



The Relationship between Visuospatial Working Memory and Mathematical Performance in School-Aged Children: a Systematic Review

Katie Allen¹  · Steve Higgins¹ · John Adams²

Published online: 8 February 2019
© The Author(s) 2019

Abstract

The body of research surrounding the relationship between visuospatial working memory (VSWM) and mathematics performance remains in its infancy. However, it is an area generating increasing interest as the performance of school leavers comes under constant scrutiny. In order to develop a coherent understanding of the literature to date, all available literature reporting on the relationship between VSWM and mathematics performance was included in a systematic, thematic analysis of effect sizes. Results show a significant influence of the use of a standardised mathematics measure, however, no influence of the type of VSWM or mathematics being assessed, on the effect sizes generated. Crucially, the overall effect size is positive, demonstrating a positive association between VSWM and mathematics performance. The greatest implications of the review are on researchers investigating the relationship between VSWM and mathematics performance. The review also highlights as yet under-researched areas with scope for future research.

Keywords Visuospatial · Working memory · Mathematics performance · Numerical operations · Mathematical reasoning

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10648-019-09470-8>) contains supplementary material, which is available to authorized users.

✉ Katie Allen
katie.allen@durham.ac.uk

¹ School of Education, University of Durham, Durham, UK

² Department of Psychology, University of Durham, Durham, UK

Introduction

Development of Children's Mathematical Skills

Informal mathematical development begins much before children reach the age of formal education with the development of number sense. Inherent in this is the existence of a mental number line (Berch 2005; Schneider et al. 2009). As a precursor to the development of this mental number line, research postulates an innate sense of number by which humans are able to distinguish between sets to judge which has more (the approximate number system; Dehaene 2001). Additionally, young children demonstrate the ability to perceptually determine the exact number of items in small sets (Clements 1999), an ability known as subitizing (Ginsburg 1978; Benoit et al. 2004). This innate sense of number is a skill that evolutionary psychologists attribute survival to, for example where one can find more food (De Cruz 2006). Many habituation studies (e.g. Xu and Spelke 2000; Starkey et al. 1990) have provided evidence for number sense in young infants, demonstrating a renewed interest upon alteration of the number of items in the presented array, as long as a critical ratio criterion is met (according to Weber's Law; Feigenson et al. 2004).

Once children become verbal, they learn a counting list which functions in the form of a 'placeholder structure' (Sarnecka and Wright 2013), carrying little numerical context. This suggests that children develop a knowledge of a specific set of number words, in a fixed order, before their knowledge develops into a deeper understanding of number as an abstract principle (Sarnecka and Gelman 2004). A further milestone in the development of number sense occurs when young children are taught to attribute specific quantities to Arabic numerals (Krajewski and Schneider 2009). Wynn (1990) previously described specificity as the knowledge that every number word describes a specific numerosity. Importantly, the attribution of specific quantities to individual numerals paves the way for children establishing understanding of a set of rules: cardinality (the final numeral used represents the total number in the set), abstraction (sets of any nature can be counted, including entirely mental constructs), one-to-one correspondence (each item in a set should be counted once and only once), stable order (numerals should be used in a fixed order) and order irrelevance (items in a set can be counted in any order without changing the cardinality of the set; Gelman and Gallistel (1986); Dehaene (1992); Thompson (2010)). Upon reaching this stage, children are deemed to have developed a 'mental number line', which, over time, becomes increasingly linear after initially following a somewhat logarithmic structure (Siegler and Booth 2004; Dehaene 2003), whereby numbers outside of the child's counting range may be viewed only as 'big' or 'lots'. From this foundation, children can begin to understand the formal manipulations of numbers required to gain proficiency in mathematics through formal instruction, as identified by Libertus et al. (2011) who demonstrated that ANS acuity in infants predicts early maths achievement.

The development of mathematical skills, upon the commencement of formal schooling, can be considered to pertain to two broad stylistic categories, as adopted by Weschler assessments: Numerical Operations and Mathematical Reasoning. Whilst the National Curriculum has four areas (number, measurement, geometry and statistics), it is these categories that will be considered in this review as they succinctly describe the fundamental understanding of mathematics (Numerical Operations) and its application (Mathematical Reasoning). Numerical Operations concerns procedures that may best be described as numeracy, involving number knowledge, basic numerical manipulations and mental arithmetic (Geary et al. 2007). Tests of Numerical Operations typically comprise of explicit mathematical equations with basic

operations for children to solve using a written format, as well as assessments of counting, identifying numbers and written calculations (Pearson Clinical; Wechsler 2017). By contrast, Mathematical Reasoning is defined by Thompson (1996) as the ability to carry out ‘purposeful inference, deduction, induction, and association in the areas of quantity and structure’. Such a definition aligns well with the nature of the tasks used to assess the construct, which comprise mainly of single- and multi-step contextual story problems that the children are required to solve using the information provided. Examples of such problems are those involving whole numbers, fractions and decimals, graphs and probability (Wechsler 2017).

A broad range of assessments are employed in both research and educational settings when establishing a child’s understanding of mathematics. Such assessments range from simple, individually derived series of calculations and equations to subtests of standardised test batteries. As a result of this wide-ranging variety, it is imperative to note whether the assessment in question provides a standardised score, or should only be considered in an isolated manner. One should take care to consider the structure and content of the assessment used in relation to the research question in order to determine its suitability regarding content and intended statistical analysis. This is particularly important when critiquing studies utilising non-standardised measures of mathematics over those taken from standardised batteries.

In summary, mathematical development begins before, and continues throughout, formal schooling. However, careful attention should be paid to the measures used to assess mathematics for research and educational purposes as their structure and content may influence the conclusions that can be drawn.

Theory of VSWM

Baddeley and Hitch (1974) first developed the concept of the visuospatial sketchpad as one of two slave systems in working memory (WM), outlining its responsibility for storing and manipulating visual and spatial information. Researchers in the field of WM have long since adopted the most recent revision of this model (Baddeley 2000) as it has been demonstrated to accurately conceptualise findings (e.g. Holmes and Adams 2006; Ashkenazi et al. 2013; Andersson and Lyxell 2007) and to be robust to developments in understanding resulting from neuropsychological and dual task studies (e.g. Logie 2014; Henson 2002). As such, this model still holds as an appropriate explanation of WM and is the model adopted by the studies included in this review. Currently, a focus on the emergence of simultaneous and sequential visuospatial working memory (VSWM; see Mammarella et al. 2006 and Mammarella et al. 2013 for evidence of a double dissociation) is evident, in a move to understand the finer nuances of using VSWM as an academic predictor.

Simultaneous VSWM tasks are defined as such tasks whereby all information is presented to the participant at the same time (Mammarella et al. 2006). Following this presentation, the participant is asked to recall the positions of the stimuli they saw previously; an example of this type of task is the visual patterns task. In contrast, sequential tasks involve the presentation of stimuli in a sequence to the participant (as in Passolunghi and Mammarella 2011). Participants are then required to recall the positions of the stimuli, typically in the correct order, as in the Corsi block task (Mammarella et al. 2006). There is evidence for the dissociation of these tasks (Mammarella et al. 2008), supporting the need for their independent investigation in order to assess their predictive power.

In line with these observations, a number of different VSWM tasks are used to tap into each of these components. As elements of standardised test batteries, a small number of VSWM

tasks are standardised; however, a large proportion of the tasks used are designed for the purpose of the study in question. As such, it is imperative to assess the characteristics of the test in relation to the research question and statistical procedures applied before accepting the conclusions drawn from the results. This is of particular importance when studies employ a non-standardised VSWM measure.

