and non-verbal intelligence was assessed by Ravens Progressive Matrices (1958, 2000). Reading comprehension was assessed using a standardised reading test (Lindeman, 1998) which required participants to answer questions about texts they had just read; or it was measured by filling in gaps in sentences taken from a text that had just been read (Vauras et al., 1999). Attitude towards maths was assessed by the ‘Scale of Attitude Towards Maths (Aiken, 1979), or the Beliefs and Attitudes questionnaire (Verschaffel et al., 2009). This contained two subscales: seven items about pleasure and persistence in solving word problems and 14 items about level of solution or process oriented problem solving.

Outcomes.

Studies which measured MC (n=9) found that the experimental group showed higher levels of MC at post-test than controls. The only study which measured MC and did not replicate this finding was Kapa (2007), who found that the experimental group made significantly more progress from pre to post measurement but that there was no between group difference at post-test. These changes in MC were measured using standardised measures (Kramarski & Mevarech, 2003; Kramarski & Mizrachi, 2006; Masqud, 1998; Mevarech & Amrany, 2008), a non-standardised rating scale measuring improvement in rated MC behaviours (Pennequin et al., 2010), analysis of ‘think aloud protocols’ (Jacobse & Harskamp, 2009; Teong, 2003), or the recorded process of problem solving (Desoete et al., 2003; Kapa, 2007; Mevarech & Kramarski, 2003). No studies reported follow-up measurements.

For measures of maths achievement nearly all studies (n=20) found at post-test that the MC training group had significantly higher scores than a control group. This finding was consistent across different maths assessments such as standardised tests, word problems, topic specific tasks and open ended questions. The exceptions to this were Kajamies et al. (2010) and Montague, Applegate and Marquard (1993). Montague et al. (1993), found no significant difference between groups at post-test on either standardised tests or word problems, however the experimental group was found to have made significant progress from pre to post.

It was unclear whether improvement in maths could generalise to untrained maths topics. For example, Desoete et al., (2003) found at post-test that the experimental group significantly outperformed a control group on trained maths skills such as procedural calculation, but there was no significant between group difference on untrained skills such as matching numbers. Four studies used open ended maths tasks to assess untrained skills and found that the experimental group either improved significantly from pre to post measurement or performed significantly better than a control group at post-test (Hoek et al., 1999; Kapa, 2007; Mevarech & Kramarski, 1997; Vauras et al., 1999).

Some studies separated participants into subgroups of low and high maths achievers. Differences in maths progress were found between these subgroups although these were inconsistent across studies. For example, Mevarech and Kramarski (1997) had two control groups, one with low achievers and one with high achievers, and two MC training groups, again one with low achievers and one with high achievers. They found that high achievers in the MC training group made significantly more progress on open ended maths tasks than high achievers in a control group. However, there was no significant difference in performance on these problems between control and MC training groups containing low achievers. Conversely, Hoek et al. (1999) found that low ability participants in the experimental group made more progress on an open-ended test compared with the higher ability group.

Studies with more than one experimental group provided information about what made MC training most effective. For example, Kramarski and Mevarech (2003) trained MC strategies in a group and also with individuals and found that group based problem solving led to significantly higher gains. A tangential development of this finding was a study by Kramarski and Mizrachi (2006) where two groups received the same MC instruction but the group solving maths problems via an online computer programme had significantly higher scores at post-test than the group completing this through whole class discussion. Results from Kapa (2007) indicated that MC support should be given at multiple stages of the problem solving process. For example, group 1 were given MC prompts when identifying the problem, planning, evaluating and gaining feedback, group 2 had MC prompts at all stages except feedback, group 3 only had MC prompts at feedback stage and controls had no MC support. At post-test it was shown that the more MC prompts given, the more maths performance improved. These studies show ways in which MC and maths learning may be most successfully combined, (Kramarski & Mevarech, 2003; Kramarski & Mizrachi, 2006) and at which points in a lesson MC support has the most impact (Kapa, 2007).

Follow-up measurements of maths attainment were rare (n=5), and did not provide much evidence of the sustainability of training effects. For example, Seo and Bryant (2012) found that 50% of participants (n=2) had maintained gains from post-test, but scores of the other 50% had decreased by 11%. Kajamies et al. (2010) found that at follow-up the experimental group was no longer significantly different to the control group, however Desoete et al. (2003) found that at follow-up the experimental group scored significantly higher than controls. Finally, Montague et al. (1993) found that maths scores improved from post to 3 week follow-up, but that this improvement was only sustainable on subsequent measurements if a 'booster' session was given prior to testing.

There did not appear to be a difference in results for studies using selected samples and those who had randomised participants. For example, eight RCTs and 11 selected samples found significant differences between groups at post-test, three RCTs and two selected samples showed significant differences in the experimental groups pre to post scores, one RCT and four selected samples showed significant differences between groups at follow-up, and only one selected sample showed significant differences between pre and follow-up scores for the experimental group.

In addition to the broader effects of MC training on maths achievement, improvements were also noted in a number of untrained areas. Intervention groups were found to have made significant improvements at post-test in comparison to controls in the following areas: non-verbal intelligence (Cardelle-Elawar, 1992; Masqud, 1998); attitude to maths (Cardelle-Elawar, 1992, 1995; Masqud, 1998) and reading comprehension (Vauras et al, 1999). No significant differences were found at post test for verbal intelligence (Desoete et al., 2003).

Quality assessment

Reporting subscale. All studies partially or fully reported on: hypotheses/aims, measures and outcomes, participant characteristics, type of intervention, main findings and estimates of random variability in the data. Exceptions to this were: Vauras et al., (1999) who failed to adequately report on outcomes, participants' characteristics and random variability; Kajamies et al. (2010), Mevarech and Kramarski (1997) and Seo and Bryant (2012) who all failed to adequately report their main findings. Distributions of confounders within participant groups were not adequately described in a number of studies (n= 5). All but one of the studies (Hoek et al., 1999) failed to report on potential adverse consequences of the study intervention, and all but one study (Seo & Bryant, 2012) did not report on characteristics of participants lost to follow-up. Finally, only three studies fully or partially reported exact p-values for significant findings (Hoek et al., 1999; Kapa, 2007; Mevarech, 1999).

External validity. With the exception of three studies (Fuchs et al., 2003; Kapa, 2007; Kramarski & Mevarech, 2003) there was insufficient information to report on external validity. In the majority of papers it was not possible to determine: whether participants who were invited and who assented to take part were representative of the population from which they were recruited. It was also not possible to determine whether the staff, setting, and facilities where the intervention was carried out were representative of the experience of the majority of pupils.

Internal validity (bias). The majority of studies were potentially biased in outcome measurement because all but one failed to use double blind procedures (Desoete et al., 2003), and only four studies were single blind (Fuchs et al., 2003; Kajamies et al., 2010; Kapa, 2007; Mevarech & Kramarski, 2003). It is therefore unclear as to whether detection and performance bias may have affected results as those participating and those measuring outcomes may have been influenced by their knowledge of group allocation. Statistical analyses were conducted appropriately in all studies however it was not possible to determine whether any analyses were based on data dredging. In all but one study (Kapa, 2007) there was no clear information given as to adjustments made to analyses to account for differences in follow-up periods. One of the main limitations in the reviewed studies was the widespread use of non-standardised measures with no reported reliability and validity data, therefore making it unclear if instruments measured what they purported to. Only two studies (Desoete et al., 2003; Kajamies et al., 2010) used measures which could be considered to be reliable and valid (i.e. KRT2 & 3 & RMAT), with a further seven using measures that partially met this criteria. Only seven studies either fully or partially reported intervention compliance rates.

Internal validity (confounding/selection bias). The level of confounding and selection biases varied across studies. A relative strength was that 14/22 studies recruited participants from the same population therefore minimising confounding variables. However, findings were mixed with regards to the use of measures to minimise selection bias. For example, randomisation was employed by approximately half of the papers (n=10), however in these studies it was not clear if this randomisation had been concealed from participants and those implementing the intervention until it was irrevocable. The main limitations of the reviewed studies were a failure to report on: the time period over which participants were recruited; characteristics of participants lost to follow-up, and whether adjustments were made to analyses to account for confounding variables.

Power. The majority of studies did not report conducting a power calculation (n=19), however Desoete et al. (2003); Kajamies et al. (2010); and Pennequin et al. (2010) all reported calculations indicating they had sufficient power to detect an effect if one were present (over 0.8).

Summary

Results indicated that MC interventions contributed to the primary outcome of improved maths achievement and secondary outcomes of: MC, attitudes to maths, reading comprehension and non-verbal (but not verbal) intelligence. Some evidence suggested that effects from the training programme generalised to untrained tasks such as matching numbers, but because only half of the studies provided a measure of untrained skills it is difficult to draw firm conclusions. There was limited use of follow-up measures across studies, and therefore there is insufficient evidence to conclude that improvements in maths achievement, reading comprehension and non-verbal intelligence were maintained over time. The studies that used follow-up measures reported mixed findings where some treatment effects persisted but others did not. It is unclear whether low or high maths achievers benefitted differently from MC interventions.

Results of some of the reviewed studies should be treated with caution due to irregularities in outcome measurement. For example, a number of studies (Kramarski, Mevarech & Arami, 2002; Kramarski & Mizrachi, 2006; Mevarech & Amrany, 2008; Mevarech & Kramarski, 2003) did not use the same measures for pre and post-test thus it was unclear if groups improved over time and only three studies used a standardised measure of maths achievement.

Table 2. Review of Metacognition Study Characteristics and Findings

Key for abbreviated test names
Metacognition: EPA2000 = Evaluation Prediction & Assessment 2000 MAI = Metacognitive Awareness Inventory SMQ = Swanson Metacognitive Questionnaire TAP = Think Aloud Protocol (whilst solving maths problem)
Maths: WPT = Word Problem Task (devised by authors of study) MTDR = Maths Test Devised by Researchers RLPS = Real Life Problem Solving SMT = Standardised Maths Test MT = Maths test (devised by authors of study) KRT = Kortrijk Arithmetic Test OEP = Open Ended maths Problem MRAT = Mathematical Reasoning Ability Test
Attitudes: BAQ = Beliefs and Attitudes Questionnaire SATM = Scale of Attitude Towards Maths
Achievement: RPM = Ravens Progressive Matrices

Abbreviations:
EG = Experimental Group MC = Metacognition CT = Controlled Trial WM = Working Memory CG = Control Group RCT = Randomised Controlled Trial
MCML = Metacognitive guidance given in the context of a Maths Lesson MCCP = Metacognitive guidance given whilst solving maths problems on a Computer Programme *(NB: Difference in favour of experimental group unless stated otherwise)* 1 = Significant difference between groups at post –test2 = Significant difference between groups at follow-up
3 = Significant difference between pre and post in experimental group 4 = Significant difference between pre and follow-up in experimental group
5 = EG improved significantly more from pre to post than CG 6 = EG improved significantly more from pre to follow-up than CG

Metacognition (MC) Training study	Design	Target Sample	Intervention	Outcome Measures	Key Results
Seo & Bryant (2012)	Quasi-experimental No control Pre-test (tested once a week for 3 weeks) post-test & follow-up (2 weeks)	***Characteristics:*** Low SES, grades 2-3 in Texas, USA; identified difficulties in maths, ethnicity= White, Hispanic and African American; ***N (% male):*** 4 (75%) ***Age range:*** 7-8yrs ***Mean age:*** 7yrs 9mth	***EG (n=4)*** MCCP (Maths Explorer)	***Maths*** MTDR (18 items ‘Maths Explorer’ word problems). ***No MC measure***	Scores on MTDR improved by: 75%; 79%; 85% & 86%. At follow-up, 2 students’ scores on MTDR dropped by 11%, but 2 remained the same

Kajamies, Vauras & Kinnunen (2010)	CT Single blind (researcher) Pre, post & follow-up (6 months later) Passive and active control	***Characteristics:*** Attended 12 schools in 21 classes in schools in Finland. ***N (% male):*** 429 (no gender split given) ***Age range:*** Not given ***Mean age:*** 10yrs 4mths (s.d. 4 months) (no info given on EG or CG)	***EG (n=8)*** MCCP (Quest of the Silver Owl, Vauras & Kinnunen, 2003). ***CG 1 (n=8)*** Normal lessons ***CG 2 (n=8)*** Reading comprehension intervention (Quest for Meaning) for 20 hours. ***Other (n=405)*** 'Typical' achievers -	***Maths*** RMAT – (no full name given) WPT (15 one step and multi-step problems). ***No MC measure*** ***Reading Comprehension*** Finnish standardised reading test (Lindeman, 1998). ***Non-verbal intelligence*** RPM (Raven et al., 2000).	3. Maths (WPT) 5. Maths (WPT) 1. Maths (RMAT), but difference only significant compared to 'other' group, not CG.

Pennequin, Sorel, Nanty & Fontaine (2010)	RCT Pre & post test Active control	***Characteristics*:** School in Touraine region of France, low achievement in maths ***N (% male)*:** 48 (48%) ***Age range:*** 8.4 years – 10.1 years ***Mean age*:** EG: 8.11yrs (s.d. 0.4yrs) CG: 8.9yrs (s.d. 0.3yrs)	***EG (n=12)*** Over 7 weeks, 5 x 1 hour training sessions in groups of 6. N=12 (low maths, MC) N=12 (average maths, MC) Based on Schraw (1998) Strategy Evaluation Matrix & Regulatory Checklist. ***CG (n=12)*** N=12 (low maths) N=12 (average maths) Same number of teaching hours, but taught study skills.	***Maths*** Pre: 12 WPT (based on textbook about MC in problem solving, Côtè 2000) Post: Similar word probs. ***MC*** Rate 9 MC actions & predict how many problems you would solve.	3.MC (low achievers only); prediction of performance; maths (WPT), but lower achievers improved more than normal achievers.
Jacobse & Harskamp (2009)	CT Passive control Pre & post-test	***Characteristics:*** Two classes (in one Netherlands school). Comparable SES, maths and word problem solving. ***N (% male):*** 49 (45%) ***Age range:*** Not given ***Mean age:*** 11 years (no s.d. and no EG/CG information)	***EG (n= 23)*** MCCP ***CG (n= 26)*** Normal teaching in class.	***Maths*** WPT (15 items) ***MC (only used on EG)*** TAP	1.Maths 3.MC (TAP - increase in use of planning and orientation activities)

Verschaffel et al. (2009)	CT Passive control Pre, post and follow-up (no time given)	***Characteristics:*** Belgian, spoke Dutch. ***N (% male):*** Not given ***Age range:*** 6-12 years ***Mean age:*** Not given	***EG (n=86)*** 20 lessons lasting 50-60 minutes over 4 months. Taught MC strategies to approach Worked in mixed ability groups of 3-4 pupils. ***CG (n=146)*** Followed regular maths curriculum in normal lessons.	***Maths*** 10 item WPT SMT (Stinissen, Mermans, Tistaert & Vander Steene, 1985). Videoed maths problem solving and rated on scale devised by researchers. ***Attitudes to maths*** BAQ Two subscales: 7 items (pleasure and persistence in solving word problems); 14 items (level of solution or process oriented problem solving). ***No MC measure***	1.WPT, BAQ, SMT 2.WPT 3.WPT, BAQ, SMT No difference in achievement for low, medium and high ability.
Mevarech & Amrany (2008)	RCT Pre & post test Interviews (2mths after post-test, but	***Characteristics*:** High school students; Israeli; ***N (% male)*:** 61 (0%) ***Age range:*** Not given ***Mean age*:** Overall: 16yrs 8mths	***EG (n not given)*** Taught maths module (over a month) using MC instruction method called IMPROVE. ***CG (n not given)*** Studied same content	***Mathematics*** MTDR (different versions at pre & post) ***MC*** 24 item questionnaire adapted from MAI (Schraw &	1.Maths & MC

	no pre-test interviews)	(s.d. not given) (No info for each group)	with no explicit MC instruction	Dennison, 1994)	
Kramarski & Mizrachi (2006)	RCT 2 active controls Pre & Post (but different measures)	***Characteristics:*** Seventh grade Israeli students. ***N (% male):*** 86 (no gender split given) ***Age range:*** None given ***Mean age:*** 13.4yrs (no s.d. given & no information given about mean age in EG and CG)	***Online:*** MC (IMPROVE) ***Face to Face discussion(FtF)*** Maths prob. solving discussion in a whole class group. ***EG 1 (n=20)*** (Online & MC) ***EG 2 (n=22)*** (FtF & MC) ***CG 1 (23)*** (Online) ***CG 2 (21)*** ***(***FtF)	***Maths*** *Pre-test:* MTDR (38 item multiple choice - factual knowledge and open ended computations). *Post-test:* MTDR (56 item multiple choice algebra questions. RLPS ***MC*** Adapted questionnaire from Montague & Bos (1990) and Kramarski & Mevarech (2003).	**EG1:** 1.Maths (MTDR & RLPS) & MC (sig. better than EG2 & CG) **EG2:** 1.Maths (MTDR & RLPS) & MC (sig. better than CG)
Kapa (2007)	RCT Single blind (participant) Active control Pre & post-	***Characteristics:*** Studied at 4 public junior high schools in Tel-Aviv (Israel). ***N (% male):*** 231 (49.4%) ***Age range:*** 13-14 years ***Mean age:*** Not given	***EG 1 (n=62)*** MC support given at each problem solving phase and at conclusion of task. ***EG 2 (n=61)*** MC support given at each problem solving phase but no	***Maths & MC*** MTDR OEP Product score = how many correct Process score: 0 -5, with a point awarded for each MC stage completed.	1. MC (process score on MTDR), NB: EG1 significantly better than all groups, EG2 significantly better than all but EG1. Maths (MTDR) 3.Maths (OEP

