UNIVERSITY OF CALIFORNIA,
IRVINE

The role of dual-task experiments in working memory and arithmetic performance

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Edward Harry Chen

Dissertation Committee:
Associate Professor Drew Bailey, Co-Chair
Professor Susanne Jaeggi, Co-Chair
Professor Lindsey Richland

2022

DEDICATION

To

my friends, family, and students

for their unwavering support

When all is said and done,

Strive to do good science

&

Be a good person

Table of Contents

LIST OF FIGURES

Page

Study 1

Study 2

Study 3

LIST OF TABLES

Page

Study 1

Study 2

Study 3 85

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my co-chairs, Professors Drew Bailey and Susanne Jaeggi for their expert guidance throughout my graduate career. It is with their scholarship and mentorship that I attempt to model my scientific endeavors after. Without their continued support, this project and my completion of the program would not have been possible.

I would also like to thank Professor Lindsey Richland for her work on my committee and for her service in helping all graduate students of the School of Education obtain the same quality of experience that I received.

Thank you Professor Janice Hansen for revitalizing my love for teaching and learning. I hope I carry with me the drive to help students the way you do every day.

In addition, a thank you to Professor Barbara Sarnecka, my undergraduate research advisor. Your lab always seems to produce fantastic scholars.

Finally, I wanted to thank my team of research assistants: Gladys Aguilar, Yu Jiang, Stacy Le, Shannon Le, Zoe Liuag, Flora Song, and Muxi Zheng. You are all destined for great things. I am proud of you all and grateful for your help in making this project a reality.

VITA

Edward Harry Chen

2016	B.A.s in Psychology & Education, University of California, Irvine
2017-2018	Graduate Student Researcher, UCLINKS, University of California, Irvine
2019-2020	Graduate Student Researcher, SPARCS Project, University of California, Irvine
2021	Instructor, School of Education, University of California, Irvine
2017-2022	Teaching Assistant, School of Education, University of California, Irvine
2022	Instructor, School of Education, University of California, Irvine
2022	Ph.D. in Education, University of California, Irvine

FIELD OF STUDY

Human Development in Context, Mathematical Cognition

PUBLICATIONS

Chen, E. H., Jaeggi, S.M., & Bailey, D. H. (in press). No clear support for differential influence of visuospatial and phonological resources on mental arithmetic: A Registered Report. *Journal of Numerical Cognition*.

Qiu, K., Chen, E. H., Wan, S., & Bailey, D. H. (2021). A multilevel meta-analysis on the causal effect of approximate number system training on symbolic math performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Chen, E. H., & Bailey, D. H. (2021). Dual-task studies of working memory and arithmetic performance: A meta-analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(2), 220.

ABSTRACT OF THE DISSERTATION

The role of dual-task experiments in working memory and arithmetic performance
by
Edward Harry Chen
Doctor of Philosophy in Education
University of California, Irvine, 2022
Professor Drew H. Bailey, Co-chair
Professor Susanne M. Jaeggi, Co-chair

This dissertation looks across theories of working memory to evaluate the competing claims regarding its influence on mental arithmetic. I attempt to reconcile differences between correlational and experimental literatures to better understand the specific impacts of working memory on arithmetic through meta-analytic and experimental methods following a dual-task paradigm. In order to investigate whether working memory is causally linked to math, I conducted a meta-analysis on dual-task experiments, summarizing the effects of secondary task load on mental arithmetic performance (Chapter 1). In addition, I tested a number of relevant moderators, including the secondary task type, primary arithmetic task type, difficulty, and combinations of primary and secondary tasks. While results supported a robust causal effect of working memory, it was unclear if arithmetic performance was affected purely by the cognitive demands of the tasks or if they were affected by similarly shared resources with the secondary task. Thus, I conducted a registered report in which I probed whether arithmetic operations are differentially impacted by various types of working memory secondary tasks by replicating an influential dual-task experiment and testing other relevant factors that predicted differences in dual-task performance (Chapter 2). The within-subject experiment tested whether there was differential interference of verbal and visuospatial loads on multiplication and subtraction performance and whether it could be generalized to sub-populations and difficulty levels. Prior results from the influential experiment were not replicated, so I conducted further analyses including an additional task to test whether different theories of working memory could reliably predict any kind of interaction between working memory and arithmetic tasks (Chapter 3). I tested main effects and interactions between secondary task type, arithmetic operations, and difficulty and explored whether such task features are implicated in arithmetic strategy choice. I found strong evidence for main effects of the aforementioned factors on arithmetic but not interactions between them except in the case of an arithmetic-based secondary task load. Overall, this dissertation found more evidence to support domain-general models of the influence of working memory on arithmetic performance rather than domain-specific models discussed within many previous dual-task studies. Lastly, I discuss the implications of these results as well as an outlook for future research.