Relationship between VSWM and Mathematics

Importantly, VSWM is described by Ashkenazi et al. (2013) as a ‘source of domain general vulnerability in arithmetic cognition’, indicating its position as one of a number of mechanisms in the brain which function to support learning in a broad range of areas. Such a definition also follows that knowledge is cumulative and so builds up over time to form our overall knowledge structure. As evidenced by the results of previous studies, age appears to be crucial to the extent of the involvement of VSWM in mathematics performance (Li and Geary 2013), with the suggestion of a cyclical pattern of involvement between VSWM and verbal WM. One could reasonably question the potential for an emerging relationship between novelty and mastery inherent in a cyclical relationship. VSWM is more strongly predictive of mathematics performance in younger children (Holmes et al. 2008; Holmes and Adams 2006) which is, arguably, the period in which children are acquiring new mathematical skills at an increased rate. Therefore, it is possible that VSWM is employed to a greater extent during the procurement of new skills, and to a lesser extent once children achieve mastery of such skills (Andersson 2008).

It may be possible to identify the age at which young children’s mathematics ability is most strongly influenced by VSWM, and hence use this information to make predictions regarding future attainment. Research is currently moving to exploit this relationship further in order to train WM to improve academic attainment (e.g. Holmes and Gathercole 2014; see Sala and Gobet 2017 for a review); however, this will only be possible when the intricacies of the relationship between the two factors are fully understood. Similarly, the potential to mediate vulnerability to mathematical difficulties as a result of poor WM before they occur is hindered by a lack of detailed knowledge in this area. Before research in this area can progress, a clear representation of what is currently known in the literature is necessary. This review aims to provide this comprehensive picture.

In doing so, it is necessary to ensure that confounding factors are limited as far as possible. Often, studies employ tasks previously designed either to investigate a particular aspect of VSWM or mathematics, or those which form a component of a standardised battery. When appraising potential measures for a study, the age group for which the task was designed and, potentially standardised, is crucial. Only by considering the target age and that of the participants is it possible to make reasonable adjustments to prevent floor and ceiling effects. This is of particular importance when considering appropriate mathematics tasks as it is imperative that tasks administered align with concepts children have been exposed to through the curriculum. More leniency can be afforded to VSWM tasks as such tasks present fewer barriers to achievement should a child not have completed a similar task before. Further, given the nature of the research seeking to extend scientific understanding of the components of VSWM, novel tasks are required to access each component individually.

In summary, using VSWM as a means to predict pupil’s future attainment in mathematics is a topic that has gained a significant amount of traction in recent years. Driven by the desire to improve academic performance, it is necessary to first ensure a clear understanding of the

relationship between the two components before steps can be taken to use VSWM as a predictive tool.

Importance of this Review

Given the relative infancy of this field of research, no other reviews concerning the relationship between VSWM and mathematics attainment have been identified. Szűcs (2016) completed a review on a similar field, identifying the relationships between subtypes of mathematical difficulties and elements of working and short-term memory. The available literature demonstrates both comparable and contrasting results which can only be adequately understood by appraising the results of the studies alongside their methodologies. In doing so, it is possible to begin to explain the variations in results as features of the methodological differences. To this end, this review is necessary to consolidate the findings of previous research in order to provide a comprehensive understanding of the relationship between VSWM and mathematical performance. The results have a number of implications with regard to using VSWM as a predictive tool for future mathematical attainment, something which cannot be achieved without a streamlined understanding of the relationship central to forming these predictions, including, but not limited to, early intervention to improve attainment.

Objectives of the Review

The aim of this review is to examine the literature surrounding the relationship between VSWM and mathematical attainment in children. Four key issues will be addressed; these are the influences of the age of the participants, the type of mathematics being assessed, the type of VSWM being assessed and the nature of the tasks used (standardised/non-standardised). It is broadly understood that VSWM plays both an influential and predictive role in children's mathematical performance (Holmes and Adams 2006; Bull et al. 2008); however, the exact relationship between these elements remains, as yet, unclear. The existing literature alludes to a number of factors that are influential in establishing a clear and coherent understanding of the role VSWM plays in mathematical development. This review will explore these potential confounds in a move to consolidate the existing knowledge on this issue. Focusing on the age of the participants, the components of mathematics being assessed and the components of VSWM being measured, it is possible to begin to develop a more detailed understanding of the specific influences of each of these elements.

Method

Criteria for Study Inclusion

Studies eligible for inclusion in the analysis met all of the criteria outlined below.

Study Design

Studies utilising all methodological designs were included in the review due to the nature of both the current literature and the review. Before inclusion, researchers must, however, have

explicitly stated their intention to investigate the relationship between VSWM and mathematical attainment. Despite using studies with any design, before a study was included in the review, sufficient control and operationalisation of the variables must have been established. Testing should have been conducted in a controlled environment, with an emphasis on maintaining consistency between sessions in order to exert control in the absence of randomised control trials.

Type of Participants

Studies of children attending mainstream schools, between the ages of 0 and 16 years, were considered in this review. Three exclusion criteria applied: those investigating atypical populations, adults and young people over the age of 16 years old and preterm children specifically. All ethnicities, socio-economic statuses and genders were included.

Mathematics Measures

The review included mathematics measures assessing elements of mathematics relating to numerical operations, mathematical problem solving and/or mathematics as a whole; those utilising measures of number sense, numerosity and other such related components were excluded. Whilst the majority of studies included in the review used standardised measures of mathematics, including the WIAT and WOND, a proportion used specifically designed measures. Studies of this nature were included so long as an observable, clear focus on one or more of the aforementioned components of mathematics was present. Where mathematics measures had been derived for the purpose of the study, this was typically in line with the curriculum outlined for children of the specified age in the given country.

Memory Measures

Only studies published reporting VSWM as an explicit individual concept met the criteria for the review. Those reporting on WM as a whole only, without further subdivision, were not included in the final sample. A number of standardised VSWM measures were employed; however, as a result of attempts to further subdivide VSWM, many measures were designed for the purposes of the study. As such, all measures specifically of VSWM were accepted.

Location of Study

Studies may have been conducted in any country utilising an alphabetic language system to be eligible for inclusion; however, the final paper must be available in English. Only nations with alphabetic language systems were included due to the potential influence of logographic writing systems on the development of VSWM (Tan et al. 2001).

Additional Criteria

Criteria were identified which led to the exclusion of a study. These exclusion criteria were studies concerning:

- Neuroimaging
- Mathematics anxiety
- Number sense/numerosity
- Visual perception
- Working memory training
- Strategy use
- Interventions/teaching methods
- Transcoding

Additional criteria for exclusion were texts from book chapters (serving only to summarise findings from included empirical studies) or other review articles.

Search Methods for Study Identification (Search Strategy)

Electronic Searches

Searches were conducted of the databases listed below (using the ‘all databases’ option for each), with search terms defined as ‘visuospatial’, ‘working memory’ and ‘math*’. Only articles where the full text was available were included. These terms were defined so as to identify all available studies that use these terms either in the title, abstract or main body. Given the specificity of the desired work, simple, clearly defined search criteria were most appropriate:

- Web of Science
- JSTOR
- Science Direct
- Medline/ NCBI
- Scopus
- FirstSearch
- EBSCOhost

Search of Other Sources

Reference lists of the included papers were scrutinised to identify any further appropriate papers.

Data Collection and Analysis

Determining Eligibility and Data Extraction

All data was extracted by the same author. Before any coding began, a stringent set of inclusion and exclusion criteria were clearly defined and periodic checks throughout data extraction were carried out to ensure criteria were adhered to at all times. Should a study be found to be ineligible upon full reading, the reasons for its exclusion were documented. Before beginning synthesis of results, the main statistic for each study was extracted and recorded.