	test (2 months after intervention)		correcting feedback at end of task. ***EG 3 (n=46)*** MC support given at task conclusion. ***CG (n=62)*** No MC support.		product score)
Desoete, Roeyers & DeClerq (2003)	Double blind RCT Pre, post & follow-up testing (But: Follow-up testing only on intelligence and maths)	***Characteristics*:** Third grade; White; Dutch speaking; Attended 7 Belgian mainstream elementary schools ***N (% male)*:** 237 (52%) ***Age range:*** Not given ***Mean age*:** Overall: 8yrs 3mths (s.d. 3.27mths) (No info for each group)	**All groups except control:** Groups of 10, 5 times a week for 2 weeks, for 50 minutes. **MC group (n=49):** 'Number Town': Maths problems with cognitive & MC instruction. **Algorithmic group (n=50):** 'Count City' no MC instruction. **Motivation group (n=38)** Maths problems on computer – no instruction. **Math group (n=42)** Five paper and pencil exercises of common	***Metacognition*** EPA2000 (trained and untrained cog. & MC maths problem solving) ***Maths achievement:*** KRT2 & KRT3 (due to different norm periods) ***Verbal Intelligence*** CIT-34 ***Motivation*** 5 point rating scale for lessons	1.Maths 2. Maths 3. MC 4. Maths

			maths problems. **Control (n=58)** Same instruction time as other groups, but over 5 sessions:		
Fuchs et al. (2003)	CT Pre & post test Active control group	***Characteristics:*** Attended 6 schools in USA. Pupils chosen as taught by teachers selected to receive training ***N (% male):*** 375 (49%) ***Age range:*** 8-9 years ***Mean age:*** Not given	***EG 1 (n=89)*** 20 sessions solution training ***EG 2 (n=94)*** 10 solution sessions and 10 MC sessions ***EG 3 (n=97)*** 20 solution sessions + 10 MC sessions ***CG (n=95)*** Taught students using 'Math advantage' programme, as a normal class.	***Maths*** WPT (10 items identical to the intervention content) WPT (7 items similar to the intervention content, but one presented in a novel format) WPT (4 items - different types of word problem). ***No MC measure***	1.Maths (WPT, 10 items) (All EGs) 1. Maths (WPT 7 items) EG3 significantly better than EG1 1. Maths (WPT 10 items) (EGs 2 & 3)
Mevarech & Kramarski (2003)	Quasi-experimental Pre & post-test. Follow-up after 1 yr. Active control	***Characteristics*:** Eighth grade Israeli students from five classes ***N (% male)*:** 122 (gender split not given) ***Age range:*** Not given ***Mean age*:** 14.12yrs (s.d. 0.62yrs)	***EG (n=70)*** Worked in a group of 5 (mixed ability based on pre-test scores) to solve maths problems using the IMPROVE method. ***CG (n=52)*** Worked in a group of	***Maths*** *Pre-test*: MTDR (algebra) *Post-test*: MTDR (time-distance-speed unit). ***MC*** *Videotape of problem solving:* Coded for	1.Maths 2. Maths ***Videotape coding:*** EG had a higher number of MC statements than CG.

		(no info given on EG and CG	5 using worked out examples of maths problems.	different MC behaviours, e.g. monitoring, reflection.	
Kramarski & Mevarech (2003)	CT Passive (control 2) & Active (control 1) Pre & post-test	***Characteristics:*** Attended 12 classes in 4 junior high schools in Israel ***N (% male):*** 384 (47%) ***Age range:*** Not given ***Mean age:*** 13.3yrs	***EG 1 (n=105)*** MCML (worked in groups - IMPROVE method). ***EG 2 (n=95)*** MCML (worked individually – IMPROVE method). ***CG 1 (n=91)*** Co-operative discussions in class ***CG 2 (n=93)*** Worked individually to solve problems.	***Maths*** Graph interpretation test (based on Mevarech & Kramarski, 1992/3). Graph construction test (transfer task – non taught skill). ***MC*** General & domain specific MC questions: questionnaire adapted from study by Montague & Bos (1990).	1.Graph interpretation test (NB: EG1 performed significantly better than all groups, but EG 2 performed significantly better on graph interpretation test than CG). 1.Graph construction test & domain specific MC
Teong (2003)	Quasi-experimental Active control Pre, post & follow-up (6	***Characteristics:*** Attended a Singapore school. Low achievers in maths based on end of year exam results. ***N (% male):*** 40 (no gender split given) ***Age range:*** 11-12 years	***EG (n=20)*** CRIME & word problems on WordMath computer software. ***CG (n=20)*** Solved word problems on WordMath	***Maths*** WPT ***MC*** TAP (didn't compare EG & CG – case study approach)	NB: EG & CG had diff. pre-test scores so means were adjusted. 1.Maths 2.Maths MC – Presence of MC behaviours

	weeks later)	***Mean age:*** Not given	computer software.		didn't guarantee success in WPT.
Kramarski, Mevarech & Amrani (2002)	CT Passive control Pre & post testing	***Characteristics:*** Israeli. Attended junior high schools. ***N (% male):*** 91 (48%) ***Age range:*** 12-13 years ***Mean age:*** 12.3 years	***EG (n=60)*** Five lessons a week for 6 weeks. Index cards and teacher prompts of 4 MC questioning styles, following IMPROVE. Worked in small missed ability groups to discuss solutions. ***CG (n=31)*** Pupils had co-operative discussions to solve maths problems.	***Maths*** WPT – scored 0-3 (pre-test) sored 0-5 (post-test), based on organisation of information and justification for answer. MT - 41 items (pre-test): whole numbers, fractions, decimals and percentages OEP (maths reasoning) MT - 22 items (post-test) rational numbers, positive and negative numbers, algebra. ***No MC Measure***	1.MT (both low and high achievers, but bigger effect size for high ability.
Hoek et al. (1999)	CT Pre & Post-test Active	***Characteristics:*** Attended two Dutch secondary schools. ***N (% male):*** 444 (53%)	***EG (n=222)*** MCML (working in groups) ***CG (n=222)*** Teacher instructed	***Maths*** MRAT (Horn, 1969). OEP (Measures & measurement; Information	1.Maths (MRAT & OEP) 3.MRAT (Higher ability children) 3. OEP (information

	control	***Age range:*** 12-13 years ***Mean age:*** Not given	pupils to work together, but gave no further special assistance.	gathering) ***No MC measure***	gathering) (Lower ability children)
Mevarech (1999)	RCT Active and passive control Pre & post test	***Characteristics:*** Attended one Israeli junior high school. ***N (% male):*** 174 (51%) ***Age range:*** 12 – 14 years ***Mean age:*** 13.4yrs (no info given about EG and CG)	***EG (n= 71)*** MCML (IMPROVE method) Worked in small groups ***CG 1 (n=32)*** Used a diagrammatic strategy to solve maths probs. ***CG 2 (n=71)*** No MC prompts or strategy training.	***Maths*** MTDR (20 item test - calculation problems and 4 transfer problems) ***Reading comprehension*** Exam set by Israeli education board. ***No MC measure***	1.Maths (MTDR - transfer problems and less straightforward algebra task)
Vauras, Kinnunen & Rauhanummi (1999)	CT Passive control Pre & post testing	***Characteristics:*** Involved in a longitudinal study by authors, from pre-school. ***N (% male):*** 58 (gender split not given) ***Age range:*** 9 - 10yrs ***Mean age:*** No info given	***EG (n=22)*** MCCP & text comprehension strategies (Quest for the Golden Chalice). ***CG (n=22)*** Took normal maths classes. ***CG2 (n=16)*** High achievers – took normal classes.	***Maths*** WPT OEP ***No MC measure*** ***Reading comprehension*** Read a text then fill in the gaps in sentences taken from the text.	1. Maths (OEP - more realistic answers) 5. Maths (finding correct solutions and using correct methods of calculation).
Maqsud (1998)	RCT	***Characteristics:*** Attended middle school in South	***EG:(n=20)*** Taught maths in class	***Maths*** MTDR (15 items)	1.All measures

	Passive control Pre & Post	Africa. Lower than average maths compared to others in school. ***N (% male):*** 40 (no gender split given) ***Age range:*** 15 – 17 years ***Mean age:*** 15.4yrs (no s.d or information on EG and CG given)	– experimenter looked at written work and interviewed students to teach MC strategies. ***CG: (n=20)*** Taught maths in class – normal teacher help given.	***MC*** Modified SMQ ***General ability*** RPM (1958). ***Attitude towards maths*** SSTM (Aiken, 1979).	
Mevarech & Kramarski (1997)	CT Passive control Pre & post measures (study 1)	***Characteristics:*** Junior high schools in Israel: EG from different school district to CG. ***N (% male):*** 247 (no gender info given) for study 1 & 265 for study 2 ***Age range:*** 12 years ***Mean age:*** Not given	***EG (n=99, study 1; n=164, study 2)*** Index cards and teacher prompts/use of 4 MC questioning styles, following IMPROVE. Worked in small groups to discuss solutions. ***CG (n=148, study 1; n=101, study 2)*** Traditional maths teaching.	***Maths*** Study 1 (post-test after a semester): MTDR (25 items algebra) OEP (11 items) Study 2 (post-test after an academic year): MTDR (48 items – algebra).	Study 1 & 2: 1.Maths (MTDR & OEP) For analysis students split into low, mid and high maths achievers in maths. 1.All abilities: Maths (OEP); Mid and high achievers (MTDR)
Cardelle-Elawar (1995)	RCT Passive control	***Characteristics:*** Elementary & junior high school in USA. 87% Hispanic. Low SES. Below average maths scores ***N (% male):***	***EG (n=297)*** MCML ***CG (n=172)*** Normal maths teaching	***Maths*** MTDR (20 items - content of maths units taught in study).	1.All measures

	Pre & Post measures	489 (no gender info given) ***Age range:*** 8 – 13 years ***Mean age:*** Not given		***No MC measure*** ***Attitudes towards maths*** Aiken's (1974) scales E (enjoyment) & V (value)	
Montague, Applegate & Marquard (1993)	RCT Pre & post testing (follow-up at 3, 5 & 7 weeks) Active & passive controls	***Characteristics:*** Learning Difficulties; visual, auditory or language difficulty; IQ = or > 85 ***N (% male):*** 96 (68%) ***Age range:*** 12-15yrs ***Mean age:*** EG1 = 14.3yrs EG2 = 14.5yrs EG3 = 13.9yrs CG = 13.7yrs	***EG 1 (n=23)*** MCML, 50 minutes a day for 7 days, tested after 8 days, then 5 days of instruction, tested on 14th day. ***CG 1 (n=25)*** 50 minute scripted lesson in a small group. Memorise problem solving strategies, then teacher modelled them. ***CG 2 (n=24)*** Did what EG 1did. ***CG 3 (n=24)*** Tested at pre-test and then on 14th day.	***Maths*** Six sets of WPT (10 items) ***No MC measure***	3.Maths (WPT) No sig between group diffs (CG = much higher starting point) 4. Maths (3 week follow-up) (slight increase at 7 weeks when given short booster session)
Cardelle-Elawar (1992)	RCT Passive control	***Characteristics:*** Sixth grade, attended elementary school in Arizona. Low SES, mostly Hispanic, below average maths.	***EG (n=60)*** MC approach based on Mayer's (1987) problem solving.	***Maths*** MTDR (20 items - content of maths units taught in study)	1.All measures

	Pre & Post measures	***N (% male):*** 90 (no gender split given) ***Age range:*** Not given ***Mean age:*** 11 years (no s.d)	Feedback on errors based on Elawar & Corono (1985). ***CG (n=30)*** Taught maths by their normal teacher.	***No MC measure*** ***Attitudes towards maths*** Aiken's (1974) scales E (enjoyment) & V (value) ***General ability*** RPM (1958)	

Discussion

The current review explored the efficacy of interventions that targeted WM and MC with the primary aim of improving maths achievement. Results provided some evidence that WM training improved maths attainment, as four (/7) studies showed significant between group differences following training. Two studies found this result at follow-up and two at post-test (but they did not conduct follow-up measures). One study which found no significant between group difference did report significant within group differences for the experimental group but not for a control group, from pre to post test. In all studies, training improved WM significantly more in an experimental than a control group. Two of these studies conducted follow-up measures which found sustained WM improvements. The reviewed studies did not examine the statistical link between improvements in maths achievement and improvements in WM over the course of the training therefore it is not clear whether changes in WM underpinned changes in maths achievement Findings from WM studies are likely to be relatively robust as the quality of the WM studies was generally good, although there were no instances of double blind procedures to ensure minimisation of detection and performance bias and only half of the studies reviewed used randomised allocation. Due to the small number of studies available, their findings would benefit from further replication.

Results for nearly all MC studies (n=20) found that the experimental group had significantly higher maths scores at post-test than a control group. Follow-up measurements were rare and only three of four studies which took follow-up measurements showed sustained improvement in maths. The majority of MC interventions tested maths skills taught during the intervention, and therefore it was unclear whether numeracy was generally improved by the MC intervention and could be generalised to untrained maths topics. Four studies used open ended maths tasks which assessed untrained maths skills and provided some evidence to support generalisation because they all found that the MC group performed significantly better than a control group at post-test. As the MC studies were delivered in the context of a maths lesson this provided more information than the WM studies as to how to combine cognitive (MC) instruction with maths teaching. For example, it was found that MC instruction was more effective when incorporated into group discussion than when given individually (Kramarski & Mevarech, 2003), but lone working on a computer was more effective than whole class MC support (Kramarski & Mizrachi, 2006). Research also showed that MC prompts had the most impact on learning when given at all stages of problem solving as opposed to just the beginning or end. Only half of the MC studies tested MC, but those that did found that it improved significantly in the experimental group as compared to a control group. Those studies which did measure MC did not routinely statistically examine the link between maths achievement and

MC development. This is a significant omission because without evidence of a statistical link between change in MC and improvement in maths scores it is unclear whether MC underpinned the observed changes in maths achievement. Pennequin et al. (2010) were the only study to statistically examine the link of MC to maths problem solving. Using regression analysis they found that MC accounted for 52% of the variance in progress made in maths problem solving. The direct link of MC and maths skill development is therefore an area that requires further examination.

Any effect of MC training on maths needs to be viewed in light of methodological problems in these studies. For example, there were a lack of standardised measures and validity/reliability data was often not reported. Furthermore, maths tests used in the reviewed studies often measured skills taught in the intervention therefore results may have just reflected memorisation of maths content taught during the intervention rather than assessing the transferable effect of MC on maths. In a third of the studies reviewed pre and post-measurements of maths scores were taken using different tests at each time point. Progress from pre to post could therefore not be ascertained because a difference in scores may have reflected an individual's aptitude for the post-test over the pre-test measurement. There was limited use of blinding and many MC studies (n=12) did not use randomisation thus creating the possibility of selection bias, although it should be noted that similar results were found in RCTs and studies using selected samples.

Findings from this review are consistent with the theoretical framework proposed by Butterworth et al. (2011). This model proposed that if the correct educational experiences are not provided this indirectly affects the development of cognitive functions such as WM and MC, which subsequently affects processing of mathematical information and development of maths skills. This is demonstrated in this review because participants given the educational experiences of WM and MC training tended to experience greater growth in maths skills than participants who had not been given these opportunities. The mechanisms through which MC and WM training affected maths achievement are likely to be different. As MC training was delivered in the context of a maths lesson the 'educational experience' facilitated more direct links to maths content (Dignath & Büttner, 2008), therefore the MC prompts may have allowed a pupil to work out 'what to do next' Kirsh (2005). Research suggests that over time support from a teacher allows MC abilities to become more fluent as young people become more able to articulate their thinking processes and the reasoning behind them (Hennessey, 1999). The development of fluency as seen in the MC interventions is consistent with Demetriou et al., (2002), whose model proposes that the more difficult a task the more MC resources it requires thus potentially 'overloading' WM and interfering with the application of MC until the skill becomes ‘automatic’. Therefore, as demonstrated by Kapa (2007) for each new problem pupils may have needed prompts to know when to use MC strategies until this became a consolidated skill that could be used automatically (Schoenfeld, 1992).

Correlational research suggests that WM capacity is lower in individuals with maths difficulties and that specifically central executive functions such as inhibition are strongly linked to maths achievement (Bull et al., 2008; Geary et al., 2007). It is likely therefore that the observed increase in WM capacity contributed to higher maths scores through an improved ability to inhibit irrelevant information (Bull et al., 2008) and an increased capacity to process verbal information (Gathercole et al., 2008). As this link has not previously been explicitly explored this would benefit from further research. Research indicates that the impact of improved WM capacity on maths scores may be mediated by affective variables such as anxiety and, therefore, it is important to consider this relationship when interpreting the results of the reviewed studies. Empirical evidence indicates that WM capacity can be reduced by both general and maths specific anxiety as ‘worry’ consumes cognitive capacity leaving less available to process information (Ashcraft & Kirk, 2001; Eysenck & Calvo, 1992). This can have negative effects on achievement in maths as reflected in higher error rates during problem solving (Ashcraft & Moore, 2009; Kellog et al., 1999) and significantly slower processing of numbers increasing in value as compared to a control group (Maloney et al., 2010). As none of the research studies reviewed measured anxiety it is unclear whether an anxious individual if given WM training would also show improved WM or maths achievement, therefore this is an area for future research.

There was a risk of publication bias in the current review as only studies from peer reviewed journals were eligible for inclusion; therefore, it is possible that studies which found contrary or non-significant results were omitted. Additionally, a broader scope considering affective factors such as self-efficacy would have developed a fuller understanding of what factors affected maths achievement. For example, research suggests that low self-efficacy in maths can produce intrusive thoughts (similar to the effects of anxiety) which then interfere with WM capacity as evidenced by performance on dual maths tasks (Bonnot & Croizet, 2007). This effect was found even when maths ability was controlled for, suggesting that individuals with higher self-efficacy may benefit more from WM training.