Overview of studies

Working memory is a limited-capacity, mental workspace, involved in the temporary storage and active processing of relevant information to perform complex cognitive tasks, such as mathematical cognition (Baddeley, 1992). Mathematical cognition involves processing different types of mathematical content, such as arithmetic, algebra, and geometry. Often, problem solving within these domains involves temporarily holding onto partial information and processing new information to find a solution, which may require working memory resources. While it is well established that a relation between working memory and math exists (e.g., Peng, Namkung, & Barnes, 2016) and may contribute to our understanding of individual differences in mathematical development, a causal link between working memory and math has been much more difficult to establish from cognitive training studies (Sala & Gobet, 2017, 2020). More so, the exact nature of this causal relationship – how and in what ways does working memory influence mathematics – is not well understood either.

An alternative experimental paradigm that allows researchers to test these specific effects is the dual-task experiment. By having participants complete a primary cognitive task (e.g., mental arithmetic) simultaneously with a secondary working memory task (e.g., recalling letters or numbers), researchers can examine how arithmetic performance varies by the cognitive demands of the working memory tasks. If the primary and secondary tasks are relying on the same pool of cognitive resources (i.e., working memory), then reaction times and accuracy are likely to suffer. What makes this paradigm particularly useful is that it allows researchers to isolate specific components of working memory (e.g., verbal and visuospatial memory) and test how each component differentially impacts primary task performance. However, there is some debate as to whether such differential effects can be attributed to having to difficulties in processing semantically similar content simultaneously or to difficulties in maintaining attentional control on increasingly difficult cognitive loads (Pashler, 1994). Likely, both are true to some extent, but this divide circles back to the larger issue of determining the specific ways that working memory influences mathematical processing. Therefore, the goal of this dissertation is to reconcile the differences in the working memory and arithmetic literature by systematically reviewing and experimentally investigating the dual-task arithmetic literature.

In Chapter 1, I investigated the proposed causal relationship between working memory and mathematics performance by meta-analyzing a sample of dual-tasking literature. Dual-task experiments have been used in math cognition research as a way to show how mental arithmetic taps into different pools of working memory resources, thereby providing a supplementary account of the link as opposed to a direct account from cognitive training literature. I tested whether dual-task studies showed a robust effect on arithmetic performance as well as investigated whether and how study design features contributed to the effects' heterogeneity. Key moderators of interest included the type of secondary task load, type of arithmetic task, difficulty, and a proxy for primary and secondary task combinations using author predictions. Based on my meta-analysis of 400 effect sizes from 21 studies and 1,049 participants, I concluded that there was clear evidence for a causal effect of working memory on arithmetic performance. However, the type of working memory task strongly moderated arithmetic performance; specifically, tasks that were more cognitively demanding, requiring participants to mentally manipulate larger amounts of information, produced the largest effects. Finally, the effects of author-hypothesized combinations of working memory and arithmetic tasks were surprisingly small (although not exactly 0) – suggesting that general cognitive resources may underlie the associations between working memory and math performance reported in previous work. However, the results from the moderator analyses suggest that previously proposed models of working memory that focus on component-specific influences may not adequately explain dual-task performance, but rather models emphasizing more general, attentional control. Thus, questions remained as to whether arithmetic draws on content-specific working memory resources.