Study Coding Categories

Any study that met the criteria for inclusion based on title, abstract and full text reading was coded to extract the same information. This information included details regarding methodology, measures taken, participant details, the area of VSWM and mathematics being assessed, statistical method used and the main reported statistic, and a quality judgement of the study fit for the review (1 = very good fit; 2 = good fit; 3 = not very good fit, e.g. Vanbinst et al. (2018) = 1, very good fit, Caviola et al. (2012) = 3, not very good fit). Once this information was compiled for each study, where not already given by the paper, an effect size was calculated and a quality judgement of the effect size calculation noted (1 = exact calculation, 2 = good approximation, 3 = rough approximation, e.g. Campos et al. (2013) = 1, exact calculation, Pina et al. (2014) = 2, good approximation).

Determining Effect Sizes

Common effect sizes were calculated (r) for each paper so as to allow for direct comparison between studies. R was chosen as an appropriate effect size due to the assessment of overlap between the variables, rather than the difference between experimental groups. Where this was reported in the paper, this is the effect size reported; however, in other cases, this was calculated using an accessible effect size calculator from the Campbell Collaboration (Wilson n.d.), alongside a second freely available calculator from Psychometrica (Lenhard and Lenhard, 2016). These calculators allow calculation of effect sizes from a comprehensive range of study designs and so provide the most appropriate calculations of any given effect size. The use of two independent calculators allowed calculation of effect sizes from a greater variety of study design, where one calculator provided a means to convert a statistic in its absence from the other, as well as corroboration of calculations by using both calculators.

Dealing with Missing Data

In cases where sufficient data were not available to calculate effect sizes, where possible this information was calculated from other available data, for example the use of reported correlations from studies using multi-level models. As such, it was possible to calculate all of the required effect sizes, though the basis of such calculations on good approximations of the exact data was recorded in the quality judgements of the effect size calculations made for each study.

Assessment of Heterogeneity

Due to the varied nature of the research available, it was unlikely that a meta-analysis would be possible. Studies included in the review demonstrated important differences between crucial aspects of their design, such as measures, methods used and participants included. As a result, a thematic analysis using inferential statistics was concluded to be the most appropriate method for synthesis so as not to introduce error through drawing comparisons between dissimilar studies. An I^2 statistic of 89.81%, much higher than the recommended maximum of 25% when undertaking a meta-analysis, supports not completing a meta-analysis on the current data. Following findings by Von Hippel (2015), it is important to consider the I^2 statistic in relation

to the number of studies included in the meta-analysis; however, 35 studies should be sufficient to mitigate the potential for bias with a small number of studies.

Data Synthesis

Due to the large variation in the studies included, and a number of confounding factors, including sample size and the use of unstandardised measures, a quantitative synthesis was completed using inferential statistics. As such, thematic analysis of the components of interest was completed, addressing issues of participant age, type of mathematics being assessed and component of VSWM being assessed, sample size and the use of standardised measures.

Detecting and Adjusting for Publication Bias

Most of the studies in this review concentrate on correlational relationships between the measured variables. As such, it is not unreasonable to suggest that publication bias may affect publication of these studies to a lesser extent as there is less of a drive to demonstrate a particular outcome. One must remain vigilant, however, as it remains the case that negative or more difficult to interpret results will be less easy to publish. In order to reduce publication bias introduced to this review, databases that include work such as theses and dissertations were also included when the literature search was conducted (see below for funnel plot). The non-significant Egger's regression ($p=0.21$), alongside the randomly distributed funnel plot, suggests there is no evidence of publication bias.

Results

Description of the Studies

Results of the Search

The search of the above-listed databases returned 590 records (search terms: 'visuospatial' 'working memory' and 'math*'). Along with the electronic database searches, an additional 34 records were found as a result of the manual searches of reference lists completed.

Fifty-two of the records identified throughout the entire search process were duplicates and so were removed; 538 remained after this stage. Following screening of the titles and abstracts for irrelevant records, 469 records were excluded in accordance with the exclusion criteria, leaving 69 records. The remaining articles were read in full, and the relevant data from 35 articles deemed appropriate, according to the inclusion and exclusion criteria, was extracted for analysis in the current review. Data was extracted from 35 articles in total (Tables 1 and 2).

Description of Included Studies

The studies included were conducted in a number of countries, and as such allow for a clearer understanding of the relationships between VSWM and mathematics performance globally, as opposed to solely in relation to the National Curriculum followed in the UK. Further, the broad age range of participants allows for an understanding to be established regarding the potential

Table 1 Number of study participants, age of participants, mathematics measures used and VSWM measures used for each study included in the analysis

Author(s)	Date	Number of participants	Age (in years)	Mathematics measures	VSWM measures
Vanbinst et al.	2018	51	5–8	Standardised addition and subtraction task	Corsi block task
Vandenbroucke et al.	2018	107	6–7	Standardised achievement test	Dot matrix, block recall, odd one out, Mr X
Bresgi, Alexander and Seabi	2017	80	7–8	Group mathematics test	Spatial recall, spatial processing recall
Li and Geary	2017	145	12–15	Weschler individual achievement test (WIAT)	Block recall, mazes memory
Mammarella et al.	2017	72	9–10	AC-MT 11–14 standardised arithmetic battery, AC-FL, BDE-2 battery	Visual memory houses/balloons, spatial-simultaneous and spatial-sequential matrices
Mix et al.	2016	854	6–11	Place value, word problems, calculation, missing term problems/algebra, number line estimation, fractions	Adaptation of dot matrix
Wiklund-Hörnqvist et al.	2016	597	9	Swedish national test in mathematics for grade 3 pupils	Adaptation of WISC-IV block span
Soltanlou, Pixner and Nuerk	2015	77	8–11	Multiplications	Corsi block task forwards/backwards
Van de Weijer-Bergsma et al.	2015	4337	5–10	Arithmetic tempo test	Lion game
Martin et al.	2014	193	6–7	Procedural counting, conceptual counting, symbolic number identification (K), small sums addition and subtraction, WJ-3 calculation, WRAT-3 arithmetic, WJ-3 applied problems subtest, single-digit story problems (1st grade)	Adaptation of dot matrix
Nath and Szucs	2014	66	7	Numerical operations subtest (WIAT)	Dot matrix, odd one out
Pina et al.	2014	102	9–10	Fluency and quantitative concepts tests from Spanish version of WJ-III ACH, arithmetic test from Spanish WISC	Computerised corsi block task forwards and backwards
Ashkenazi et al.	2013	34	7–9	Numerical operations and mathematical reasoning subtests (WIAT)	Block recall
Li and Geary	2013	177	6–11	WIAT	Block recall, mazes memory
Szucs et al.	2013	24	9	Mathematical assessment for learning test, numerical operations subtest (WIAT)	Dot matrix, odd one out
Campos et al.	2013	103	8–9	Arithmetic word problems, measurements	Block recall, mazes memory
Caviola et al.	2012	263	8–9	Standardised arithmetic battery	Dot matrix

Table 1 (continued)