Future Research

There are clear theoretical links between maths achievement and WM/MC and an increasing literature base considering the extent to which maths abilities improve following WM and MC interventions. There is some evidence to suggest that WM and MC interventions lead to improvements in maths achievement and improvements in the targeted cognitive factor (e.g.

WM capacity or MC abilities). However, it is currently unclear whether there is a direct association between these intervention effects. Therefore it is not possible to determine whether improvements in WM capacity or MC abilities are mechanisms through which training leads to improvements in maths achievement. Future intervention studies should test statistical models in which cognitive factors act as mediators or moderators for improvement in maths achievement.

Intervention studies that directly compare the effects of MC and WM training on maths achievement would provide a greater understanding of the interrelationships between WM, MC and maths achievement. Research should monitor the change in these variables over time in both WM and MC interventions to ascertain the impact of these variables on observed outcomes. The moderating effect of anxiety should also be considered as this has clear empirical and theoretical links to the effect of WM and MC on maths achievement (Ashcraft & Kirk, 2001; Eysenck & Calvo, 1992; Veenman et al., 2000). Training in WM did not have an effect on untrained areas such as response inhibition, motivation and attention and hyperactivity, but MC intervention groups significantly improved in non-verbal intelligence, attitude to maths and reading comprehension. It is unclear why MC skills appeared more generalisable than WM skills, therefore this may benefit from future research.

Future research should aim to establish the transferability of maths skills by testing untrained maths content and follow-up assessments should be conducted to assess sustainability of measured effects. An understanding of the interaction of cognitive and affective variables and their effects on maths outcomes will help to clarify the type of support that should lead to improved maths attainment. Effective support for pupils that raises maths achievement is an important area of research given the significance of maths skills to life opportunities. Further research in this area therefore has implications for pupil achievement and for the wider picture of youth employment.

Understanding the Role of Metacognition and Working Memory in Maths Achievement: An Empirical Study

Achievement in maths has always been a key item on the political agenda but with the accession of the coalition government focus on this area has sharpened, with calls for greater school accountability in raising attainment of under/low achieving pupils (Department For Education, 2012). These reforms are underpinned by research highlighting the link between maths skills and future earning potential. For example, Crawford and Cribb (2013) found that maths scores at primary school have a greater direct association with earnings at the age of 30 than reading test scores. Moreover, a labour market lacking in maths skills has serious financial implications. The cost to the taxpayer of unemployment due to numeracy difficulties was estimated by the Every Child A Chance (2009) report to be in the region of £1872.7million through lost national insurance, income tax and payment of benefits. Recent increases in youth unemployment have driven reforms to educational provision, resulting in an increased focus on gaining a formal qualification in maths because this is often used as a ‘sifting criteria’ to award jobs or course places (Wolf, 2011). The importance of providing programmes that develop the maths skills of low or under-achieving pupils has been highlighted as crucial to improving employability (Wolf, 2011). To reflect this focus on improving outcomes for low achievers, performance tables that list the attainments of pupils in primary and secondary phases of education have been amended to highlight the progress of low achieving children (Department For Education, 2011b).

Researchers have argued that raising maths achievement may best be addressed through the provision of appropriate, evidence based interventions (Grouws & Cebulla, 2000). The importance of providing appropriate learning opportunities is highlighted in a model of maths development by Butterworth et al. (2011). This framework suggests that educational experiences affect maturation of maths specific brain areas which in turn affects development of cognitive functions such as Working Memory (WM) and Metacognition (MC) that are linked to maths performance. Demetriou et al. (2002) similarly proposed a theoretical framework that considered the role of MC and WM in problem solving. This model suggested that MC skills are used to organise the selection and application of problem solving strategies and then to monitor and evaluate their use. WM capacity is linked to how able a person is to use these MC skills as it provides the space to ‘hold’ these thoughts whilst they are being processed.

Consistent with frameworks that have focused on WM and MC in achievement, empirical research has found positive associations between maths achievement with WM (Alloway, Gathercole, Kirkwood & Elliott, 2009; Bull & Scerif, 2001; Gathercole & Alloway, 2008) and MC (Desoete, 2007; Desoete et al., 2006; Van der Walt et al., 2008). Based on the established link in the literature between these variables, research studies have implemented interventions to train WM and MC and have empirically tested the impact of these on maths achievement (see Dignath & Büttner, 2008; Morrison & Chein, 2010 for reviews). Although both WM and MC interventions aim to improve maths achievement the mechanisms through which they achieve this are theoretically quite different.

Working Memory Interventions

WM represents a limited storage control system that coordinates verbal and visuospatial material and manages the transfer of information in and out of long term memory (LTM) (Baddeley & Hitch, 1974). In this model the phonological loop is the component that stores phonological/acoustic information for up to 2 seconds unless refreshed via sub-vocal rehearsal (Baddeley & Hitch, 1994). The visuospatial sketchpad is theorised to act as a 'workspace' for holding and manipulating visual and spatial information (Baddeley & Hitch, 1994). These stores are thought to be controlled by the central executive which is suggested to co-ordinate functions including dual task performance, attentional focusing/switching and inter-facing with LTM (Baddeley, 1998; Bull & Scerif, 2001). The original conceptualisation of the WM model had no clearly defined mechanism or store by which information from both the visuospatial sketchpad and the phonological loop could be held and integrated with information from LTM, therefore Baddeley (2000) proposed the episodic buffer. Input from the phonological loop and the visuospatial sketchpad were thought to be bound by the central executive and passed to the episodic buffer. The buffer is argued to retrieve relevant information from LTM to give meaning to the stored verbal and visuospatial information - a process that can be observed in phenomenon such as 'chunking' (Baddeley, 2002). The buffer therefore acts as a modelling space for information before it is passed to LTM, but it represents a separate system as it is not capable of long term storage (Baddeley 2000, 2002).

In the research literature WM training is suggested to act on two different mechanisms - plasticity (Klingberg, 2010) and strategy training (Leedale et al., 2004). Memory strategy training is based on the principle of improving retrieval of information from memory (Witt, 2011). Although the links between use of memory strategies and improvement in WM have been documented in research (e.g. McNamara & Scott, 2001) the links between this type of training and improvements in maths achievement are rare, (see St-Clair-Thompson et al., 2010; Witt, 2011). St-Clair-Thompson et al., (2010) used a computer programme called 'Memory Booster' with participants aged 5-8 years to teach them three memory strategies: verbal rehearsal, mental imagery (i.e. linking displayed objects to a mental image) and visual chaining (linking objects in a narrative). Witt (2011) supported pupils of a similar age (9-10 years) with practical activities to teach these strategies. The results in both studies show improvement in maths skills both over time (St-Clair-Thompson et al., 2010) and between the experimental and a control group (Witt, 2011).

WM training based on the principle of plasticity has a more established research base. Plasticity is the notion that brain connections improve as a result of training which alters the structure of the brain thus changing the way in which information is processed (Takeuchi et al., 2010). Research has found a number of brain areas to be linked to WM tasks. For example, participants aged 9-18 years had brain scans during completion of a WM task and these showed that a number of brain areas including the frontal lobes and the parietal cortex - were activated (Klingberg, Forssberg & Westerberg, 2002). This finding is consistent with the notion that the central executive would not be located in a single anatomical location because its co-ordinating role requires links to other parts of the brain (Baddeley, 1998). Five studies that measured maths as an outcome used these programmes to train visual-spatial and verbal Short Term Memory (STM) and WM by displaying tasks that involved temporary storage and/or manipulation of sequential verbal or visuospatial information, or a combination of both (Alloway, 2012; Gray et al., 2012; Holmes et al., 2009; Packiam-Alloway et al., 2013; Van der Molen et al., 2010). Results from studies provide evidence of the positive effect of this training on maths achievement (e.g. Alloway, 2012; Holmes et al., 2009). For example, Alloway (2012) found that at post-test mathematical ability had improved significantly more for the WM training group compared to a control.

Metacognition (MC) Interventions

The role of MC has been defined as monitoring the application of cognitive strategies (Flavell, 1979), whereby an individual conducts a ‘self-interrogation’ to ensure their approach is effective (Brown, 1978). MC is broadly thought to consist of metacognitive knowledge and regulation of performance (Lai, 2011). Metacognitive knowledge refers to an individual’s understanding of their own cognition or about cognitions in general (Schraw & Moshman, 1995). This is split into three components: procedural knowledge about use of relevant strategies, conditional knowledge about when and why to use a particular strategy and declarative knowledge about what affects performance (Schraw & Moshman, 1995). Regulation of cognition refers to “metacognitive activities that help control one’s thinking or learning” (Schraw & Moshman, p.354). Regulation is also split into three components: planning which involves selecting appropriate strategies and allocating resources, monitoring through levels of comprehension and task performance and evaluation of the output and process of your performance (Legg & Locker, 2009; Schraw et al., 2006; Schraw & Moshman, 1995).

MC is suggested to develop based on feedback about the efficacy and outcome of strategies to inform how and if they should be used in the future (Kuhn, 2000) and therefore builds up over time (Lai, 2011). Brown (1978) suggested that MC strategies should be taught in an applied context so that individuals can understand how they may be employed and thus appreciate their usefulness. The importance of contextualising MC skills is emphasised in the literature on maths interventions, where MC training is typically delivered in the context of a maths lesson. For example, a number of studies asked participants to solve maths word problems whilst teachers/researchers asked questions and gave MC prompts including: comprehension, to check understanding of the task, connection, to link the current problem to previous knowledge, generation and evaluation of strategies (Hoek, et al., 1999; Kajamies et al., 2010; Mevarech & Kramarski, 1997; Teong, 2003). Studies have predominantly taught MC alongside maths content with the main variation being the method through which maths problems were delivered, such as whole class teaching (Cardelle-Elawar, 1995), small group co-operative working (Verschaffel et al., 2009) or computer programmes (Kramarski & Mizrachi, 2006).

The UK government has developed a one-to-one tutoring programme to support learners struggling in maths which is essentially based on MC principles that equip the learner with self-help routines (Department for Children, Schools and Families, 2009). In one-to-one tutoring the MC processes of a learner are developed through ‘think aloud protocols’ where the learner verbalises their thinking process and the teacher gives prompts to develop MC strategies (Department for Children, Schools and Families, 2009). Despite the fact that this is one of the only maths interventions available in the UK to students aged 14 to 16 years, research has only evaluated this intervention when delivered by less effective, unqualified teachers (Brown, Ireson, Shepherd, Bassett & Rushforth, 2010) or as one of a number of concurrently delivered initiatives to raise performance (Humphrey & Squires, 2011).

The role of anxiety in understanding maths performance

Maths anxiety has been defined as “feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of maths problems” (Richardson & Suinn, 1972, p.551). Factor analysis of maths anxiety measurements has highlighted that this construct relates to two main areas: learning maths anxiety and maths evaluation anxiety (which broadly relates to test anxiety) (Hopko, Mahadevan, Bare & Hunt, 2003). Theoretical frameworks linked to maths achievement have highlighted the detrimental effect on performance of heightened negative emotions (Ashcraft & Kirk, 2001; Ashcraft & Krause, 2007; Prevatt et al., 2010; Rubenstein & Tannock, 2010). Some theoretical frameworks have focused on the association between performance and cognitive processing (e.g. WM) to understand this relationship. For example, processing efficiency theory suggests that anxious thoughts occupy cognitive ‘space’, thus lowering WM capacity (Eysenck & Calvo, 1992). Consistent with this framework, Ashcraft and Kirk (2001) gave undergraduate students a dual task maths problem involving a WM component, where participants were shown a set of letters to be recalled following completion of an addition problem using the carrying operation (thus taxing WM). The results showed that students with higher levels of maths anxiety made more errors on the addition problems. Improving WM skills has also been found to decrease test anxiety. For example, Roughan and Hadwin (2011) showed that WM training with participants aged 12 – 13 years significantly reduced levels of test anxiety. Research examining the link of maths anxiety, WM and maths performance has mostly been conducted with young adults, despite evidence that this anxiety is also present in adolescents (Chinn, 2009).

Aims of this research

The aim of the current study was to run a randomised control trial (RCT) to test the relative effectiveness of a WM intervention and a MC intervention (in the form of one-to-one tutoring) on changes in MC, WM and maths achievement. In addition, it explored whether these interventions would serve to reduce individuals' anxious feelings about maths and whether any reduction in maths anxiety or anxiety more broadly was linked to increased performance on a maths task. Its objective was to inform the development of future interventions to support maths attainment. Participants were randomly allocated to one of the intervention groups and key variables (WM, MC, anxiety (maths anxiety and generalised anxiety) were measured at three time points (pre- and post-intervention and follow-up). Following previous research, it was anticipated that pupils in both groups would improve their maths scores but that this would be achieved via change in different underlying mechanisms. It was hypothesised that the WM training group would improve their maths skills via improved WM capacity and that the one-to-one tutoring group would also improve their maths skills but that this would be via improved metacognition. If the interventions served to increase participants' maths achievement then both interventions should be associated with a decrease in maths anxiety.

Method

Design

A randomised, controlled between groups design was used to investigate whether intervention type, of which there are two levels: computerised Working Memory (WM) training programme (CogMed) or one-to-one maths tutoring (Department for Children, Families and Schools, 2009), could impact on maths achievement, WM, maths anxiety, general anxiety and MC. A between group, repeated across time design was used to explore the impact of time (three levels: pre – T1, post – T2 and follow-up – T3) on the four target variables. Power was calculated using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) which is a software programme which calculates power using Cohen's d. The effect size (ES) used for this calculation was taken as an average of Cohen's d effect sizes from meta-analyses looking at academic achievement following WM training (Melby- Lervåg & Hulme, 2013) (ES = 0.18) and one-to-one tutoring (Higgins et al., 2013) (ES = 0.44) which gave an average Cohen's d effect size of 0.31, thus requiring 22 participants. Therefore if only the two interventions are accounted for in the power calculation then the current study may be slightly underpowered.

Participants

The school from which participants were selected has a demographic of predominantly white British with a low (below national average) proportion of pupils from a low income family and a below average number of pupils with SEN. Inclusion criteria were that a pupil had been chosen by their school to receive one-to-one tutoring which was an intervention that the school were already running and was thus 'treatment as usual'. Selection for this intervention was therefore conducted by the school and was based on the following criteria: pupils who had made age appropriate progress by the end of primary school (Key Stage 2, age 11 years) by achieving National Curriculum (NC) level 3a - 4c . but had not made expected progress from the start of secondary school (age 11/12 years) to the end of KS3 (NC level 5)(NB. KS3 normally finishes at the end of year nine, 14 years old, but this is accelerated in the current school to year eight, 13 years old). Participants were all on roll at one secondary school and were in year nine (age 13/14) of the UK education system during the intervention. Pupils (n=1) were excluded for persistent school absence (below 80%). The remaining thirty participants that had been selected by the school to receive one-to-one tutoring as part of the schools' typical provision, were invited to participate via a letter and participant information sheet sent to their parents. Consent was received for 20 participants.

Following randomisation, unforeseen circumstances led to one-to-one tutoring only being delivered to eight (instead of 10) participants. The WM training group contained eight females and two males (mean age = 13 years 1.7 months, standard deviation = 3.26 months). The one-to-one tutoring group contained three females and five males (mean age = 13 years 2 months, standard deviation = 3.74 months).

Measures

Wide Ranging Achievement Test, 4th Edition (WRAT4) (Wilkinson & Robertson, 2006). This test was used to measure participants' maths attainment. The WRAT4 is designed for use with participants aged 5-94 years. Due to the tiered difficulty of this test's questions it has been shown to be advantageous with maths anxious individuals because it can accommodate lower levels of maths competence often found in this population (Ashcraft & Krause, 2007). There are alternate versions of the test form to minimise practice effects. The test has two sections: oral maths (15 items) and computation (40 items). A point is awarded for each correct answer. For participants who are eight years and older who correctly answer the first five questions of the computation subtest, the oral maths subtest is not administered and credit is given for these 15 items. The computation subtest is timed (15 minutes) and contains questions of increasing difficulty on problems of addition, subtraction, multiplication, division, rounding numbers, decimals, fractions and algebra. The raw scoring range is 0 – 55, which converts into standardised scores. The WRAT4 math computation subtest has been shown to have good concurrent validity (Wilkinson & Robertson, 2006).

Automated Working Memory Assessment (AWMA) (Alloway, 2007). This test is a computerised WM assessment for participants aged 4 -22 years based on the Baddeley and Hitch (1974) model of working memory. Two sub-tests were included - the listening recall sub-test (verbal-WM: VWM) which requires classification of spoken sentences as 'true' or 'false' and when all sentences had been heard the last word in each sentence was recalled. This subtest generates two scores: a memory score which indicates how many words were recalled and a processing score which indicates how many sentences were correctly classified as true or false. The odd one out sub-test (visuospatial-WM: VSWM) displayed three shapes in a grid and asked participants to pick the 'odd-one-out'. At the end of each trial the shape's previous location was located on a blank grid. This subtest also generates two scores: the memory score indicates how often the correct location of the shape was remembered and the processing score indicates how often the 'odd-one-out' was correctly identified. The AWMA has good test-retest reliability, internal validity and diagnostic validity (Alloway, 2007; Alloway, Gathercole, Kirkwood & Elliott, 2008; Alloway, Gathercole & Pickering, 2006).