Author(s)	Date	Number of participants	Age (in years)	Mathematics measures	VSWM measures
Maennamaa et al.	2012	723	8–9	Maths test designed in line with the third grade Estonian curriculum	Figure recognition test
Alloway and Passolunghi	2011	206	7–8	AC-MT, WOND	Dot matrix, mazes memory, block recall, odd one out, Mr X, spatial recall
Geary	2011	177	7–10	Numerical operations subtest (WIAT)	Block recall, mazes memory
Meyer et al.	2010	98	7–8	Numerical operations and mathematical reasoning subtests (WIAT)	Block recall
Passolunghi and Mammarella	2010	59	9	12-item standardised mathematics test, WRAT calculation subscale	Corsi block task, spatial matrix, houses recognition task, pathway span
Alloway et al.	2009	308	5–9	Weschler objective numerical dimensions (WOND)	Odd one out, Mr X, spatial recall, dot matrix, mazes memory, block recall
De Smedt et al.	2009	106	6–7	Flemish Student Monitoring System	Block recall, visual pattern task
Andersson	2008	141	9–11	Horizontally presented addition, subtraction and multiplication problems; arithmetic fact retrieval	Visual matrix, corsi block task
Bull, Espy and Wiebe	2008	104	4–7	Performance indicators in primary school (PIPS)	Corsi block task forwards and backwards
Holmes, Adams and Hamilton	2008	107	7–10	Maths tests designed to test the 4 elements of the national curriculum	Visual patterns test, block recall
Kyttaelae and Lehto	2008	128	15–16	The Mathematics Test (Finland)	Visual patterns test, corsi block task, mental rotation
Andersson and Lyxell	2007	138	9–10	Simple addition	Dot matrix, Corsi block task
Holmes and Adams	2006	148	7–10	Maths tests designed to test the 4 elements of the national curriculum	Mazes memory
Bayliss et al.	2005	56	7–9	NFER-Nelson mathematics	Target search, adaptation of dot matrix
D'Amico and Guamera	2005	28	9–11	ABCA	Matrix task, corsi block task
Jarvis and Gathercole	2003	128	10–14	National curriculum composite results	Visual patterns test, dot matrix, spatial span task, odd one out task
Maybery and Do	2003	49	9–10	Wood and Lowther Easymark Diagnostic Mathematics Test	Fixed spatial span, running spatial span
Reuhkala	2001	115	15–16	National mathematics test	Matrix pattern task, corsi block task

Table 2 Effect size (r), confidence interval for effect size, type of mathematics and type of VSWM for each study included in the analysis

Author(s)	Date	r	Lower CI	Upper CI	Type of mathematics	Type of VSWM
Vanbinst et al.	2018	0.527	0.294	0.701	Numerical operations	Sequential
Vandenbroucke et al.	2018	0.600	0.463	0.709	Numerical operations and mathematical reasoning	Sequential
Bresgi, Alexander and Seabi	2017	0.446	0.251	0.606	Numerical operations	Sequential
Li and Geary	2017	0.428	0.285	0.553	Numerical operations	Simultaneous and sequential
Mammarella et al.	2017	0.323	0.099	0.516	Numerical operations	Simultaneous and sequential
Mix et al.	2016	0.137	0.071	0.202	Numerical operations and mathematical reasoning	Simultaneous
Wiklund-Hömqvist et al.	2016	0.230	0.153	0.305	Numerical operations and mathematical reasoning	Sequential
Soltanlou, Pixner and Nuerk	2015	0.268	0.047	0.464	Numerical operations	Sequential
Van de Weijer-Bergsma et al.	2015	0.520	0.498	0.541	Numerical operations	Sequential
Martin et al.	2014	0.163	0.022	0.297	Numerical operations	Sequential
Nath and Szucs	2014	0.357	0.126	0.551	Numerical operations	Sequential
Pina et al.	2014	0.231	0.038	0.407	Numerical operations	Sequential
Ashkenazi et al.	2013	0.469	0.156	0.697	Numerical operations and mathematical reasoning	Sequential
Li and Geary	2013	0.380	0.246	0.500	Numerical operations	Simultaneous and sequential
Szucs et al.	2013	0.610	0.274	0.813	Numerical operations and mathematical reasoning	Simultaneous and sequential
Campos et al.	2013	0.586	0.443	0.700	Mathematical reasoning	Sequential and simultaneous
Caviola et al.	2012	0.425	0.321	0.520	Numerical operations	Simultaneous
Maennamaa et al.	2012	0.280	0.211	0.346	Numerical operations and mathematical reasoning	Simultaneous
Alloway and Passolunghi	2011	0.520	0.413	0.614	Numerical operations	Simultaneous and sequential
Geary	2011	0.253	0.109	0.386	Numerical operations	Simultaneous and sequential
Meyer et al.	2010	0.365	0.180	0.526	Numerical operations and mathematical reasoning	Sequential
Passolunghi and Mammarella	2010	0.391	0.150	0.588	Mathematical reasoning	Simultaneous and sequential
Alloway et al.	2009	0.639	0.567	0.701	Numerical operations and mathematical reasoning	Sequential and simultaneous
De Smedt et al.	2009	0.510	0.354	0.639	Numerical operations and mathematical reasoning	Simultaneous and sequential
Andersson	2008	0.040	-0.126	0.204	Numerical operations	Simultaneous and sequential
Bull, Espy and Wiebe	2008	0.230	0.039	0.405	Numerical operations	Sequential
Holmes, Adams and Hamilton	2008	0.300	0.117	0.463	Numerical operations and mathematical reasoning	Simultaneous and sequential
Kyttälä and Lehto	2008	0.412	0.257	0.547	Numerical operations and mathematical reasoning	Simultaneous and sequential
Andersson and Lyxell	2007	0.360	0.205	0.497	Numerical operations	Simultaneous and sequential
Holmes and Adams	2006	0.173	0.012	0.325	Numerical operations and mathematical reasoning	Simultaneous

Table 2 (continued)

Author(s)	Date	<i>r</i>	Lower CI	Upper CI	Type of mathematics	Type of VSWM
Bayliss et al.	2005	0.690	0.521	0.806	Numerical operations and mathematical reasoning	Simultaneous and sequential
D'Amico and Guamera	2005	0.595	0.285	0.792	Numerical operations	Simultaneous and sequential
Jarvis and Gathercole	2003	0.320	0.155	0.468	Numerical operations and mathematical reasoning	Simultaneous and sequential
Maybery and Do	2003	0.381	0.112	0.598	Numerical operations and mathematical reasoning	Sequential
Reuhkala	2001	0.574	0.437	0.685	Numerical operations and mathematical reasoning	Simultaneous and sequential

fluctuations in this relationship as children mature and undergo more formal schooling in mathematics.

The studies included adopted a number of methodological designs; however, no specific inclusion and exclusion criteria were defined regarding methodology as it was anticipated that a broad range of designs would be used. As a result, all study designs were included. Owing to the variety of methodological designs used, the resulting statistical analyses employed by the included studies also varied greatly. Whilst a vast majority of studies employed, at least as part of their analysis, ANOVA, correlation and regression techniques, additional techniques including factor analysis, structural equation modelling and multi-level modelling were used to further explain the data gathered. For this review, the main result from each study was converted to a correlation co-efficient, *r*, in order to make accurate comparisons between studies.

Quantitative Synthesis of Results

Overall Findings

Sufficient data were provided by each of the studies included in the analysis to be included in the quantitative synthesis. As noted above, a full meta-analysis of the data was not conducted due to the vast differences inherent in the study designs. It was deemed that there were insufficient similarities within the studies for a meta-analytical comparison to be tangible due to the impact on the subsequent interpretation of the results. Rather, inferential statistics were employed, where possible, to achieve an objective assessment of the relationships within the data. Analyses were conducted on a number of subsections of the data by way of identifying the possible sources of the aforementioned heterogeneity in order to better understand the relationship between VSWM and mathematics performance.

As previously mentioned, effect sizes were calculated based on the most relevant result to the review topic, with an average effect size taken in situations when more than one statistic was equally relevant. Since all effect sizes calculated resulted from different studies, they can be considered independent. The results demonstrated an overall positive relationship between VSWM and mathematics, as evidenced by the forest plot in Fig. 2 below. From the funnel plot, Fig. 1, publication bias appears minimal in the studies available on this subject.