Junior Metacognition Awareness Inventory (JMAI) (Sperling, Howard, Miller & Murphy, 2002). This questionnaire measures participants' levels of MC and has two versions (A & B) (see Appendix D). Version B is for pupils aged 12-15 years and contains 18 items concerning use of metacognitive strategies, such as, 'I try to use ways of studying that have worked for me before'. The frequency with which these strategies are used is rated on a Likert scale from one to five, where one is never and five is always. Points awarded correspond to the number on the Likert scale, for example a rating of one would earn one point. Points are summed to give a total in the range of 18–90. A higher score indicates a higher level of MC. Research has shown the JMAI has good construct validity (Sperling, Richmond, Ramsay & Klapp, 2012). Internal consistency in the present study was good for pre, post and follow-up testing, $\alpha = .83$, $\alpha = .88$ and $\alpha = .93$.

Abbreviated Maths Anxiety Scale (AMAS) (Hopko Mahadevan, Bare & Hunt, 2003). This is a nine item questionnaire that measured participants' level of maths anxiety (see Appendix E). Respondents are required to rate anxiety levels in maths related situations on a 1-

5 Likert scale, where 1 is low anxiety and 5 is high anxiety. Points awarded correspond to the number on the Likert scale, for example a rating of one would earn one point. Points are summed to give a total in the range of 9–45. A higher score indicates a higher level of anxiety. Research has shown that the AMAS has good test-retest reliability and construct validity (Hopko, 2003). Internal consistency in the present study was good for pre, post and follow-up testing all $\alpha > .8$.

Spence Children's Anxiety Scale (SCAS) (Generalised anxiety items) (Spence, 1998). This test was used to screen for generalised anxiety. This measure is appropriate for children aged 7 to 16 years and the generalised anxiety subscale consists of 6 items which are responded to by indicating the frequency with which they experience the symptom; never, sometimes, often, always (see Appendix F). Responses are scored as follows: never (zero), sometimes (one), often (two) and always (three). Points are summed to give a total in the range of 0–18. A higher score indicates higher levels of anxiety. The SCAS has been found to have good convergent validity and test-re-test reliability (Essau, Sasagawa, Anastassiou-Hadjicharalambous, Guzmán & Ollendick, 2011; Spence, Barrett & Turner, 2003). Internal consistency in the present study was good for pre, post and follow-up testing $\alpha = .85$, $\alpha = .88$ and $\alpha = .93$.

Intervention programmes

Working memory (WM) training (CogMed). A software programme called CogMed (Klingberg et al., 2005) was used to train WM. The researcher received training from Pearson to become a 'CogMed Coach 'and therefore delivered all of the WM training to participants. The coach monitors participant progress through the online tracking system supplied by CogMed. The researcher also took on the role of 'Training Aide' who motivated participants during the programme.

Training was delivered daily for 30–45 minutes over 5-8 weeks. All participants completed 20 out of a possible 25 sessions. The training programme began with an introductory session run by the researcher which explained the key features of the programme. The programme version designed for school aged children was used in this project (CogMed- RM). This consisted of 12 different games that trained visual-spatial and verbal/auditory STM and WM. For example, a verbal/auditory task involved listening to a series of digits and then recalling them in reverse order by selecting the corresponding numbers on an on-screen display. Visual-spatial tasks included a game where a grid displayed a series of lamps that lit up, the grid then rotated 90 degrees and the participant had to recall the order and position the lamps had been in when alight. Eight games were presented during a session and approximately every 5 days different intervention games were selected by the programme (from 12 available) and were rotated to maintain interest. Each game had 15 trials and difficulty level automatically adapted dependent on previous performance.

The programme's motivational features included a computer based 'reward' game called 'Robo-racing' played at the end of each session and positive verbal feedback on performance from the researcher. Users could track progress using an indicator at the side of the screen which indicated the number of remembered items, including the highest level reached during training.

Metacognition intervention (One-to-one tutoring). Ten sessions were delivered twice a week across 5/6 weeks by a qualified teacher who was employed by the participating school to deliver one-to-one tutoring as part of the school's normal provision for children requiring extra support. Sessions lasted approximately 1 hour and content targeted a set of objectives devised by the participating school based on common 'gaps' in low/underachieving pupils' maths knowledge. Sessions covered: place-value, angles, fractions, decimals, percentages, algebraic equations, pie charts, area, volume, real life problem solving (e.g. working out the best mobile phone deal) and shape construction. Sessions followed the same format which included an introduction to the topic area via a maths problem on computer software called 'MyMaths, followed by teacher instruction on relevant procedures and skills. Taught skills were applied by completing questions and then a 'real life' word problem. For example, during the session on area pupils found the area of presented triangles and then applied these skills to find the area of a bathroom. During these tasks the teacher gave MC prompts (e.g. "Today we're looking at X. What do you know about X?", Department for Children, Schools and Families, 2009) to develop comprehension through understanding what the task required and connection through linking the current problem to previous knowledge (Mevarech & Kramarski, 1997). Specific types of metacognitive knowledge (knowledge about your cognitions) were also developed through teacher prompts (Schraw et al., 2006; Schraw & Moshman, 1995), for example, procedural knowledge about what strategies were available was developed through questions such as "Have you covered X in class? Can you remember what you did?" (Department for Children, Schools and Families, 2009), and conditional knowledge about when and why to use a particular strategy were developed through questions such as "What are the important things you need to remember?" (Department for Children, Schools and Families, 2009).

Procedure

Prior to data collection consent was obtained from the school's Headteacher. Opt-in consent letters (see Appendix J) and participant information sheets (see Appendix K) with details about the study, confidentiality and right to withdraw were sent to parents of pupils selected for one-to-one tutoring. Participants were administered pre-test measures in July 2012 (end of academic year) to provide a baseline score of performance. All measures were taken for all participants within a period of 3 days. The WRAT4, JMAI, AMAS and SCAS were administered in paper based form in a small group (five to seven pupils) by the researcher in a quiet room at school. The AWMA is a computerised assessment and was administered one-to-one by the researcher, also in a quiet room at school. The order in which measures were administered at each time point was randomised to reduce potential bias. Based on maths scores from the WRAT4 participants were placed into ability matched pairs by the researcher. An external researcher split the pairs and randomised them into separate groups to create an approximately equal maths ability split across the experimental (n=10) and control (n=8) groups.

The experimental and control interventions began in September 2012 in the first full week of the new academic year. One-to-one tutoring took place at school in a non-teaching room. The experimental intervention (CogMed) was delivered in groups of five in a computer suite on school premises. All participants had their own computer with a set of headphones to minimise environmental disruption and maximise the opportunity to hear the programme's verbal instructions. All sessions were facilitated by the researcher. There were two cohorts of intervention groups. The first ran from September 2012 to mid/late October 2012. The second cohort ran from mid/late October 2012 until late November/mid December 2012.

The researcher gave participants a 50p sticker, collated on a reward chart, for each session attended (template from CogMed). Each week participants would earn in vouchers the face value of stickers collected and if all sessions had been attended a bonus 50p was earned in the first week, a £1.50 bonus in weeks two, three and four and a £2.50 bonus in the final week. A performance based incentive was introduced in week four to boost participant motivation. Performance was reviewed weekly and a 50p sticker was given for each one of the following four criteria that were met: each session was 30+ minutes, each game played that week stayed at or above three items (to be remembered) and improvement over the week could be seen on four or eight games.

Within one to two weeks of intervention completion participants completed all outcome measures in the same manner as at T1. Follow-up (T3) measures were completed within 8-9 weeks of intervention completion and were administered in the same way as at T1/T2. All participants (n=18) completed all pre, post and follow-up measures in the designated time frame. All participants in this study had been selected by the participating school to receive one-to-one tutoring therefore upon completion of follow-up testing by the researcher, participants who had completed the WM training received one-to-one-tutoring from a teacher employed by the school.

Ethics

Ethical approval was obtained from the University of Southampton’s ethics committee (see Appendix G) and Research Governance Office (see Appendix H). Permission to use questionnaires was obtained from test authors (See Appendix I).

Results

Descriptive Statistics

Parametric assumptions were tested for pre (T1), post (T2) and follow-up (T3) data for each dependent variable and group and all were found to be within acceptable limits for normal distribution and homogeneity of variance. Means, standard deviations and range of scores for WM, maths achievement, MC and anxiety at T1, T2 and T3 are displayed in Table 3.

Table 3 *Means, Standard Deviations and Range of Scores for the Standardised Scores for Working Memory and Maths Achievement and for Raw Scores for Metacognition and Anxiety at T1 (pre-intervention), T2 (post-intervention) and T3 (follow-up) for the Working Memory and One-to-One Tutoring Groups*

	Working Memory Training			One to One Tutoring		
Variables	T1	T2	T3	T1	T2	T3
	Mean(S.D.)[Range]	Mean(S.D.)[Range]	Mean(S.D.)[Range]	Mean(S.D.)[Range]	Mean(S.D.)[Range]	Mean(S.D.)[Range]
Maths Achievement	78.40(±6)[69–89]	82.40(±4) [73–89]	82.10 (±6) [71–89]	76.75 (±5) [60–85]	80.75 (±4) [72–86]	83.50 (±5) [74-93]
Working Memory						
Visuospatial	90.80(±11)[72-106]	102.01(±11)[81–116]	110.44(±16)[91–132]	90.65(±12)[75–116]	101.60(±11)[86 –121]	97.86(±14)[78–110]
Verbal	81.99(±13)[63–108]	89.50(±15)[73–131]	89.04 (±14)[68–117]	87.16(±10)[69–99]	89.28(±14)[64–113]	93.74 (±8) 82-108]
Metacognition	53.90 (±9) [39–72]	59.40(±11) [40–80]	56.40(±13) [37–88]	57.50(±10)[42–72]	59.88(±10) [43–71]	56.75(±15) [39-81]
Anxiety						
Maths	26.40 (±7) [19–40]	22.90 (±6) [15–35]	23.20 (±7) [12–37]	27.25 (±9) [15–41]	23.25 (±6) [15–33]	21.25(±10) [12-43]
General	8.50(±4) [3–14]	8.60 (±5) [1–15]	8.70 (±6) [1–17]	5.88 (±3) [2–10]	5.50 (±2) [1–16]	6.00 (±5) [0–18]

Correlations between all T1 measures were calculated and only VSWM and MC were significantly correlated, (*r* = .42, *p* = .04, Table 4 and Figure 3). VSWM and VWM approached significance, (*r* = .34, *p* = .09), as did maths anxiety and general anxiety, (*r* = .38, *p* = .06, see Table 4). All other correlations were non-significant, with all *rs* < .30, and *ps* > .1.

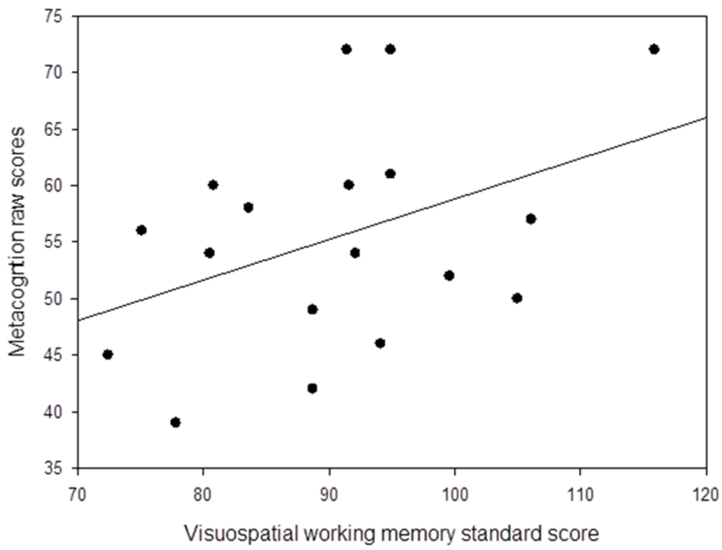
Table 4 *Correlations Between T1 Standardised Scores for Working Memory and Maths with Raw Scores for Metacognition and Anxiety*

Variables	1	2	3	4	5	6
1. Maths achievement	-	.25	-.00	-.14	.20	.04
Working memory						
2. Visuospatial		-	.34#	.42*	.30	.05
3. Verbal			-	.11	-.30	.14
4. Metacognition				-	.29	-.04
Anxiety						
5. Maths					-	.38#
6. Generalised						-

* one-tailed *p* < .05   # one-tailed *p* < .1

An independent samples t-test was conducted on all T1 data to assess if there were pre-existing significant differences between the experimental and control groups. Results indicated no significant differences between scores on any measure with all *ts* <1.5 and all *ps* > .1.

Figure 3 Correlation of T1 (pre-intervention) Visuospatial Working Memory Standardised Scores with Metacognition Raw Scores



Approach to analysis

In order to examine the impact of WM training and one-to-one tutoring on maths achievement, WM (visuospatial, verbal), MC and anxiety (maths, general), group differences were explored over the three time points (T1, T2 and T3). Four Group (2) x Time (3) repeated measures ANOVAs (repeated for time) were conducted on the dependent variables (maths-achievement, VSWM, VWM and MC). In addition, exploratory analyses were conducted to assess the effect of the interventions on maths anxiety and general anxiety. Standard scores were analysed for WM and maths achievement, and raw scores from questionnaire data were analysed for MC and anxiety (maths, general). In addition to considering statistical significance Effect Sizes (ES) as measured by Partial Eta Squared were reported as small (>.01), medium (>.06) and large (>.14) (Richardson, 2011, 2009), for all analyses because when sample size is small, strong and important effects can be missed (Levine & Hullett, 2002). Partial Eta Squared was used because this was the effect size calculated by SPSS.

Jacobson and Truax (1991) argued that statistical significance and ES only give information about group differences and may not capture meaningful changes that occur on an individual basis. Reliable change scores were therefore also calculated to derive a distribution of scores that would be expected if no actual change had occurred, against which observed scores could be measured to provide an indication of the magnitude of change that had occurred for individuals during the intervention (see Jacobson & Truax, 1991). Evans, Margison and Barkham (1998) define reliable change as “the extent to which the change shown by an individual falls beyond the range which could be attributed to the measurement variability of the instrument itself" (p.70). Measurement variability is measured by calculating the Standard Error (SE) of the measurement (see formula below) and Evans et al., (1998) state that "Change exceeding 1.96 times this SE is unlikely to occur more than 5% of the time by unreliability of the measure alone. Reliable change scores were calculated to consider change on each dependent variable for every participant between T1 and T2 and between T1 and T3. These scores were calculated using the following formula taken from Jacobson and Truax (1991):

$$RC = \frac{x_2 - x_1}{S_{diff}}$$

This formula is broken down into smaller steps which are as follows:

$$S_{diff} = \sqrt{2(S_E)^2}.$$

$$S_E = s_1\sqrt{1 - r_{xx}}$$

In this formula RC stands for 'Reliable Change, x^1 represents the pre-test score of an individual on a particular measure and x^2 is the corresponding post-test score, S_{diff} represents the standard error of difference between two test scores, S_E is the standard error of measurement, s^1 is the standard deviation of all participants at pre-test for the measure in question and r_{xx} represents the test re-test reliability of this measure. This formula therefore produces a measure of change using standardised units. The process through which these scores were calculated was as follows:

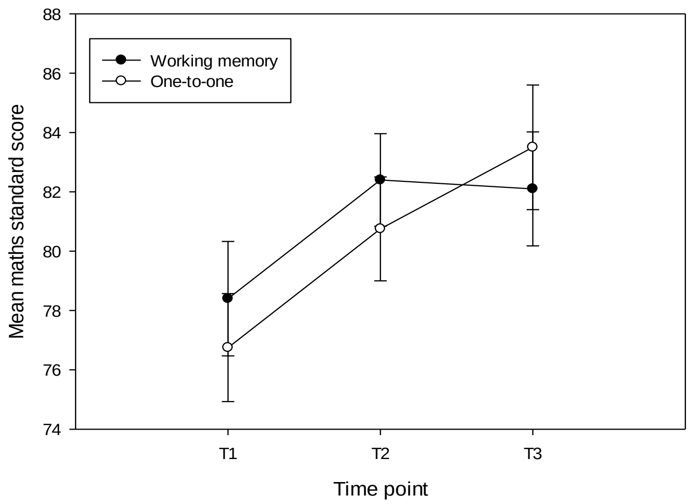
1. Test-re-test reliability was calculated by correlating pre and post scores on a particular measure (e.g. WRAT4).
2. The standard deviation of participants' pre-test scores on that measure were calculated.
3. These values were then put into the S_E formula $S_E = s_1\sqrt{1 - r_{xx}}$
4. The resulting value from this formula was then used in the S_{diff} formula $S_{diff} = \sqrt{2(S_E)^2}$.
5. Change scores were calculated for each individual as per the RC formula $RC = \frac{x_2 - x_1}{S_{diff}}$ and these values were divided by the S_{diff} value for that measure to give a standardised reliable change score for each participant.

A meaningful improvement was indicated by values greater than 1.96 and a meaningful decline in performance was indicated by scores less than – 1.96. For the purpose of clarity the formula has been re-arranged to become, 1.96*Sdiff = (Post test score - Pre-test score), in order to calculate based on the original rather than standardised score, the change score (post - pre), that is required on a particular measure in order to meet the criteria for reliable change. A change score is the difference between a particpants' pre (T1) and post (T2) or pre and follow-up (T3) scores on a measure and therefore represents the change in their scores over time.

Change in Maths Achievement

Results showed that maths scores significantly improved over time and with a large partial eta squared effect size ($F(2, 32) = 14.94$, $p < .001$, $\eta_p^2 = = .65$). Post-hoc analyses using pairwise comparisons indicated that there were significant improvements in scores from T1 to T2 and from T1 to T3 but not from T2 to T3. Results from the repeated measures ANOVA indicated that there was no significant interaction between time and group ($F(2, 32) = 1.55$, $p = .23$, $\eta_p^2 = = .09$). There was no significant between group effect, ($F(1, 16) = .07$, $p = .79$, $\eta_p^2 = = .004$, see Table 3 and Figure 4).