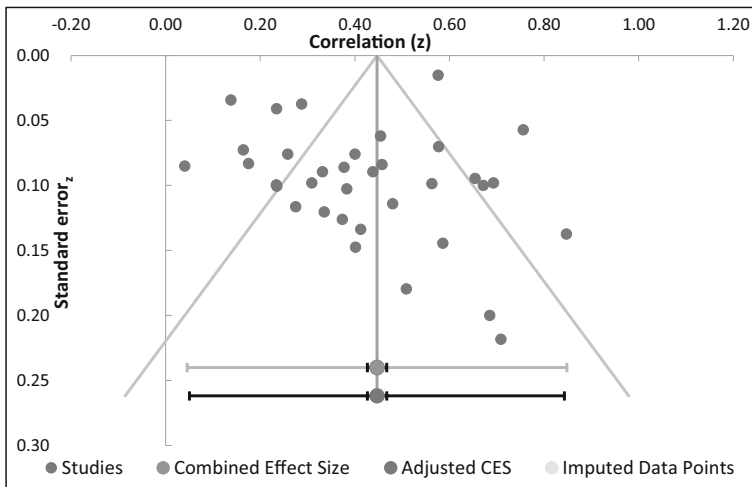


Fig. 1 Funnel plot showing a random distribution, suggesting no evidence of publication bias

Subsection Analysis

Sample Size

Upon investigation, a relationship between sample size and effect size is present within the data. This section strives to investigate this relationship further to understand the potential ways in which sample size may influence the effect sizes found.

Larger effect sizes appear concurrent with smaller sample sizes ($r_s = 0.340$, $p = 0.046$), as previously demonstrated in the literature as a common phenomenon and indicative of potential publication bias (Kühberger et al. 2014; Levine et al. 2009). The correlation between sample size and effect size is stronger following removal of one study with an extremely large sample size (Van de Weijer-Bergsma et al. 2015; $r_s = 0.404$, $p = 0.018$); however, caution should be applied when interpreting this finding due to issues of statistical power (Button et al. 2013). A medium-large effect size resulting from the study with the largest sample size (4337; Van de Weijer-Bergsma et al. 2015) sits comfortably within the range of effect sizes, hence reducing the potential influence of sample size on effect size (Button et al. 2013).

Sample sizes were divided into two groups for further analysis: small (mean = 115.82, sd = 65.84, lower bound = 24, upper bound = 308) and large (mean = 1627.75, sd = 1809.21, lower bound = 597, upper bound = 4337). No significant difference was found between the two groups ($t(33) = -1.357$, $p = 0.184$), suggesting a lesser influence on sample size than indicated by Button et al. (2013). Finally, once negative effect sizes were transformed into positive (via a reflection of the original due to the \pm difference resulting from the labels assigned to M1/M2), they did not deviate from the core cluster and, hence, show no significant differences from the remaining effect sizes.

Type of Mathematics

Approximately equal numbers of studies investigated Numerical Operations and both Numerical Operations and Mathematical Reasoning (17 and 16, respectively); however, only two

studies considered purely Mathematical Reasoning. Interestingly, the largest mean effect size was produced by studies concerning Mathematical Reasoning (mean = 0.49), with studies using small-average samples ($n = 30$ and $n = 103$), suggesting that this result cannot be explained by sample size alone. Those studies investigating both types of mathematics demonstrated the next largest effect size (mean = 0.43), followed by Numerical Operations only (mean = 0.35), indicating that VSWM may be more of an influencing factor in Mathematical Reasoning than Numerical Operations. Despite the aforementioned differences being present in the data, the between-group differences were not statistically significant ($F(2) = 1.380$, $p = 0.266$). It is evident from the data that Numerical Operations and a combination of both Numerical Operations and Mathematical Reasoning showed greater spread of effect sizes (range = 0.55 and 0.52, respectively), though only two studies looked at Mathematical Reasoning alone (range = 0.20). It is to be expected that the range of effect sizes resulting from studies of Mathematical Reasoning would have been greater if more studies had investigated Mathematical Reasoning alone.

Two studies (Maennamaa et al. 2012; Wiklund-Hörnqvist et al. 2016) investigated both types of mathematics using large samples, which may have skewed the average effect size generated for this subgroup as 13 studies used small samples. However, as suggested by Button et al. (2013), it may be the case that these larger samples provide the power to detect effects within the data and increase the likelihood that statistically significant results are reflective of true effects. Both studies assessing Mathematical Reasoning (Campos et al. 2013; Passolunghi and Mammarella 2010) had only small sample sizes (103 and 59, respectively) and as such, according to Button et al. (2013), the large effect sizes may be less likely to be representative of the true population effect.

Type of Visuospatial Working Memory

Studies were broken down according to the type of VSWM they assessed: simultaneous, sequential or both. The largest mean effect size was observed for studies concerning both simultaneous and sequential VSWM (mean = 0.44), followed by sequential (mean = 0.37), and simultaneous (mean = 0.25). Whilst this difference is marginally non-significant ($F(2) = 2.727$, $p = 0.081$), it is suggestive of a bias in the level of influence of each type of VSWM on mathematics performance.

The largest range of effect sizes can be seen in the data for both types of VSWM (range = 0.65), with smaller ranges seen for sequential and simultaneous (range = 0.44 and 0.29, respectively). Such a finding alludes to other influencing factors in studies measuring both types of VSWM due to the large range of effect sizes displayed. Further, it may suggest the more stable development of simultaneous VSWM by the age of children included in these studies (5 and 6 years, respectively, for both types of VSWM and simultaneous only). All four studies involving large sample sizes (Maennamaa et al. 2012; Mix et al., 2015; Van de Weijer-Bergsma et al. 2015; Wiklund-Hörnqvist et al. 2016) concerned both types of VSWM, which may explain, in part, the large range of effect sizes for this category (14 used small samples), whereas all studies concerning only simultaneous or sequential VSWM used small sample sizes.

Type of VSWM was measured alongside type of mathematics to ascertain further detail on more specific relationships between the two components. No studies investigated the influence of sequential VSWM on Mathematical Reasoning, highlighting a gap in the research requiring additional investigation. An ANOVA showed no significant effects when

using type of VSWM and type of mathematics as fixed effects. Simultaneous VSWM shows the lowest mean effect sizes for both Numerical Operations and both types of mathematics (mean = 0.28 and 0.23, respectively), suggesting that simultaneous VSWM has the smallest influence on mathematical performance in these areas of mathematics (simultaneous VSWM alone was not measured for Mathematical Reasoning). The largest mean effect size (mean = 0.49) was identified for both types of VSWM in Mathematical Reasoning tasks. A large mean effect size here implies a large influence of VSWM in Mathematical Reasoning tasks, in line with the additional demands of such tasks; however, only two studies (Campos et al. 2013; Passolunghi and Mammarella 2010) measured this combination and so caution should be exercised when generalising the result. As may be expected, studies measuring both types of VSWM showed the largest mean effect size, regardless of the type of mathematics being investigated (Numerical Operations, Mathematical Reasoning or both). One potential explanation for this may be the need to combine information and/or the complexity of the tasks used, particularly in the case of studies assessing Mathematical Reasoning and both types of mathematics.

Age of Participants

The age of participants at the beginning and end of each study was extracted for an in-depth analysis. Neither showed a significant correlation with effect size ($r_s(35) = -0.025$, $p = 0.885$; $r_s(35) = -0.178$, $p = 0.307$, respectively). The mean age at the beginning of the included studies was 7.89 years, with a range from 4 to 15 years ($sd = 2.44$). Once all studies had reached their conclusion, the mean age showed an increase to 9.86 years, ranging from 7 to 16 years ($sd = 2.35$). Studies concerning Numerical Operations showed the lowest mean age at the beginning of the study (mean = 7.29 years; range = 4–12); therefore, it is conceivable that the effect sizes generated for this type of mathematics might be affected by the involvement of such young participants. Further, the involvement of younger participants in studies surrounding Numerical Operations aligns with methods for teaching mathematics, whereby arithmetic skills are taught before any reference to word problems or other such questions linking to Mathematical Reasoning. Studies investigating both types of mathematics involved the largest age range of participants (mean = 8.44 years, range = 5–15 years). Such a large range of ages, and the combination of styles of mathematics questions, may have influenced the effect sizes collected due to the demand of the questions, particularly those relating to Mathematical Reasoning. Mathematical Reasoning questions requiring a high level of proficiency in reading may have proven particularly detrimental to young children's Mathematical Reasoning scores. A further potential influence on results concerns whether an age-appropriate/standardised measure of mathematical ability was taken to assess performance.