Figure 4 Mean Standardised Maths Achievement Scores and Standard Errors at T1, T2 and T3 for the Working Memory and One-to-One Tutoring Group



Calculations on individual scores from T1 to T2 suggested that two participants from the WM training group and three from one-to-one tutoring made positive reliable change as indicated by achieving a change score of 7 or more. This number fell to one participant from each group when calculated for T1 to T3 scores which was evidenced by a change score of 9 or more These findings suggest that both interventions were equally effective over time at reliably improving maths achievement.

Working Memory

Visuospatial working memory. Results from the repeated measures ANOVA[1] found no evidence of significant between group differences ($F(1,16) = .65$, $p = .43$, $\eta_p^2 = .04$).

However, findings indicated that scores on VSWM had significantly improved over time, with a large partial eta squared effect size ($F(2, 32) = 15.17$, $p < .001$, $\eta_p^2 = .49$). Pairwise comparisons indicated that there was significant improvement from T1 to T2 and from T1 to T3, but not from T2 to T3. Results also indicated a significant interaction between time and group with a large partial eta squared effect size ($F(2, 32) = 3.72$, $p = .04$, $\eta_p^2 = .19$, see Figure 5). Independent t-tests were conducted on T1, T2 and T3 scores which indicated that there were no significant differences between groups for any time point.

The effect of time was considered in each group separately using repeated measures ANOVAs. The WM group showed significant improvements in VSWM scores over time with a large effect size ($F(2,18) = 16.39$, $p < .001$, $\eta_p^2 = .65$). Pairwise comparisons highlighted that this significant improvement was from T1 – T2, T1 – T3 and from T2 – T3. VSWM also significantly improved over time in the one-to-one tutoring group, with a large partial eta squared effect size ($F(2,14) = 4.02$, $p = .04$, $\eta_p^2 = .37$). Pairwise comparisons confirmed that this significant improvement in scores was only found between T1 and T2, with non-significant differences between T1 and T3 and T2 and T3.

Calculations on individual scores from T1 to T2 suggested that three participants from the WM training group and two from one-to-one tutoring group made positive reliable change by achieving a change score of 18.1 or more. From T1 to T3 participants making positive reliable change had a change score of 20.2. Three WM training participants had made positive reliable change over this period, but this had dropped to one in the one-to-one tutoring group.

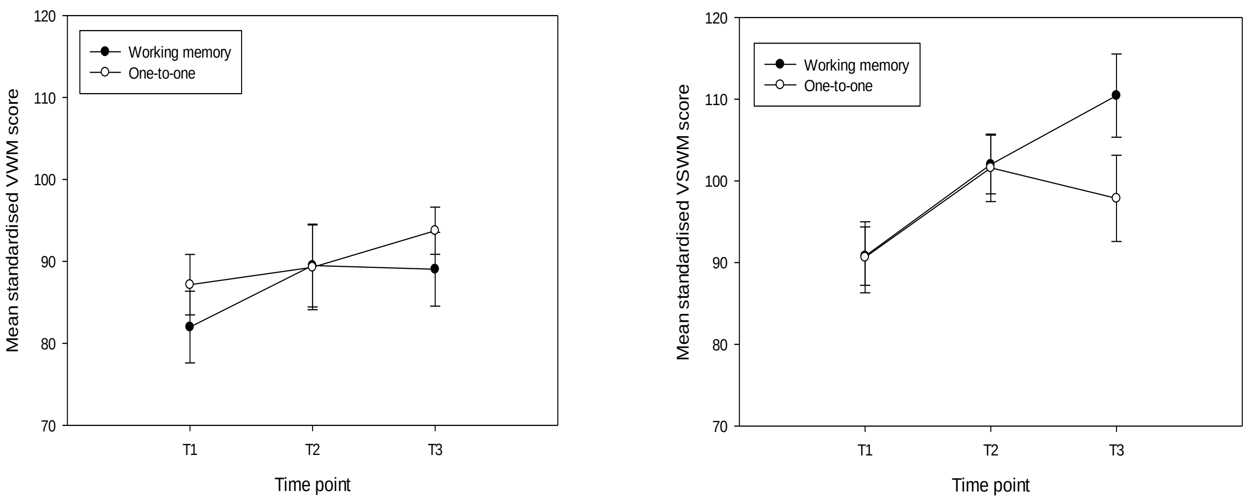
---

[1]ANOVAs were conducted on VSWM processing standard scores and results showed the same pattern of findings as ANOVAs conducted on VSWM memory standard scores.

Verbal working memory. Results from the repeated measures ANOVA[2] revealed a significant improvement over time on scores of VWM, with a large partial eta squared effect size ($F(2, 32) = 3.78$, $p = .03$, $\eta_p^2 = .19$). Pairwise comparisons indicated that there were significant improvements in scores from T1 to T3 but not from T1 to T2 or from T2 to T3 (see Figure 5). There was no significant interaction between time and group ($F(2, 32) = .69$, $p = .51$, $\eta_p^2 = .04$) and there was no significant between group difference ($F(1,16) = .33$, $p = .58$, $\eta_p^2 = .02$).

Calculations on individual scores from T1 to T2 suggested that three participants, all in the WM training group, made positive reliable change by achieving a change score of 21,2 or more. From T1 to T3 this fell to two participants who were in the WM group who all achieved a change score of 20.3 or more.

Figure 5 Mean Standard Scores and Standard Errors for Verbal and Visuospatial Working Memory at T1, T2 and T3 for the Working Memory and One-to-One Tutoring Groups



[2]ANOVAs were also conducted on processing standard scores generated during the VWM subtest. Results found no evidence of change over time, no interaction of time and group and no between group difference.

Metacognition

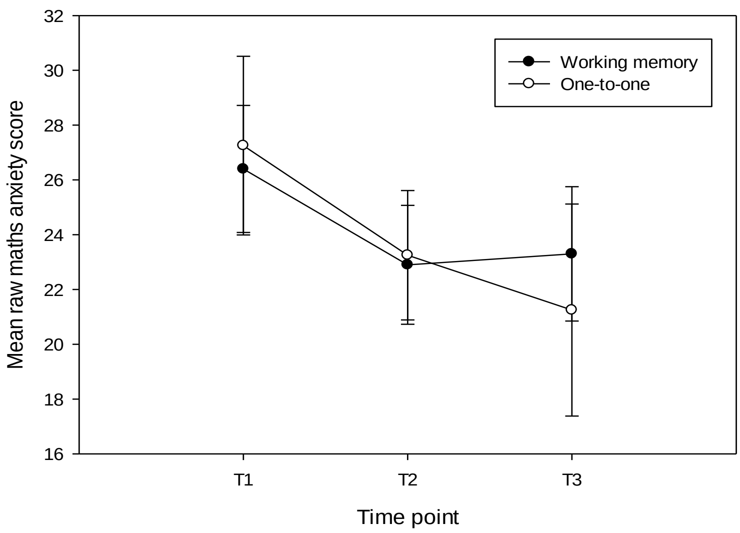
Results showed no effect of time or group on MC and no significant interaction (in all cases F < 2 and p > .1)

Anxiety

Maths anxiety. Results from the repeated measures ANOVA indicated a significant reduction in maths anxiety over time, with a large partial eta squared effect size ($F(2, 32) = 7.48$, $p = .002$, $\eta_p^2 = .32$). Pairwise comparisons indicated that there were significant reductions from T1 to T2 and T1 to T3, but not from T2 to T3 (see Figure 6). There was no significant interaction between time and group ($F(2, 32) = .76$, $p = .48$, $\eta_p^2 = .05$) and no significant between group difference ($F(1,16) = .01$, $p = .94$, $\eta_p^2 = .01$).

Calculations on individual scores from T1 to T2 suggested that two participants, one from the WM training and one from one-to-one tutoring group, had made positive reliable change by achieving a change score of 12 or more. From T1 to T3 two participants had still made reliable change (both from the one-to-one tutoring group) indicated by a change score of 10 or more.

Figure 6 Maths anxiety scores and standard errors on the AMAS for T1, T2 and T3 for the working memory and one-to-one tutoring groups.



General Anxiety. Analysis of scores from the repeated measures ANOVA found no evidence of a significant effect of time or group or the interaction on general anxiety (in all cases F < 2 and p > .1).

Understanding Change as a Result of the Interventions

It was hypothesised that MC and WM would affect maths achievement. Therefore correlations were conducted to consider whether changes in maths achievement were linked to changes in WM or MC over the course of the intervention. Change scores were calculated between T1 and T2 for each dependent variable by subtracting T1 scores from T2 scores. Change scores were calculated between T1 and T3 by subtracting T1 scores from T3 scores. A positive change score value on measures of maths achievement, MC and WM indicated an improvement in scores over time and a negative value meant a decrease in scores. For maths anxiety and general anxiety a positive value indicated an increase in anxiety whereas a negative value indicated a decrease in anxiety over time.

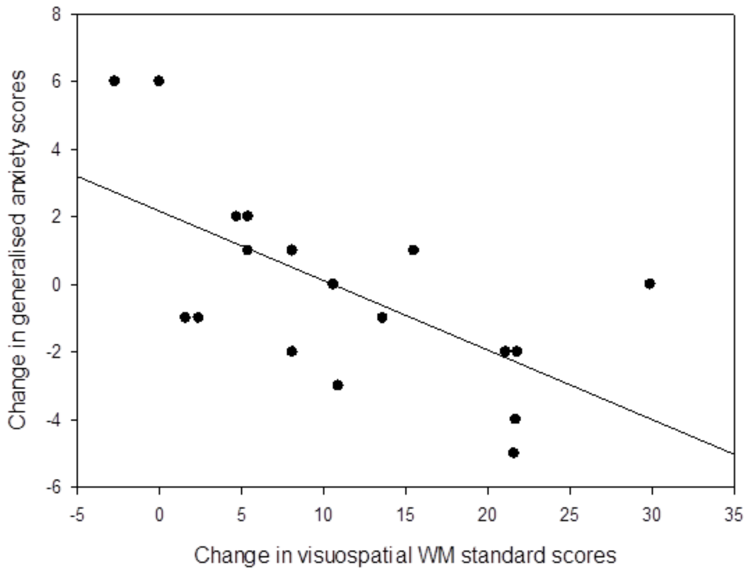
Table 5 *Correlations Between T1 – T2 Change Scores for Working Memory, Maths, Metacognition and Anxiety*

Variables	1	2	3	4	5	6
1. Maths	-	.06	-.12	-.21	-.06	-.04
Working Memory						
2.Visuospatial		-	.25	.24	-.12	-.64**
3. Verbal			-	.22	-.12	-.02
4. Metacognition				-	-.05	-.22
Anxiety						
5. Maths					-	-.02
6. General						-

** one tailed p<.01

Correlations between the T1 to T2 change scores showed no association between change in maths and any other variable (see Table 5). The only significant change was found between VSWM and general anxiety, indicating that a positive change in VSWM was linked to decreased anxiety (see Table 5 and Figure 7). (This result was also found when change scores from each group were separately correlated: WM group, r = -.60, p (one tailed) = .03, One-to-one tutoring, r = -.79, p (one tailed) = .02.)

Figure 7 Correlation of Change from T1 (pre-intervention) to T2 (post-intervention) of Raw General Anxiety Scores and Standardised Visuospatial-WM Scores.



Correlations were calculated between change scores from T1 to T3. These results indicated that change in maths improvement over time only significantly correlated with change in maths anxiety scores. Maths anxiety also significantly negatively correlated with metacognition and positively correlated with general anxiety (see Table 6).

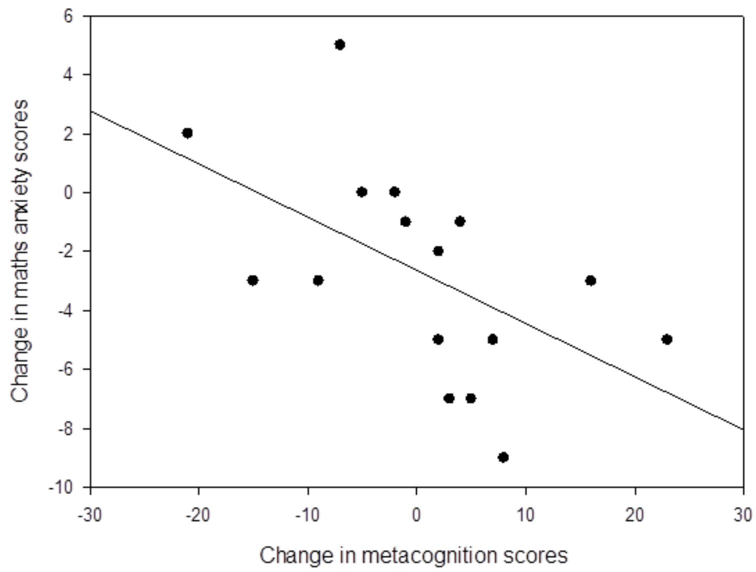
Table 6 *Correlations Between T1 – T3 Change Scores for Working Memory, Maths, Metacognition and Anxiety*

Variables	1	2	3	4	5	6
1. Maths	-	-.20	-.21	-.07	-.47*	-.27
Working Memory						
2.Visuospatial		-	.37#	.16	-.00	-.24
3. Verbal			-	-.01	-.25	-.10
4. Metacognition				-	-.40*	.36#
Anxiety						
5. Maths					-	46*
6. General						-

* one tailed *p*<.05 # one tailed *p*<.1

Two participants were subsequently removed from the analysis as their maths anxiety change scores were more than 2.5 standard deviations from the mean. After these removals the correlation between maths achievement and maths anxiety became non-significant (*r* = -.18, *p* (one tailed) = .25), but the correlation between MC and maths anxiety became more significant, (*r* = -.54, *p* (one tailed) = .02, see Figure 8), as did the correlation between maths anxiety and general anxiety (*r* = .59, *p* (one tailed) = .01). Correlations approaching significance were between MC and general anxiety (one tailed) *p* = .07), and VSWM and VWM (one tailed) *p* = .06). However, considering groups separately, for the WM training groups' scores only the VSWM change scores significantly correlated with general anxiety (*r* = -.61, *p* (one tailed) = .03), and VWM (*r* = .62, *p* (one tailed) = .03), whilst the correlation with maths anxiety approached significance (*r* = -.54, *p* (one tailed) = .06).

Figure 8 Correlation with Outliers Removed of Change from T1 (pre-intervention) to T3 (follow-up) of Raw Maths Anxiety and Metacognition Scores



Discussion

The aim of the current study was to explore whether maths achievement could be improved in individuals with Maths Difficulties (MD) through a WM or MC intervention (one-to-one tutoring). It was expected that both types of intervention would improve scores in maths but that change would be achieved via different pathways; through improved WM capacity in the WM training group and through increased MC in the one-to-one tutoring group. Results showed that both interventions were associated with improvements in maths achievement, increases in VSWM and VWM and with reductions in anxiety. In general, the magnitude of improvement was similar for both interventions; however group-based and individual analyses indicated that improvements in WM were greater and more sustained in the participants who received the WM intervention. The results also found no evidence to suggest that MC improved following either intervention. Furthermore, findings did not indicate that change in WM capacity or MC was directly linked with improvements in maths achievement. However, there was evidence to suggest that improvement in WM and MC following the interventions was linked to reductions in anxiety. The findings suggest that WM training and one-to-one tutoring are effective interventions to improve maths achievement and WM and to lower anxiety. The focus in the current study on the effect of maths achievement in two distinctive interventions represents a novel contribution to the research literature.

The finding in the current study that maths achievement significantly improved over time was consistent with some previous WM training studies, for example Holmes et al. (2009), St-Clair-Thompson et al. (2010) and Van der Molen et al. (2010) (but see studies by Gray et al., 2012 and Packiam-Alloway et al., 2013 that found no improvement in post-intervention or at follow up respectively following a WM intervention). Similar to the current study, previous studies have also found improvements in WM that were significantly greater over time in the WM training group than a comparison (passive control) group (Holmes et al., 2009; Alloway, 2012). No previous WM training studies that measured maths achievement have also measured anxiety, however some studies have found that WM training can significantly reduce test anxiety (e.g. Roughan & Hadwin, 2011). The current study extended this research by considering the impact of WM training on anxiety (anxiety specific to maths and anxiety more generally) in pupils with maths difficulties.

Findings from the current study indicate that WM capacity is important in maths achievement. The WM training group made comparable improvements in maths achievement as the one-to-one tutoring intervention, suggesting that WM capacity is equally beneficial to maths achievement as explicitly developing subject specific maths knowledge. This finding may be understood in relation to the theoretical frameworks of Butterworth et al. (2011) and Demetriou et al. (2002) which propose that WM provides the cognitive capacity to process new mathematical knowledge, and teaching new mathematical concepts will lead to academic benefits if there is sufficient cognitive ‘space’ to process them.

Increases in WM following training can be explained in relation to the hypothesised plasticity principle, which suggests that the capacity of WM can be increased via training due to improved brain connections that facilitate changes in information processing (Klingberg et al, 2002; Takeuchi, et al., 2010). Increased brain activity observed during WM training has been found in the caudate nucleus, prefrontal and parietal cortex (Klingberg, 2010). Links have also been found between these brain areas and performance on maths tasks using WM. For example, Stocco, Lebiere, Randall, O’Reilly and Anderson (2012) gave undergraduate students previously learned algebra problems and novel algebraic equations that required WM to process the new task information. Measurements using fMRI scans showed that the parietal cortex and the caudate nucleus were more active whilst processing novel problems which required WM. Plasticity of brain connections is only a tentative explanation of WM capacity changes because as noted by Holmes et al. (2009) “the nature of the cognitive and neural changes underpinning gains in WM with this adaptive training program are yet to be fully understood” (p.14).