Studies assessing sequential VSWM involved the youngest mean age of participant (mean = 6.85 years), with an age range of 4–9 years. The mean age at the beginning of studies for those assessing simultaneous VSWM was non-significantly higher than sequential VSWM (mean = 7.25 years, $p = 0.955$). It would not be expected that an age difference as small as can be observed in the given data would have a significant impact on VSWM and mathematics performance. The largest observable age range can be seen for studies examining both types of VSWM (range = 5–15 years), with a mean age of 8.78 years. This is also the group of studies with the oldest mean age. All studies using older children, of secondary school age, fall into this category, which would be expected to influence the results as it is expected that older children will have a larger WM capacity (Gathercole et al. 2004).

Standardised Measures

Studies were examined according to whether they had employed a standardised measure of VSWM or not. The mean effect size for studies using a standardised measure was higher, but not statistically significantly so, than that found for those using non-standardised measures (mean = 0.40 and mean = 0.38, respectively; $t(33) = 0.212$, $p = 0.833$). A non-significant finding here indicates that standardised and non-standardised measures appear to be equally effective at measuring VSWM in relation to mathematics performance. Further, there was no significant relationship between the size of the sample used and the use of a standardised measure ($t(12.154) = -1.143$, $p = 0.275$). As such, the use of a standardised measure and sample size is unlikely to have a compound influence on effect size.

Studies were then examined according to their use of a standardised mathematics measure. The mean effect size gathered for studies using a standardised mathematics measure (mean = 0.44) was significantly higher than those using unstandardised measures (mean = 0.25, $t(33) = 3.587$, $p = 0.001$). Such a finding highlights the importance of using standardised mathematics measures in order to uncover the true extent of any relationship between mathematics performance and VSWM. As with measures of VSWM, the data do not show a significant relationship between the size of the sample used and the use of a standardised measure ($t(33) = 0.125$, $p = 0.901$). Therefore, the size of the sample is unlikely to have a compound effect on the already significant influence of the use of a standardised measure.

Overall, the results indicate the importance of using a standardised measure of mathematics when investigating the relationship between VSWM and mathematics performance. However, they also suggest that the use of a non-standardised measure of VSWM does not necessarily prove detrimental to the integrity of the study.

Discussion

Systematic Review Results Summary

This review concerned 35 independent studies, following thorough examination of each document to ensure no overlaps between studies were present. The review was conducted with the aim of producing a comprehensive overview of the current knowledge base relating to the relationship between visuospatial working memory and mathematics. The included studies comprised of a number of designs, and involved a variety of assessments of both mathematics and VSWM. Whilst this is a relatively small sample of studies for the purposes of a review, there were a sufficient number to conduct further analysis. A forest plot and funnel plot (Figs. 1 and 2) were generated to give an overview of the data before inferential statistics were applied in the absence of a meta-analysis.

The number of studies analysed for this review is reflective of the current understanding of the relationship between VSWM and mathematics. Research remains in its relative infancy; therefore, the intricacies of the relationship are as yet unknown. For example, the earliest study in this review demonstrates the first published study documenting the specific relationship as taking place in 2001 (Reuhkala 2001).

No other systematic reviews on this area of the research have been published, to our knowledge, up to the date of writing; hence, there is great scope for collating the findings of the research to date. As such, it is not possible to draw comparisons with the findings of

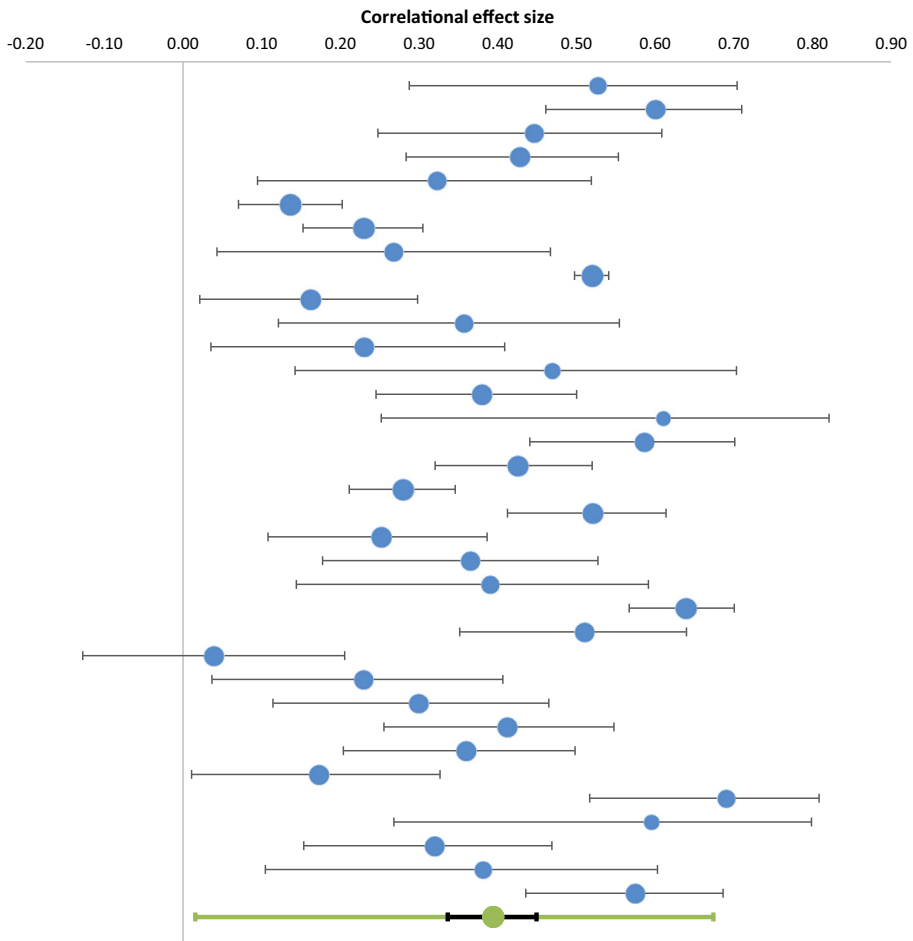


Fig. 2 Forest plot showing an overall significant positive relationship between VSWM and mathematics

reviews of other aspects of this area of research. The lack of reviews previously completed in this area indicates the need to develop a comprehensive understanding of the given relationship before continuing with further research.

Quantitative Analysis Results Summary

The review results highlight the importance of a sufficiently large sample in order to detect any effect within the data and accurately determine its significance, as evidenced by the negative correlation identified between effect size and sample size. The inclusion of only two studies exploring solely Mathematical Reasoning demonstrates an evident lack in the literature of such focused work; however, a further 16 studies investigated Mathematical Reasoning in conjunction with numerical operations.

From the evidence, it appears that numerical operations and mathematical reasoning are both influenced to a similar extent by VSWM; however, the level of influence within each of these types of maths is variable. The greatest variation can be seen for numerical operations. A

bias in the amount of influence of the type of VSWM is suggested from the data. Nevertheless, once the type of mathematics being assessed is included in the analysis, the difference is not significant. Further, age did not have a significant impact on the effect sizes generated, nor did the use of a standardised VSWM measure. On the contrary, the use of a standardised mathematics measure resulted in a significantly larger effect size. One possible reason for such a difference may be the design of standardised measures to rigorously assess specific areas of mathematics and address all areas of the curriculum.

Quality of the Evidence

Five hundred and ninety studies were screened before arriving at the final sample of 35, suggesting that a sufficiently scoping search was completed to identify all relevant available literature, in line with the inclusion criteria. This suggestion is supported as all relevant studies were available in full.

As previously mentioned, the studies included employ a number of designs, measures and methods of analysis. This emphasises the need to apply caution when attempting to directly compare across studies. However, sufficient data was provided within each manuscript to allow for the calculation of effect sizes, thus allowing less problematic, direct comparisons due to the common scale. Comparisons have also been drawn regarding the variance accounted for, in order to examine the extent of the influence, as well its significance, so as to reduce the probability of making type 2 errors, given the potential for the small studies included to be underpowered. As a result, the conclusions drawn from the data in this review seem relatively robust.

Conclusions

This review analysed the available literature on the relationship between VSWM and mathematics and proposes that the type of VSWM and mathematics being assessed do not have significant influence; however, the use of a standardised mathematics measure demonstrates significant influence on the effect size generated. Overall, a positive influence of VSWM on mathematics attainment is evident.

Implications for Research

The findings presented above suggest the greatest implications for those seeking to develop VSWM research in relation to mathematics performance. Since there is the suggestion that the use of a standardised mathematics measure significantly influences the estimation of the level of effect, researchers ought to be cautious of devising their own measures of mathematics attainment where a suitable standardised measure is available.

There is a great deal of scope for further research suggested by the findings of this review, as well as the gaps in the research identified throughout. Additional research is necessary to determine the stability of the relationship as identified over the years children spend at school. For example, in order for preventative measures for mathematical difficulties to be devised, it is first necessary to understand the intricacies of the relationship. Additionally, further research should seek to identify whether the relationship identified throughout is specific to components of mathematics, or whether the explanation satisfies mathematics in general.

Acknowledgements This research was funded via a PhD studentship from the North East and Northern Ireland Doctoral Training Partnership, which is in turn funded through the Economic and Social Research Council.

Compliance with Ethical Standards This research is funded by the Economic and Social Research Council. There are no known conflicts of interest, financial or otherwise, and no data was gathered from human participants to warrant obtaining informed consent.

OpenAccess This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Studies included in the analysis are denoted by an asterisk.

- *Alloway, T. P., 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. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ and mathematical skills in children. *Learning and Individual Differences*, 21(1), 133–137. <https://doi.org/10.1016/j.lindif.2010.09.013>
- *Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: the importance of central executive functions. *British Journal of Educational Psychology*, 78(2), 181–203. <https://doi.org/10.1348/000709907X209854>
- *Andersson, U., & Lyxell, B. (2007). Working memory deficit in children with mathematical difficulties: a general or specific deficit? *Journal of Experimental Child Psychology*, 96(3), 197–228. <https://doi.org/10.1016/j.jecp.2006.10.001>
- *Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S., Swigart, A. G., & Menon, V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia*, 51(11), 2305–2317. <https://doi.org/10.1016/j.neuropsychologia.2013.06.031>
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). Academic Press.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2).
- *Bayliss, D. M., Jarrod, C., Baddeley, A. D., & Gunn, D. M. (2005). The relationship between short-term memory and working memory: complex span made simple? *Memory*, 13(3–4), 414–421. <https://doi.org/10.1080/09658210344000332>
- Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development*, 19(3), 291–307. <https://doi.org/10.1016/j.cogdev.2004.03.005>.
- Berch, D. B. (2005). Making sense of number sense: implications for children with mathematical disabilities. *Journal of Learning Disabilities*, 38(4), 333–339. <https://doi.org/10.1177/00222194050380040901>.
- *Bresgi, L., Alexander, D. L. M., & Seabi, J. (2017). The predictive relationships between working memory skills within the spatial and verbal domains and mathematical performance of Grade 2 South African learners. *International Journal of Educational Research*, 81(Supplement C), 1–10. <https://doi.org/10.1016/j.ijer.2016.10.004>
- *Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, 33(3), 205–228. <https://doi.org/10.1080/87565640801982312>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>.

- *Campos, I. S., Almeida, L. S., Ferreira, A. I., Martinez, L. F., & Ramalho, G. (2013). Cognitive processes and math performance: a study with children at third grade of basic education. *European Journal of Psychology of Education*, 28(2), 421–436. <http://www.jstor.org/stable/23421902>.
- *Caviola, S., Mammarella, I. C., Cornoldi, C., & Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *Journal of Experimental Child Psychology*, 112(2), 141–160. <https://doi.org/10.1016/j.jecp.2012.02.005>
- Clements, D. H. (1999). Subitizing: what is it? why teach it? *Teaching Children Mathematics*, 5(7), 400.
- *D'Amico, A., & Guamera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences*, 15(3), 189–202. <https://doi.org/10.1016/j.lindif.2005.01.002>
- De Cruz, H. (2006). Towards a Darwinian approach to mathematics. *Foundations of Science*, 11(1), 157–196. <https://doi.org/10.1007/s10699-004-5916-z>.
- *De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: a longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*, 103(2), 186–201. <https://doi.org/10.1016/j.jecp.2009.01.004>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1), 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N).
- Dehaene, S. (2003). *The neural basis of the Weber-Fechner law: a logarithmic mental number line*. *Trends in Cognitive Sciences* 7(4). Netherlands: Elsevier Science. [https://doi.org/10.1016/S1364-6613\(03\)00055-X](https://doi.org/10.1016/S1364-6613(03)00055-X).
- Dehaene, S. (2001). Précis of the number sense. *Mind & Language*, 16(1), 16–36. <https://doi.org/10.1111/1468-0017.00154>.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177–190. <https://doi.org/10.1037/0012-1649.40.2.177>.
- *Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology*, 47(6), 1539–1552. <https://doi.org/10.1037/a0025510>
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, 78(4), 1343–1359 <http://www.jstor.org/stable/4620706>.
- Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Harvard University Press.
- Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy. *The American Journal of Psychology*, 91(2), 267–273. <https://doi.org/10.2307/1421536>.
- Henson, R. (2002). Working memory in perspective. In J. Andrade (Ed.). Taylor & Francis.
- *Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: implications for mathematical development and mathematics curricula. *Educational Psychology*, 26(3), 339–366. <https://doi.org/10.1080/01443410500341056>
- *Holmes, J., Adams, J. W., & Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology*, 20(2), 272–289. <https://doi.org/10.1080/09541440701612702>
- Holmes, J., & Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educational Psychology*, 34(4), 440–450. <https://doi.org/10.1080/01443410.2013.797338>.
- *Jarvis, H. L., & Gathercole, S. (2003). Verbal and non-verbal working memory and achievements on National Curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, 20, 123–140.
- Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: findings from a four-year longitudinal study. *Learning and Instruction*, 19(6), 513–526. <https://doi.org/10.1016/j.learninstruc.2008.10.002>.
- Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: a diagnosis based on the correlation between effect size and sample size. *PLoS One*, 9(9), e105825. <https://doi.org/10.1371/journal.pone.0105825>.
- *Kyttaelae, 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. <https://link.springer.com/content/pdf/10.1007%2FBF03173141.pdf>
- Lenhard, W., & Lenhard, A. (2016). Calculation of effect sizes. https://www.psychometrica.de/effect_size.html.
- Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: evidence and implications of a publication bias against non-significant findings. *Communication Monographs*, 76(3), 286–302. <https://doi.org/10.1080/03637750903074685>.
- *Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PLoS One*, 12(2). <https://doi.org/10.1371/journal.pone.0172046>