The findings of the current study are consistent with previous research which has found that previous MC interventions showed improved maths achievement (e.g. Mevarach & Kramarski, 1997). The significance of the current paper is that it has shown that a MC intervention is equally effective as one which targeted cognitive factors known to be associated with achievement in school. The current study therefore adds to the evidence for use of the Department for Children Families and School’s one-to-one tutoring programme and supports the government’s move to evidence based practice in schools (Goldacre, 2013). The current study’s finding that there were no significant improvements in MC either over time or between groups is contrary to findings of previous MC interventions, although a number of these studies did not include a measure of MC.

The current study hypothesised that both interventions (WM training and one-to-one tutoring) would improve maths achievement. It was also hypothesised that any increase in maths would be best explained by significant changes in the factors that were specifically targeted, i.e., that achievement in the MC group would be linked to changes in participants’ reports of their MC skills and changes in the WM groups would show associations with increased WM capacity. The results showed that maths was improved in both interventions, but contrary to prediction WM had also improved in both interventions and MC had not significantly improved in either WM training or one-to-one tutoring.

ROLE OF METACOGNITION AND WORKING MEMORY IN MATHS

Improvement in maths scores in the one-to-one tutoring group may be explained by the development of a self-regulated approach to learning through encouragement of pupils to check understanding of task requirements and to draw on prior knowledge (Department for Children, Schools and Families, 2009). For example, research by Eshel and Kohvi (2007) with participants aged 12-13 years showed that if an individual had high levels of self-regulation they also had high levels of maths achievement. Self-regulation encompasses both MC defined as 'regulation' of knowledge and strategies and control of behaviour based on motivation to engage in learning tasks (Zimmerman, 2001). It is therefore possible given the links between motivation and maths achievement (Luo et al., 2011) that one-to-one tutoring developed the motivation aspect of self-regulation more than the MC aspect and that this increased motivation led to the observed improvements in maths achievement. The teacher's sole focus on one pupil during one-to-one sessions may have helped to keep the pupil on task and supported them to learn effective strategies. This is supported by research which has demonstrated that teacher intervention during students' maths tasks led to improved pupil concentration and use of more appropriate strategies, thus ultimately improving performance (Chui, 2004).

An alternative interpretation of the observed improvement in maths scores is that over time good teaching in maths lesson led to statistically significant improvements in participants' competence in maths. Due to the fact that the current study did not have a passive control group who only received in class maths teaching it is not possible to conclusively state that the observed results were not a product of maturational improvements in maths. However when considering this interpretation of results it is important to note that participants were chosen for the current study because after receiving two years of whole class maths teaching at the participating school they had failed to make age appropriate progress. It is therefore perhaps unlikely that whole class teaching alone would significantly accelerate maths achievement however this interpretation cannot be discounted. It must also be considered that an improvement in maths achievement may have been the result of participants becoming more familiar with the researcher administering the tests and that this reduction in anxiety may have contributed to improved test scores over time.

The non-significant change in MC found in the current study may be understood through theoretical frameworks developed to understand academic performance. For example, the 'Good Information Processor' framework by Pressley, Borkowski and Schneiders (1989) suggests that effective information processors have good MC strategies but that use of these strategies is based on levels of self-efficacy. Therefore because low levels of maths achievement are associated with low self-efficacy (Barkatsas, Kasimatis & Gialamas, 2009), participants may not have felt sufficiently confident to give themselves a high MC rating. This is supported in research by Kleitman and Stankov (2011) who found a significant positive correlation between undergraduate students' self-reported levels of MC and self-efficacy, suggesting that the more confident an individual the higher their rated MC. Furthermore, because the current study tested a population of low/under achieving pupils who can have difficulty processing verbal information, the language based nature of MC may have made it more difficult for them to learn. For example, Abedi and Lord (2001) asked 11 -12 year olds to complete standard maths word problems and simplified word problems with shortened sentences containing minimal information. It was found that low-achievers in maths scored significantly better on problems with simplified language than on standard problems, but this difference was not found for high-achievers.

Improvements in WM in the one-to-one tutoring group may be explained through the theoretical framework proposed by Butterworth et al. (2011), which suggests that provision of appropriate educational experiences (such as one-to-one tutoring) facilitates the development of cognitive functions such as WM. However, findings indicate that one-to-one tutoring had more of an impact on VSWM than VWM based on the small effect found for change in VWM compared to the large effect size found for VSWM. As pupils who are low/underachieving in maths can have difficulties processing language this may explain participants' less significant improvement in VWM (Abedi & Lord, 2001). Improvements in WM may also be explained by one-to-one tutoring facilitating a change in the metacognitive appraisal of task difficulty. For example, Autin and Croizet (2012) conducted a RCT where participants aged 11 years completed anagrams, the difficulty of which had been gauged so that no participant would complete them in the time allocated. Only participants in the experimental condition were given feedback that reframed their metacognitive appraisal of task difficulty by communicating the expectation that no-one should have been successful. A listening span WM test was then given and the experimental group scored significantly higher than a control group, demonstrating that the changing of MC task appraisal had improved WM capacity. A further explanation of the improvement in WM during one-to-one tutoring may be that it improved the episodic buffer component of WM. The episodic buffer is proposed to conduct some central executive roles such as updating, through the processing of visuospatial and verbal information (Baddeley, 2000). It is an under-researched area and therefore links to this aspect of WM are speculative, however there is some evidence for its connection to MC. For example the MC activity of planning which highlights relevant information and strategies (Schraw & Moshman, 1995) strongly resembles the buffer's role of retrieving relevant information from LTM to give meaning to stored verbal and visuospatial information (Baddeley, 2002). This role raises the possibility that MC strategies taught during one-to-one tutoring, such as error detection through monitoring, could link to executive functions controlled by the episodic buffer such as updating task knowledge (Fernandez-Duque, Baird & Posner, 2000; Shimamura, 2000) and these types of improvements may have led to a measurable improvement in WM.

The current study found a significant link between reductions in general anxiety and improvements in WM capacity and a significant link between improvements in MC and a reduction in maths anxiety. The finding that maths anxiety reduces over time represents a novel contribution to the literature and suggests that WM training and one-to-one tutoring have wider benefits that lead to fewer self-reported worries around performance in maths. The link between improved WM and reduced general anxiety may be understood through the processing efficiency theory (Eysenck & Calvo, 1992) where reductions in anxiety may have led to less cognitive space being occupied by anxious thoughts, therefore leaving more capacity to process WM task demands. This may have contributed to the increases in WM capacity found over time in both intervention groups. The link between improved MC and reduced maths anxiety may be interpreted as MC improvements allowing individuals to direct their attention away from anxiety provoking thoughts. For example, the MC process of planning is thought to occur through inhibition of irrelevant information such as anxious thoughts (Fernandez-Duque et al., 2000). This is demonstrated empirically in a study by Veenman, Kerseboom & Imthorn (2000). Metacognitive skilfulness was measured for participants aged 12 to 13 years through 'think aloud protocols', where participants verbalised their thought process during completion of a maths problem and researchers rated this for evidence of MC skills such as prediction and evaluation. Highly anxious participants were rated as lower in metacognitive skilfulness compared with less anxious peers, but when metacognitive cues were given such as prompts to review strategy use, effects of anxiety were ameliorated so that significant improvements were made.

Limitations

The current study found positive benefits of WM training and one-to-one tutoring on maths achievement, however there are limitations in this research. The sample size used was small and therefore the study may have been underpowered to detect additional or more significant effects than those found (Field, 2009). Furthermore, the current study lacked a waitlist control although the finding that maths improvements were evident on standardised scores makes results more robust. An alternative research design that may have strengthened the ability to infer a causal relationship between the interventions and observed improvements in measured variables would have been to compare the WM and one-to-one tutoring groups with a group which had completed both interventions. This would have helped highlight if the interventions had additional or different benefits. Comparison of these groups to a waitlist control would have helped to further understand the effects of WM training and one-to-one tutoring. Methodological limitations also included a lack of blinding about condition allocation therefore participants may have developed expectations about their performance which may have created performance bias, and the researcher may have been subject to detection bias (Higgins & Green, 2011). The non-significant finding of change in MC may have been related to the chosen measurement instrument because the JMAI measures MC generally and thus the specific application of MC to maths may not have been captured. A final limitation was the language used in the maths anxiety questionnaire, because it was perhaps more suited to an American than a UK population due to the inclusion of phrases such as 'pop-quiz'. This type of language may have been unfamiliar to UK pupils and therefore may have affected their responses, although the researcher did encourage participants to ask if they did not understand anything.

Future Research and Implications

The current study has shown that relatively short term interventions can have important implications for increasing achievement and lowering subject specific worries in school for low achieving pupils. A further study with a larger sample size will be important to further understand mechanisms underpinning significant changes in maths. Future research should also aim to examine a wider range of factors (such as self-efficacy) that might be important in understanding change. To inform future interventions further research is needed to better understand the link between anxiety and MC and to explore what contributes to a reduction in general and maths anxiety.

Based on results of the current study future research may consider creating an intervention that combines WM and MC training in order to deliver key benefits from both training programmes. Demetriou et al (2002) suggested in their framework that WM capacity must first be increased to allow sufficient cognitive processing capacity to effectively use MC strategies. As WM training has been shown to be effective at developing sustained improvements in VSWM that is important for maths tasks (Holmes & Adams, 2006; Kyttälä & Lehto, 2008), pupils may first undertake WM training before receiving MC support alongside maths instruction. This approach may support the growth of MC skills by providing sufficient WM capacity to process MC strategies (Demetriou et al., 2002). Furthermore, given the links in the current study of WM training and reductions in general anxiety and MC training and reductions in maths anxiety, an intervention that explicitly supports both WM and MC may serve to reduce both types of anxiety.

The current study has highlighted the efficacy of WM training and one-to-one tutoring as interventions that can significantly improve maths achievement and WM, whilst significantly decreasing maths anxiety. The potential role of MC in this process is perhaps more linked to MC knowledge (knowing what you know) rather than MC strategies (knowing what strategy would work best) because the observed executive function improvements link most closely to this definition of MC. The conceptualisation of MC as relating to the episodic buffer highlights a new avenue of research to investigate how this component of WM may be explored in future intervention studies, thus addressing some previous criticisms of WM training (e.g. Apter, 2012). This research contributes to government policy to increase evidence based practice (Goldacre, 2013) and given the increased emphasis on schools to provide effective interventions (Department For Education, 2013), this research may inform educational provision for young people to raise maths skills which are crucial for future employability (Wolf, 2011).

Appendices

Appendix A: List of search terms used in the systematic review

Searches were conducted in each database for each type of intervention (Working Memory or Metacognition). The search terms were first entered and then limiters were applied to retrieve studies which met inclusion criteria, for example 'peer reviewed journal' and 'English language'. Search terms were initially generated by the researcher with additional terms added based on keywords from relevant articles found during the search process.

1. PsycInfo (via Ebsco; 1887-2013): All search results from the search terms below were filtered by age: childhood (birth – 12) and adolescence (13-17), type of journal: 'peer reviewed' and language: 'English language'.

Working memory training AND Mathematics AND Executive Function

Working memory AND Intervention AND Mathematics

Metacognition AND Mathematics AND Intervention

Metacognition AND Mathematics achievement

Training AND Metacognition

Training AND Mathematics

Metacognition AND Word problem solving

2. Web of Science (via Web of Knowledge; 1950-2013): All search results from the search terms below were filtered by type of publication: 'peer reviewed' and 'article', and language: 'English'

Working memory training AND Mathematics AND Executive Function

Working memory AND Intervention AND Mathematics

Metacognition AND Mathematics AND Intervention

Metacognition AND Mathematics achievement

Training AND Metacognition

Training AND Mathematics

Metacognition AND Word problem solving

3. EMBASE (via OVID; 1980-2013): All search results from the search terms below were filtered by publication type: 'Journal, article'.

Working memory training AND Mathematics AND Executive Function

Working memory AND Intervention AND Mathematics

Metacognition AND Mathematics AND Intervention

Metacognition AND Mathematics achievement

Training AND Metacognition

Training AND Mathematics

Metacognition AND Word problem solving

4. MEDLINE (via EBSCO: 1973-2013): All search results from the search terms below were filtered by language: 'English Language', publication type: 'Academic Journals' and age: 'Child, preschool: 2-5years', 'Child: 6-12 years', 'Adolescent: 13-18years'.

Working memory training AND Mathematics AND Executive Function

Working memory AND Intervention AND Mathematics

Metacognition AND Mathematics AND Intervention

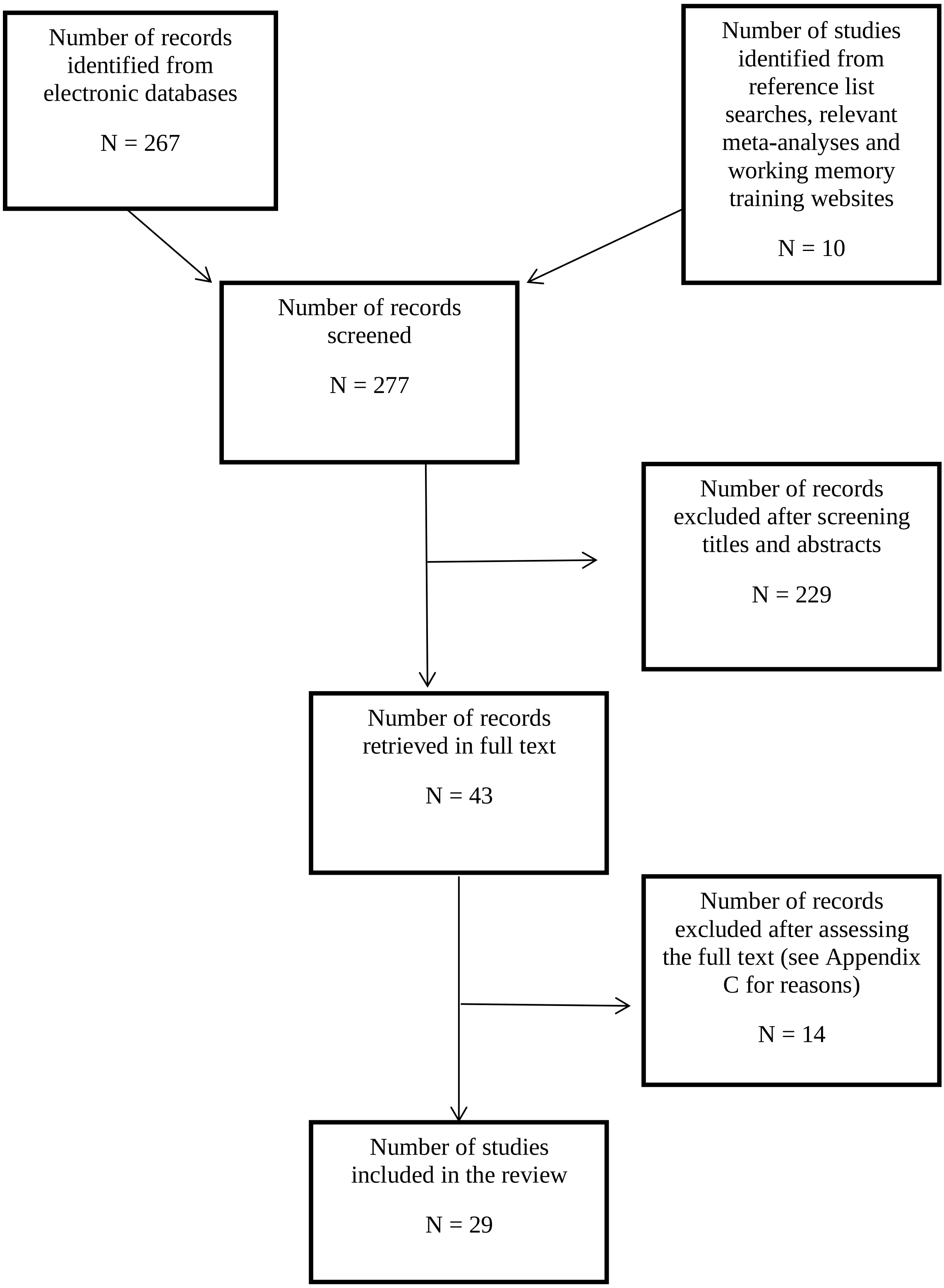
Metacognition AND Mathematics achievement

Training AND Metacognition

Training AND Mathematics

Metacognition AND Word problem solving

Appendix B: Flow chart of inclusion and exclusion of records from the systematic review



Appendix C: Criteria for excluding papers for which full text was obtained

Following the screening of titles and abstracts, 43 papers were identified as relevant and retrieved in full text. After obtaining full text, a further 14 papers were excluded for the following reasons.

1. Papers presenting a review of research rather than original research (n = 1)
2. Papers that included data from adult participants (n = 1)
3. Papers in which the training did not fit inclusion criteria (n = 7)
4. Studies that did not measure maths as an outcome (n = 3)
5. Duplication of records (n = 2)

Appendix D:
Junior Metacognitive Awareness Inventory (Sperling, Howard, Miller & Murphy, 2002)

We are interested in what learners do when they study. Please read the following sentences and circle the answer that relates to you and the way you are when you are doing school work or home work. Please answer as honestly as possible.
1 = Never 2 = Seldom 3 = Sometimes 4 = Often 5 = Always

1. **I know when I understand something**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

2. **I can make myself learn when I need to.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

3. **I try to use ways of studying that have worked for me before.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

4. **I know what the teacher expects me to learn.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

5. **I learn best when I already know something about the topic.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

6. **I draw pictures of diagrams to help me understand while learning.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

7. **When I am done with my schoolwork, I ask myself if I learned what I wanted to learn.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

8. **I think of several ways to solve a problem and then chose the best one.**

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

9. I think about what I need to learn before I start working.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

10. I ask myself how well I am doing while I am learning something new.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

11. I really pay attention to important information.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

12. I learn more when I am interested in the topic.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

13. I use different learning strengths to make up for my weaknesses.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

14. I use different learning strategies depending on the task.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

15. I occasionally check to make sure I'll get my work done on time.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

16. I sometimes use learning strategies without thinking.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

17. I ask myself if there was an easier way to do things after I finish a task.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

18. I usually decide what I need to get done before I start a task.

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

Appendix E:

Abbreviated Math Anxiety Scale (Hopko, Mahadevan, Bare & Hunt, 2003)

Rate on a scale from 1- 5 how anxious you would feel in each of these situations, where: 1 is 'not at all anxious', 2 is 'a little anxious', 3 is 'fairly anxious', 4 is 'quite anxious' and 5 is 'very anxious'

1. **Having to use the tables in the back of a math book.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

2. **Thinking about an upcoming math test one day before.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

3. **Watching a teacher work an algebraic equation on the blackboard.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

4. **Taking an examination in a math course.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

5. **Being given a homework assignment of many difficult problems which is due the next class meeting.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

6. **Listening to a lecture in math class.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

7. **Listening to another student explain a math formula.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

8. **Being given a "pop" quiz in a math class.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

9. **Starting a new chapter in a math book.**

1	2	3	4	5
Not anxious	**Little anxious**	**Fairly anxious**	**Quite anxious**	**Very anxious**

Appendix F:

Spence Children's' Anxiety Scale, Generalised Anxiety Subscale (Spence 1998)

Please put a circle around the word that shows how often each of these things happen to you. There are no right or wrong answers.

1. I worry about things

Never	Sometimes	Often	Always
1	2	3	4

2. When I have a problem, I get a funny feeling in my stomach

Never	Sometimes	Often	Always
1	2	3	4

3. I feel afraid

Never	Sometimes	Often	Always
1	2	3	4

4. When I have a problem, my heart beats really fast

Never	Sometimes	Often	Always
1	2	3	4

5. I worry that something bad will happen to me

Never	Sometimes	Often	Always
1	2	3	4

6. When I have a problem, I feel shaky

Never	Sometimes	Often	Always
1	2	3	4

Appendix G: Ethical Approval

ERGO [DoNotReply@ERGO.soton.ac.uk]

To:
Walker E.

19 June 2012 19:40

Submission Number 2133:
Submission Title Understanding pathways to maths achievement in year nine pupils: An exploration of working memory and metacognitive skills:
The Research Governance Office has reviewed and approved your submission

You can begin your research unless you are still awaiting specific Health and Safety approval (e.g. for a Genetic or Biological Materials Risk Assessment) or external ethics review (e.g. NRES).The following comments have been made:

"There are no issues with this project - We will issue you with a letter confirming our approval shorty, but due to high demand there may be a little delay in sending this out. In the meantime, this email is sufficient to commence your study."

ERGO : Ethics and Research Governance Online
http://www.ergo.soton.ac.uk

DO NOT REPLY TO THIS EMAIL

Appendix H: Research Governance Office Approval

UNIVERSITY OF Southampton

Miss Emma Walker
School of Psychology
University of Southampton
University Road
Highfield
Southampton
SO17 1BJ

RGO Ref: 8613

25 June 2012

Dear Miss Walker

Project Title Understanding Pathways to Maths Achievement in Year Nine Pupils: An Exploration of Working Memory and Metacognitive Skills

This is to confirm the University of Southampton is prepared to act as Research Sponsor for this study, and the work detailed in the protocol/study outline will be covered by the University of Southampton insurance programme.

As the sponsor's representative for the University this office is tasked with:

1. Ensuring the researcher has obtained the necessary approvals for the study
2. Monitoring the conduct of the study
3. Registering and resolving any complaints arising from the study

As the researcher you are responsible for the conduct of the study and you are expected to:

1. Ensure the study is conducted as described in the protocol/study outline approved by this office
2. Advise this office of any change to the protocol, methodology, study documents, research team, participant numbers or start/end date of the study
3. Report to this office as soon as possible any concern, complaint or adverse event arising from the study

Failure to do any of the above may invalidate the insurance agreement and/or affect sponsorship of your study i.e. suspension or even withdrawal.

On receipt of this letter you may commence your research but please be aware other approvals may be required by the host organisation if your research takes place outside the University. It is your responsibility to check with the host organisation and obtain the appropriate approvals before recruitment is underway in that location.

May I take this opportunity to wish you every success for your research.

Yours sincerely

MPrude.

Dr Martina Prude
Head of Research Governance

Tel: 023 8059 5058
email: rgoinfo@soton.ac.uk

Corporate Services, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0) 23 8059 4684 Fax: +44 (0) 23 8059 5781 www.southampton.ac.uk

Appendix I: Authorisation from Questionnaire Authors

Rayne Sperling [rsd7@psu.edu]

To:
Walker E.

11 March 2012 17:11

Dear Emma,

I have used the Jr Mai as have others in several studies. One just came out in Journal of Educational Research within the last couple of months.
You may certainly use the Jr. Mai that is found in the CEP publication your referenced below. I only request that you properly cite the instrument and if you make any changes that you document and report them clearly in any write ups in your work.

I have most often used the scores on the measure as a continuous scale but others have used +/- one SD as low and high metacognition.

Good luck with your study!

Best,
Rayne

Hopko, Derek R [dhopko@utk.edu]

To:
Walker E.

12 May 2012 19:45

Hi Emma:
Yes, you have my full permission to use the AMAS. Good luck with your research!
Derek

Sent from my iPhone

Appendix J: Parent Information Letter

Dear Parent/Guardian,

I am a Trainee Educational Psychologist currently carrying out my Doctoral research at xxxx School. This project will assess the impact on maths achievement of one to one maths tutoring and a computerised working memory intervention, as research has found that both are linked to improvements in maths achievement. In order to better understand how to support learning in maths this study will also investigate if there are any effects on maths achievement of metacognition (thinking skills), working-memory, maths anxiety and generalised anxiety. As your child has been chosen to receive one to one tutoring, I am writing (with the permission of the Headteacher) to invite them to take part in this research.

Twenty students will be randomly chosen from the names of all pupils whose parents have consented for them to participate. They will be randomly allocated to either receive (in early September or mid-October) one to one tutoring with a teacher from xxxxx school, or a computerised working memory programme with me at school, for 35 minutes a day, 5 days a week for 5 weeks. Some of these sessions will take place during tutor time and will involve pupils being at school at the slightly earlier time of 8.20am in order to fit in all of the sessions. Pupils who complete the working memory training will still receive one to one tutoring, but this will be later in the school year.

Pupils will complete standardised assessments of maths achievement and working memory and questionnaires about metacognition, maths anxiety and general anxiety. These would all be carried out with me, at school, in July 2012 (before they start one to one tutoring), October/November 2012 and January 2013, to see if anything changes over time. Data collected will be kept confidential and analysed on a group basis (rather than closely examining data from individuals), in order to identify general trends.

I have designed the study in order to make it as enjoyable for the students as possible, so if you are happy for your child to take part, then please complete and return the slip below to XXXXXX, on or before XXXXXX. If your child decides that they would prefer not to participate in the study, or if they begin and decide that they do not wish to complete it, they may withdraw at any time and there will be no consequence for their treatment in school.

This project has received ethical approval from the School of Psychology, University of Southampton. If you have any queries about this research then please feel free to contact me by email: ew1g10@soton.ac.uk or if you have a complaint then please contact the Chair of the Ethics Committee Sarah Boak, on 023 8059 4663 or s.l.boak@soton.ac.uk.

Yours Faithfully

Emma Walker (Trainee Educational Psychologist, Southampton University)

I would like my child to take part in the research examining the effect of one to one tutoring and working memory training, on maths achievement.

Child's name (print name): ....

Name of parent (print name):..

Signature of parents:.......................................   Date:...

Appendix K: Participant Information Sheet sent to Parents and Pupils

Study Title: 'Understanding pathways to maths achievement in year nine pupils: An exploration of working memory and metacognitive skills'

Researcher: Emma Walker **Ethics number**: 2133

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.

What is the research about?
My name is Emma Walker. I am currently training at Southampton University to be an Educational Psychologist. As part of my training, I am working with your school to carry out a research project to investigate the impact of one to one maths tutoring and of a computerised working memory training intervention, on maths attainment. This research also looks at factors that previous research has linked to maths achievement to see if they affect outcomes of the two interventions (tutoring and working memory training). Specifically this study looks at metacognition (thinking skills), working memory (the part of your memory that stores and processes things you have just seen or heard), maths anxiety and general anxiety to see if they have an effect on how much maths skills improve. The aim of this research is gather information about how effective the tutoring and working memory training are and what factors affect this.

Why have I been chosen?
You have been asked to be a part of this study because you have been chosen by your school to have one to one maths tutoring, and your name has been randomly chosen from those of all the children receiving maths tutoring this year.

What will happen to me if I take part?
You will be randomly allocated to either receive one to one maths tutoring or to complete computerised working memory training. One to one tutoring will take place at school, starting either in September 2012 until mid-October 2012, or from mid October 2012 until late November 2012, and will be with a teacher, for one hour, twice a week, for five weeks. Computerised working memory training (a programme called CogMed) will take place at school, with me with five other pupils who are also in the study, starting either in September 2012 until mid-October 2012, or from mid October 2012 until late November 2012, for 35 minutes a day for at least 20 days over five weeks. The programme asks you to complete memory tasks on the computer that are designed to develop verbal (speaking skills) and visuo-spatial (things you can see and your perception of space and shape). As you are successful with each task, it moves you on to the next level.

As part of this study you will be asked in July 2012, then again at the end of your tutoring/working memory training (mid-October or late November 2012) and again in January 2013 to work with me to complete some measures. The reason for testing you at these different points is to see if anything changes over time. The measures will assess: maths attainment, where you have 15 minutes to answer maths questions; working memory, where you complete two tests: one where you judge if spoken sentences are 'true' or 'false' and then recall the last word of that sentence; and odd one out, where you pick the 'odd one out' of three shapes in a grid, and then at the end of the trial, identify where it was on a blank grid. There will also be three questionnaires which will measure: level of maths anxiety, by rating from one to five how anxious you feel in various maths related situations; general anxiety, where you rate how often you feel different types of anxiety (never, sometimes, often, always) and; metacognition, where you rate how often you use a particular metacognitive strategy.

Are there any benefits in my taking part?
Your data will help us to better understand what affects maths achievement and which interventions support this.

Are there any risks involved?
Participation in the research is unlikely to involve any risks, but should you feel uncomfortable at any time and no longer want to take part, then let me know and you can stop at any time and withdraw from the study.

Will my participation be confidential?
Any information that you give will be kept on a password protected computer. On completion of the study, the data will be stored in a secure location, on disk for ten years, before being destroyed, in compliance with the Data Protection Act 1998 and University of Southampton policy. Any information that you provide which might identify you will be removed from my report and only I will be able to identify you.

What happens if I change my mind?
You have the right to withdraw at any point in the study. If you do choose not to participate, there will be no resulting consequences.

What happens if something goes wrong?
If you have any concerns about your participation in this study, then please feel free to contact Sarah Boak on 023 80 598101 or s.l.boak@soton.ac.uk. to discuss these further.

Where can I get more information?
If you have any more questions about this research, then please feel free to contact me through the school office.

References

**References marked with an asterisk indicate intervention studies included in the systematic literature review*

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219-234. doi:10.1207/S15324818AME1403_2

Aiken, L. R. (1979). Attitudes toward mathematics and science in Iranian middle school. *School Science and Mathematics, 3*, 229-234.

Alloway, T. (2007). *Automated Working Memory Assessment.* London: Pearson Assessment.

*Alloway, T. (2012). Can interactive working memory training improve learning? *Journal of Interactive Learning Research, 23*(3), 197-207.

Alloway, T. P., & Alloway, R. G. (2008). *Jungle Memory Training Program.* [Computer software]. United Kingdom: Memosyne Ltd.

Alloway, T., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology, 28*, 725-734. doi.10.1080/01443410802243828

Alloway, T., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*(2), 606-621.

Alloway, T., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698-1716.

Apter, B. J. B. (2012). Do computerised memory training programmes designed to improve working memory work? *Educational Psychology in Practice, 28.*257-272. doi:10.1080/02667363.2012.712915

Ashcraft, M. H., & Kirk, E. P. (2001). The relationship among working memory, math anxiety and performance. *Journal of Experimental Psychology, 130*, 224-237. doi:10.1037//0096-3445.130.2.224

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance and math anxiety. *Psychonomic Bulletin and Review, 14*(2), 243-248.

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, *27*, 197-205. doi: 10.1177/0734282908330580

Autin, F., & Croizet, J. C., (2012). Improving working memory efficiency by reframing metacognitive interpretation of task difficulty. *Journal of Experimental Psychology: General*, *141,* 610-618. doi: 10.1037/a0027478

Ayotola, A., & Adedeji, T. (2009). The relationship between gender, age, mental ability, anxiety, mathematics self-efficacy and achievement in mathematics. *Cypriot Journal of Educational Sciences, 4,* 113-124.

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society, 4,* 523–526.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. (2002). Is working memory still working? *European Psychologist, 7,* 85–97. doi: 10.1027//1016-9040.7.2.85

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *The Psychology of learning and motivation: Advances in research and theory, Volume 8* (pp. 47-89). New York, USA: Academic Press.

Baddeley, A., & Hitch, G. J. (1994). Developments in the Concept of Working Memory. *Neuropsychology, 8*(4), 485-493.

Barkatsas, A., Kasimatis, K., Gialamas, V. (2009). Learning secondary mathematics with technology: Exploring the complex interrelationship between students' attitudes, engagement, gender and achievement. *Computers & Education, 52,* 562–570. doi:10.1016/j.compedu.2008.11.001

Bonnot, V., & Croizet, J. C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology 43,* 857–866. doi:10.1016/j.jesp.2006.10.006

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 367–406). New York, USA: Halsted Press.

Brown, P., Ireson, J., Shepherd, D. L., Bassett, P., & Rushforth, K. (2010). *One-to-One Tuition Pilot Course Evaluation Final Report.* (Report No. T35429) Retrieved from Department For Education website: https://www.gov.uk/government/publications/one-to-one-tuition-pilot-course-evaluation-final-report

Bull, R., Andrews-Espy, K., Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in pre-schoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 3,* 205-228. doi: 10.1080/87565640801982312

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematical ability: Inhibition, switching and working memory. *Developmental Neuropsychology, 19*(3), 273-293.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*, 1049-1053. doi: 10.1126/science.1201536

*Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and Teacher Education, 8*(2), 109-121.

*Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11*(1), 81-95.

Chinn, S. (2009). Mathematics anxiety in secondary students in England. *Dyslexia*, *15*, 61– 68. doi:10.1002/dys.381

Chui, M. M. (2004). Adapting teacher interventions to student needs during cooperative learning: How to improve student problem solving and time on-task. *American Educational Research Journal*, *41*(2), 365–399.

Cippolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, *124*(4),375-390.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cracco, J., Baudonck, M., Debusschere, A., Dewulf, B., Samyn, F., & Vercaemst, V. (1995). *Kortrijkse Rekentest[Kortrijk Arithmetic Test]*. Kortrijk, Belgium: Revalidatiecentrum Overleie.

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. CAYT Impact Study: REP03. Retrieved from: https://www.education.gov.uk/publications/standard/publicationDetail/Page1/CAYT-REP03-2013

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., Kaufmann, L., Benke, T., & Felber, S. (2005). Learning by strategies and learning by drill: Evidence from an fMRI study. *NeuroImage*, *25*, 838–849. doi: 10.1016/j.neuroimage.2004.12.009

Demetriou, A., Christou, C., Spanoudis, G., & Platsidou, M. (2002). Introduction: The information processing, the differential, and the developmental traditions in the structure of the development of the mind. *Monographs of the Society for Research in Child Development*, *67,* 1-38. doi: 10.1111/1540-5834.671174/pdf

Department for Children, Schools and Families (2009). *Developing one to one tuition: Guidance for tutors*. Retrieved from https://www.education.gov.uk/publications/eOrderingDownload/8067-DCSF-Developing%20One%20to%20One-TUTORS.pdf

Department For Education (2010). *Children with special educational needs 2010: An analysis.* (Research Report No. DFE-00553-2010). Retrieved from https://www.gov.uk/government/publications/children-with-special-educational-needs-2010-an-analysis

Department For Education (2011a). *How do pupils progress during Key Stages 2 and 3?* (Research Report No. DFE-RR096). Retrieved from https://www.gov.uk/government/publications/how-do-pupils-progress-during-key-stages-2-and-3

Department For Education (2011b). *School and college performance tables: Statement of intent – 2011*. Retrieved from http://www.education.gov.uk/schools/performance/archive/index.shtml

Department For Education (2012). *Support and aspiration: A new approach to special educational needs and disability. Progress and next steps.* (Research Report No. DFE-00046-2012). Retrieved from https://www.gov.uk/government/publications/support-and-aspiration-a-new-approach-to-special-educational-needs-and-disability-progress-and-next-steps

Department For Education (2013). *Indicative draft: The (0-25) special educational needs code of practice.* Retrieved from http://www.cambridgeshire.gov.uk/NR/rdonlyres/A46D5569-8506-4DC0-A38D-E4429C7DCBE0/0/sencodeofpracticeindicativedraftforcommittee.pdf

Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology, 5*(3), 705-730. Retrieved from http://www.investigacion-psicopedagogica.com/revista/articulos/13/english/Art_13_186.pdf

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and Mathematical Problem Solving in Grade 3. *Journal of Learning Disabilities,34,*435-447. doi:10.1177/002221940103400505

*Desoete, A., Roeyers, H., De Clercq. A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology, 95,* 188-200. doi: 10.1037/0022-0663.95.1.188

Desoete, A., Roeyers, H., & Huylebroeck, A. (2006). Metacognitive skills in Belgian third grade children (age 8 to 9) with and without mathematical learning disabilities. *Metacognition Learning, 1,* 119–135. doi 10.1007/s11409-006-8152-9

Dignath, C., & Büttner, G. (2008). Components of self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning, 3,* 231-264. doi: 10.1007/s11409-008-9029-x

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of healthcare interventions. *Journal of Epidemiology and Community Health, 52*, 377–384

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84,* 253-262.