- *Li, Y., & Geary, D. C. (2013). Developmental gains in visuospatial memory predict gains in mathematics achievement. *PLoS One*, 8(7). <https://doi.org/10.1371/journal.pone.0070160>, e70160
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, 14(6), 1292–1300. <https://doi.org/10.1111/j.1467-7687.2011.01080.x>.
- Logie, R. H. (2014). *Visuo-spatial working memory*. Taylor & Francis.
- *Maennamaa, M., Kikasb, E., Peets, K., & Palu, A. (2012). Cognitive correlates of math skills in third-grade students. *Educational Psychology*, 32(1), 21–44. <https://doi.org/10.1080/01443410.2011.621713>
- Mammarella, I. C., Borella, E., Pastore, M., & Pazzaglia, F. (2013). The structure of visuospatial memory in adulthood. *Learning and Individual Differences*, 25, 99–110. <https://doi.org/10.1016/j.lindif.2013.01.014>.
- *Mammarella, I. C., Caviola, S., Giofrè, D., & Szűcs, D. (2017). The underlying structure of visuospatial working memory in children with mathematical learning disability. *British Journal of Developmental Psychology* 36(2). <https://doi.org/10.1111/bjdp.12202>, 220, 235
- Mammarella, I. C., Pazzaglia, F., & Cornoldi, C. (2008). Evidence for different components in children's visuospatial working memory. *British Journal of Developmental Psychology*, 26(3), 337–355. <https://doi.org/10.1348/026151007X236061>.
- Mammarella, I. C., Cornoldi, C., Pazzaglia, F., Toso, C., Grimoldi, M., & Vio, C. (2006). Evidence for a double dissociation between spatial-simultaneous and spatial-sequential working memory in visuospatial (nonverbal) learning disabled children. *Brain and Cognition*, 62(1), 58–67. <https://doi.org/10.1016/j.bandc.2006.03.007>.
- *Martin, R. B., Cirino, P. T., Sharp, C., & Barnes, M. (2014). Number and counting skills in kindergarten as predictors of grade 1 mathematical skills. *Learning and Individual Differences*, 34(Supplement C), 12–23. <https://doi.org/10.1016/j.lindif.2014.05.006>
- *Maybery, M. T., & Do, N. (2003). Relationships between facets of working memory and performance on a curriculum-based mathematics test in children. *Educational and Child Psychology*, 20(3), 77–92.
- *Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences*, 20(2), 101–109. <https://doi.org/10.1016/j.lindif.2009.08.004>
- *Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: the latent structure of space and mathematics across development. *Journal of Experimental Psychology: General* 145(9). <https://doi.org/10.1037/xge0000182>, 1206, 1227
- *Nath, S., & Szűcs, D. (2014). Construction play and cognitive skills associated with the development of mathematical abilities in 7-year-old children. *Learning and Instruction*, 32(Supplement C), 73–80. <https://doi.org/10.1016/j.learninstruc.2014.01.006>
- *Passolunghi, M. C., & Mammarella, I. C. (2010). Spatial and visual working memory ability in children with difficulties in arithmetic word problem solving. *European Journal of Cognitive Psychology*, 22(6), 944–963. <https://doi.org/10.1080/09541440903091127>
- Passolunghi, M. C., & Mammarella, I. C. (2011). Selective spatial working memory impairment in a group of children with mathematics learning disabilities and poor problem-solving skills. *Journal of Learning Disabilities*, 45(4), 341–350. <https://doi.org/10.1177/0022219411400746>.
- *Pina, V., Fuentes, L. J., Castillo, A., & Diamantopoulou, S. (2014). Disentangling the effects of working memory, language, parental education, and non-verbal intelligence on children's mathematical abilities. *Frontiers in Psychology*, 5, 415. <https://doi.org/10.3389/fpsyg.2014.00415>
- *Reuhkala, M. (2001). Mathematical skills in ninth-graders: relationship with visuo-spatial abilities and working memory. *Educational Psychology*, 21(4), 387–399. <https://doi.org/10.1080/01443410120090786>
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75(2), 428–444. <https://doi.org/10.1111/j.1467-8624.2004.00684.x>.
- Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: a meta-analysis of the available evidence. *Developmental Psychology*, 53(4), 671–685. <https://doi.org/10.1037/dev0000265>.
- Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: preschoolers see number words as specific. *Cognition*, 92(3), 329–352. <https://doi.org/10.1016/j.cognition.2003.10.001>.
- Sarnecka, B. W., & Wright, C. E. (2013). The idea of an exact number: children's understanding of cardinality and equinumerosity. *Cognitive Science*, 37(8), 1493–1506. <https://doi.org/10.1111/cogs.12043>.
- Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *Journal of Educational Psychology*, 101(2), 359–372. <https://doi.org/10.1037/a0013840>.
- *Soltanlou, M., Pixner, S., & Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: a longitudinal investigation. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01062>

- Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, 36(2), 97–127. [https://doi.org/10.1016/0010-0277\(90\)90001-Z](https://doi.org/10.1016/0010-0277(90)90001-Z).
- Szűcs, D. (2016). Chapter 11—subtypes and comorbidity in mathematical learning disabilities: multidimensional study of verbal and visual memory processes is key to understanding. In M. Cappelletti & W. B. T.-P. In B. R. Fias (Eds.), *The Mathematical brain across the lifespan* (Vol. 227, pp. 277–304). Elsevier. <https://doi.org/10.1016/bs.pbr.2016.04.027>
- *Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 49(10), 2674–2688. <https://doi.org/10.1016/j.cortex.2013.06.007>
- Tan, L. H., Liu, H.-L., Perfetti, C. A., Spinks, J. A., Fox, P. T., & Gao, J.-H. (2001). The neural system underlying Chinese logograph reading. *NeuroImage*, 13(5), 836–846. <https://doi.org/10.1006/nimg.2001.0749>.
- Thompson, I. (2010). The principal counting principles. *National Centre for Excellence in the Teaching of Mathematics, Early Years Magazine*, 7.
- Thompson, P. W. (1996). Imagery and the development of mathematical reasoning. In *Theories of learning mathematics*.
- *Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Memory & Cognition*, 43(3), 367–378. <https://doi.org/10.3758/s13421-014-0480-4>
- *Vanbinst, K., Ceulemans, E., Peters, L., Ghesquière, P., & De Smedt, B. (2018). Developmental trajectories of children’s symbolic numerical magnitude processing skills and associated cognitive competencies. *Journal of Experimental Child Psychology*, 166(Supplement C), 232–250. <https://doi.org/10.1016/j.jecp.2017.08.008>
- *Vandenbroucke, L., Verschueren, K., Desoete, A., Aunio, P., Ghesquière, P., & Baeyens, D. (2018). Crossing the bridge to elementary school: the development of children’s working memory components in relation to teacher-student relationships and academic achievement. *Early Childhood Research Quarterly*, 42(Supplement C), 1–10. <https://doi.org/10.1016/j.ecresq.2017.08.004>
- von Hippel, P. T. (2015). The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Medical Research Methodology*, 15(35). <https://doi.org/10.1186/s12874-015-0024-z>.
- Wechsler, D. (2017). *Wechsler Individual Achievement Test—third UK edition (WIAT-III UK) | Pearson Assessment*.
- *Wiklund-Hörnqvist, C., Jonsson, B., Korhonen, J., Eklöf, H., & Nyroos, M. (2016). Untangling the contribution of the subcomponents of working memory to mathematical proficiency as measured by the National Tests: a study among Swedish third graders. *Frontiers in Psychology*, 7, 1062. <https://doi.org/10.3389/fpsyg.2016.01062>
- Wilson, D. B. (n.d.). *Practical meta-analysis effect size calculator (online calculator)*.
- Wynn, K. (1990). Children’s understanding of counting. *Cognition*, 36(2), 155–193. [https://doi.org/10.1016/0010-0277\(90\)90003-3](https://doi.org/10.1016/0010-0277(90)90003-3).
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1–B11. [https://doi.org/10.1016/S0010-0277\(99\)00066-9](https://doi.org/10.1016/S0010-0277(99)00066-9).