Eshel, Y., & Kohavi, R. (2003). Perceived classroom control, self-regulated learning strategies, and academic achievement. *Educational Psychology: An International Journal of Experimental Educational Psychology, 23*, 249-260. doi:10.1080/0144341032000060093

Essau, C. A., Sasagawa, S., Anastassiou-Hadjicharalambous, X., Guzmán, B. O., & Ollendick, T. H. (2011). Psychometric properties of the Spence Child Anxiety Scale with adolescents from five European countries. *Journal of Anxiety Disorders*, *25*, 19-27. doi: 10.1016/j.janxdis.2010.07.001

Evans, C., Margison, F & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health, 1,* 70-72. doi:10.1136/ebmh.1.3.70

Every Child A Chance Trust (2007). *The long term costs of numeracy difficulties*. Retrieved from http://www.nationalnumeracy.org.uk/resources/14/index.html

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*(6), 409-434.

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*, 174-191. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

Fernandez-Duque, D., Baird, J. A.,& Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition,9*, 288–307. doi:10.1006/ccog.2000.0447

Field, A. (2009). *Discovering statistics using SPSS (Third edition).* London, England: Sage Publications Ltd.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new are of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.

*Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., Hosp, M., &

Janceck, D. (2003). Explicitly teaching for transfer: Effects on third grade students' mathematical problem solving. *Journal of Educational Psychology, 95,* 293-305. doi: 10.1037/0022-0663.95.2.293

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59-65.

Gathercole, S. E, & Alloway, T. (2008). *Working memory and learning: A practical guide for teachers.* London: Sage.

Gathercole, S. E. , Dunning. D. L., & Holmes. J. (2012). Cogmed training: Let's be realistic about intervention research. *Journal of Applied Research in Memory and Cognition*, *1,* 201–203. doi:10.1016/j.jarmac.2012.07.007

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory ability and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology, 22,* 1019-1037. doi: 10.1002/acp.1407

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18,* 1-16. doi: 10.1002/acp.934

Geary, D. C., Hoard, M. K., Bryd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*(4), 1343-1359.

Goldacre, B. (2013). *Building evidence into education.* Retrieved from Department For Education website: https://www.gov.uk/government/publications/department-for-education-analytical-review

*Gray, S. A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., Hockenberry,M., & Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with sever LD and comorbid ADHD: A randomized controlled trial. *The Journal of Child Psychology and Psychiatry, 53*, 1277-1284. doi:10.1111/j.1469-7610.2012.02592.x

Grouws, D. A., & Cebulla, K. J. (2000). *Improving student achievement in mathematics*. The International Academy of Education. Retrieved from http://www.ibe.unesco.org/fileadmin/user_upload/archive/publications/EducationalPracticesSeriesPdf/prac04e.pdf

Hennessey, G. M. (1999). Probing the Dimensions of Metacognition: Implications for Conceptual Change Teaching-Learning. Retrieved from http://www.eric.ed.gov/PDFS/ED446921.pdf

Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory, 9,* 233-247.

Henry, L. A., & MacLean, M. (2003). Relationships between working memory, expressive vocabulary and arithmetical reasoning in children with and without intellectual disabilities. *Educational and Child Psychology, 20*(3), 51-62. doi: 10.1002/acp.934

Higgins, J. P. T., & Green., S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions.* Retreived from http://handbook.cochrane.org/chapter_8/8_4_introduction_to_sources_of_bias_in_clinical_trials.htm

Higgins, S., Katsipataki, M., Kokotsaki, D.,Coe, R., Elliot-Major, L. & Coleman, R. (2013). The sutton trust-education endowment foundation teaching and learning toolkit: Technical appendices. Retrieved from the Sutton Trust website http://educationendowmentfoundation.org.uk/uploads/pdf/Technical_Appendices_(June_2013).pdf

*Hoek, D, van den Eeden, P., & Terwel, J. (1999). The effects of integrated social and cognitive strategy instruction on the mathematics acheivement in secondary education. *Learning and Instruction, 9*(5), 427-448.

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology: An International Journal of Experimental Educational Psychology, 26*, 339-366. doi:10.1080/01443410500341056

*Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12,* 9-15. doi: 10.1111/j.1467-7687.2009.00848.x

Hopko, D. R. (2003). Confirmatory factor analysis of the maths anxiety rating scale-revised. *Educational and Psychological Measurement, 63,* 336-351. doi: 10.1177/0013164402251041

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

Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung (Handanweisung) [Testing system for school and developmental growth].* Göttingen, Germany: Verlag für Psychologie Hochrefe.

Humphrey, N., & Squires, G. (2011). *Achievement for all national evaluation: Final report.* (Research Report No. DFE-RR176). Retrieved from https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-RR176

Ischebeck, A., Zamarian, L., Schocke, M., & Delazer. M. (2009). Flexible transfer of knowledge in mental arithmetic: An fMRI study. *NeuroImage*, *44*, 1103–1112. doi:10.1016/j.neuroimage.2008.10.025

*Jacobse, A. E., Harskamp, E. G. (2009). Student-controlled metacognitive training for solving word problems in primary school mathematics. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *15*(5), 447-463. Doi:10.1080/13803610903444519

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12-19.

*Kajamies, A., Vauras, M., & Kinnunen, R. (2010). Instructing low-achievers in mathematical word problem solving. *Scandinavian Journal of Educational Research*, *54*, 335-355. doi: 10.1080/00313831.2010.493341

*Kapa, E. (2007). Transfer from structured to open-ended problem solving in a computerized metacognitive environment. *Learning and Instruction*, *17,* 688-707. doi: 10.1016/j.learninstruc.2007.09.019

Kaufmann, L., & Nuerk, H. C. (2005). Numerical development: Current issues and future perspectives. *Psychology Science*, *47*(1), 142-170.

Kellogg, J. S., Hopko, D. R., Ashcraft, M. H. (1999). The effects of time pressure on arithmetic performance. *Journal of Anxiety Disorders*, *13*(6), 591-600.

Kirsh, D. (2005). Metacognition, distributed cognition and visual design. In P. Gärdenfors, & P. Johansson (Eds.), Cognition, education, and communication technology (pp. 1–22). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kleitman, S., Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences 17*, 161–173. doi:10.1016/j.lindif.2007.03.004

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14,* 317-324. doi: 10.1016/j.tics.2010.05.002

Klingberg, T., Fernell, E., Oleson, P. J., & Johnson, M. (2005). Computerized training of working memory in children with ADHD – A randomized, controlled trial. Journal of the American Academy of Adolescent Psychiatry, 44, 177-186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience 14*(1), 1–10.

*Kramarski, B., Mevarech, Z. R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. Educational Studies in Mathematics, 49, 225-250.

*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,* 281-310. doi:10.3102/00028312040001281

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

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9,* 178-181. doi:10.1111/1467-8721.00088

Kyttälä, M., Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, *23*(1), 77-94.

Lai, E. R. (2011). *Metacognition: A literature review.* Retrieved from https://psychcorp.pearsonassessments.com/hai/images/tmrs/Metacognition_Literature_Review_Final.pdf

Leedale, R., Singleton, C., & Thomas, K. (2004). *Memory Booster*. [Computer program and manual]. Beverly, East Yorkshire: Lucid Research.

LeFevre, J., DeStefano, D., Coleman, B., Shanahan, T. (2005). Mathematical cognition and working memory. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp.361-377). Hove, England: Psychology Press Ltd.

Legg, A. M., & Locker, L. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology, 11*(3), 471-486.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612-625.

Lindeman, J. (1998). *In: ALLU: Ala-asteen Lukutesti [Primary School Reading Test]*. Turku, Finland: University of Turku, Center for Learning Research.

Luo, W., Paris, S. G., Hogan, D., & Luo, Z. (2011). Do performance goals promote learning? A pattern analysis of Singapore student's achievement goals. *Contemporary Educational Psychology, 36,* 165-176. doi: 10.1016/j.cedpsych.2011.02.003

Maloney, E. A., Risko, E. F., Ansari, D., Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition, 114*, 293-297. doi: 10.1016/j.cognition.2009.09.013

*Masqud, M. (1998). Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers. *Educational Research, 40,* 237-243. doi: 10.1080/0013188980400210

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*(1), 10-17.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49,* 270 –291. doi: 10.1037/a0028228

*Mevarech, Z. R. (1999). Effect of metacognitive training embedded in co-operative settings on mathematical problem solving. *Journal of Educational Research, 92,* 195-205. doi:10.1080/00220679909597597

*Mevarech, Z. R., & Amrany, C. (2008). Immediate and delayed effects of meta-cognitive instruction on regulation of cognition and mathematics achievement. *Metacognition Learning, 3,* 147-157. doi:10.1007/s11409-008-9023-3

*Mevarech, A. R., & Kramarski, B. (1997). IMPROVE: A multi-dimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2), 365-394.

*Mevarech, A. R., & Kramarski, B. (2003a). The effects of metacognitive training versus worked out examples on students' mathematical reasoning. *British Journal of Educational Psychology, 73,* 449-471.

*Montague, M., Applegate, B., & Marquard, K. (1993). Cognitive strategy instruction and mathematical problem-solving performance of students with learning disabilities. *Learning Disabilities Research and Practice,8*(4), 223-232.

Montague, M., & Bos, C. S. (1990). Cognitive and metacognitive characteristics of eighth-grade students' mathematical problem solving. *Learning and Individual Differences, 2*,371–388

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin and Review, 18,* 46-60. doi: 10.3758/s13423-010-0034-0

NFER Nelson, (1992). *Group Reading Test II.* London, UK: NFER Nelson.

NFER Nelson, (1999a). *Mental Mathematics 7.* London, UK: NFER Nelson.

NFER Nelson, (1999b). *Mental Mathematics 8.* London, UK: NFER Nelson.

Ofsted (2012). *Mathematics: Made to measure*. Research report no. 110159. Retrieved from http://www.ofsted.gov.uk/resources/mathematics-made-measure

Ofsted (2013). *School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education Act 2005 (as amended) from September 2012*. Research report no. 120101. Retrieved from http://www.ofsted.gov.uk/resources/school-inspection-handbook

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

*Packiam Alloway, T., Bibile, V., Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior, 29,* 632-638. doi: 10.1016/j.chb.2012.10.023

Panaoura, A. (2007). The interplay of processing efficiency and working memory with the development of metacognitive performance in mathematics. *The Montana Mathematics Enthusiast, 4*(1), 31-52.

Pelham, W.E., Milich, R., Murphy, D.A., & Murphy, H.A.(1989). Normative data on the IOWA Conners teacher rating scale. *Journal of Clinical Child Psychology, 18*, 259–262.

*Pennequin, V., Sorel, O. Nanty, I., & Fontaine, R. (2010). Metacognition and low achievement in mathematics: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Thinking and Reasoning, 16,* 198-220. doi: 10.1080/13546783.2010.509052

Pressley, M., Borkowski, J. G., & Schneiders, W. (1989). Good information processing: What it is and how education can promote it. *International Journal of Educational Research, 13*(8), 857-867.857-867.

Prevatt, F., Welles, T. L., Li, H., & Proctor, B. (2010). The contribution of memory and anxiety to the math performance of college students with learning disabilities. *Learning Disabilities Research and Practice, 25*(1), 39-47.

Räsänen, P. (2004). *RMAT: Laskutaidon testi 9–12 -vuotiaille [RMAT: A test for arithmetical skills for ages 9 to 12; in Finnish]*. Jyväskylä, Finland: Niilo Mäki Institute.

Raven, J. C. (1958). *Raven's Standard Progressive Matrices.* New York, USA: The Psychological Corporation.

Raven, J., Raven, J.C., & Court, J.H. (2000). *Standard progressive matrices.* Oxford, UK: Psychologist Press.

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review,6*, 135–147. doi:10.1016/j.edurev.2010.12.001

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551-554.

Roughan, L., & Hadwin, J. A. (2011). The impact of working memory training in young people with social, emotional and behavioural difficulties. *Learning and Individual Differences,21*, 759–764. doi:10.1016/j.lindif.2011.07.011

Rubinsten, O., & Tannock, R. (2010). Mathematics anxiety in children with developmental dyscalculia. *Behavioral and Brain Functions*. Retrieved from http://www.behavioralandbrainfunctions.com/content/6/1/46

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111–139. doi:10.1007/s11165-005-3917-8

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

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334- 370). New York, USA: MacMillan.

*Seo, Y. J., & Bryant, D. (2012). Multimedia CAI program for students with mathematics difficulties. *Remedial and Special Education, 33*, 217-225. doi: 10.1177/0741932510383322

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur. V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child & Adolescent Psychiatry*, *9*(2), 58-64.

Shalev, R. S., Manor, O., & Gross-Tsur. V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine & Child Neurology*, *47*, 121–125.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition*, *9*,313–323. doi:10.1006/ccog.2000.0450

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, *50*(3&4), 245-276.

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy, 36*(5), 545-566. doi: 10.1016/s0005-7967(98)00034-5

Spence, S. H., Barrett, P. M., & Turner, C. M. (2003). Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. *Journal of Anxiety Disorders, 17*, 605-625. doi: 10.1016/s0887-6185(02)00236-0

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,* 51-79. doi: 10.1006/ceps.2001.1091

Sperling, R. A., Richmond, A. S., Ramsay, C. M., & Klapp, M. (2012). The measurement and predictive ability of metacognition in middle school learners. *The Journal of Educational Research, 105,* 1-7. doi:10.1080/00220671.2010.514690

*St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology: An International Journal of Experimental Educational Psychology, 30*, 203-219. doi: 10.1080/01443410903509259

Stinissen, J., Smolders, M., & Coppens-Declerck, L. (1974). *Collective Verbale Intelligentie test voor derde en vierde leerjaar(CIT–34) [Collective Verbal Intelligence Test for Grades 3 and 4 (CIT–34)].* Brussels, Belgium: Centrum voor studie en beroepsorientering.

Stocco, A., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2012). Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks. *Cognitive, Affective and Behavioral Neuroscience, 12,* 611–628. doi: 10.3758/s13415-012-0117-7

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82,* 306-314.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., ... & McCleary, R. (2001).*Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN ratings scales*. Retrieved from http:// www.adhd.net/SNAP_SWAP.pdf.

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, Y., Yomogida, Y., Komuro, N.,Yamanouchi, T., Suzuki, S., & Kawashima, R. (2010) Training of working memory impacts structural connectivity. *The Journal of Neuroscience, 30,* 3297–3303. doi:10.1523/JNEUROSCI.4611-09.2010

*Teong, S. K. (2003). The effect of metacognitive training on mathematical word-problem solving. *Journal of Computer Assisted Learning, 19,* 46-55.

*Vauras, M., Kinninen, R., & Rauhanummi, T. (1999). The role of metacognition in the context of integrated strategy intervention. *European Journal of Psychology of Education, 14*(4), 555-569.

*Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. Journal of Intellectual Disability Research, 54, 433-447. doi: 10.1111/j.1365-2788.2010.01285.x

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,* 205-235. doi: 10.1080/02673843.2008.9748004

*Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. Mathematical Thinking and Learning, 1, 195-229. doi: 10.1207/s15327833mtl0103_2

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skilfulness: Availability versus production deficiencies. *Anxiety, Stress & Coping: An International Journal*, *13*, 391-412. doi.10.1080/10615800008248343

Wechsler, D. (1993).*Wechsler Objective Reading Dimensions*. London, UK: Pearson Assessment.

Wechsler, D. (1996).*Wechsler Objective Numerical Dimensions*. London, UK: Pearson Assessment.

Wechsler D. (1999). *Wechsler Abbreviated Scale of Intelligence*. London, UK: Harcourt Assessment.

Wilkinson, G., S., & Robertson, G. J. (2006). *Wide Ranging Achievement Test 4*. PAR: Florida, USA.

*Witt, M. (2011). School based working memory training: Preliminary finding of improvement in children's mathematical performance. *Advances in Cognitive Psychology, 7,* 7-15. doi: 10.2478/v10053-008-0083-3

Wolf, A. (2011). *Review of vocational education : The Wolf report*. (Research Report No. DFE-00031-2011). Retrieved from https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-00031-2011

Zimmerman, B. J. (2008). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman, & Schunk, D. H. (Eds.), *Self -regulated learning and academic achievements* (pp. 1-36). Retrieved from http://books.google.co.uk/books?hl=en&lr=&id=M3mZYcU-apkC&oi=fnd&pg=PA1&dq=systematic+learning+and+maths+achievement&ots=6OSnfIkaR-&sig=es8Gv_kdlweVHuYgJbjIRw--gI8#v=onepage&q&f=false