In Chapter 2, I further probed the possibility of domain-specific cognitive influences like verbal and visuospatial memory on arithmetic performance by re-examining the robustness of possible interactions between combinations of primary and secondary tasks (sometimes called “crosstalk” effects) in a new experiment. Prior dual-task literature examined how different secondary task loads affected arithmetic operations to mixed effect. One previous study reported a large crosstalk interaction between specific components of working memory and arithmetic operation which had not been replicated (Lee & Kang, 2002; Imbo & LeFevre, 2010; Cavdaroglu & Knops, 2016). Using data from 100 participants in a within-subject experiment, I tested the robustness of an interaction between secondary task type and arithmetic operation. The experiment consisted of 3 within-subject factors: cognitive load difficulty (easy vs. hard); arithmetic operation (multiplication v s. subtraction); and working memory load type (phonological, visuospatial, and none/arithmetic without load). However, difficulty was only used for subsample analyses in this chapter. Using a combination of frequentist and Bayesian methods as well as numerous subsample analyses, I was able to produce main effects for secondary task type but no interaction between primary and secondary tasks was found for any of my analyses The results suggest that the original findings may have been specific to whichever methods and conditions the previous authors employed and that domain-specific influences may be much more elusive than previously thought.

The final chapter of the dissertation is an extension of the experiment in Chapter 2 that experimentally investigates theories of working memory within a dual-task paradigm. In Chapter 3, I contrasted competing theories of working memory in dual-task performance which argue that performance is dictated by either shared cognitive resources, general task demands, or some integration of the two. In addition to using the same experimental design from Chapter 2, I included an addition secondary task in the analysis models and explored whether task features influenced arithmetic strategies. I then tested different sets of predictions following the general themes of a domain-specific model, domain-general model, and multiple integrated models of working memory using a similar combination of frequentist and Bayesian methods as Chapter 2. Results supported a domain-general model of working memory in dual-task arithmetic performance where task demands reliably worsened arithmetic performance. The inclusion of an addition task as a more direct line of domain-specificity did produce reliable effects above and beyond other secondary tasks, but its effects could be attributed to its overall difficulty rather than any content-dependent sharing of resources. Overall, while there is strong evidence for a domain-general model, there might still be selective effects of secondary tasks on arithmetic. However, these effects are likely much more difficult to predict and smaller than expected because of individual differences in participants' dual-task arithmetic strategies and to some extent the high degree of specificity that may be needed to produce such effects between the primary and secondary tasks.

Study 1

Dual-task studies of working memory and arithmetic: A meta-analysis

Accepted for publication at *Journal of Experimental Psychology: Learning, Memory, & Cognition*

Study 1

It is widely agreed that various aspects of arithmetic performance are dependent on working memory, but the strength of this relation and the degree to which different features of working memory contribute to this performance is difficult to study. Arithmetic procedures involve the temporary storage and manipulation of numerical elements across multiple steps (Hitch, 1978). For example, individuals solving multi-digit arithmetic problems, such as 23 × 16 or 256 + 169 must encode the problem they are working with, perform a number of calculations, and maintain these intermediate values in order to form a coherent solution to the arithmetic problem (for a review, see Raghubar, Barnes, & Hecht, 2010). Arithmetical processing, therefore, appears to rely heavily on working memory. Consistent with this theory, working memory has been found to be reliably correlated with performance on mathematical tasks and has been found to statistically predict children’s mathematics outcomes (for reviews, see Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, [2013] and Raghubar, Barnes, & Hecht, [2010]).

However, some specific questions about the nature of working memory resources influencing arithmetic performance have been difficult to address. In particular, the specificity of these effects to particular facets load types of working memory (e.g. differential impact of the visuospatial system versus the phonological system on subtraction performance), types of arithmetic, and interactions among them, is limited by the use of correlational designs. Two more causally informative approaches are training studies, where participants learn to better utilize working memory resources, and the dual-task experimental design, whereby participants perform a primary cognitive task (e.g. multiplication) concurrently with another secondary task (e.g. pressing a left key when hearing a low tone through a headset or the right key when hearing a high tone). Dual-task experiments offer an alternative to correlational or training studies by allowing the researcher to both experimentally manipulate working memory load and the overlap between features of the working memory task and the arithmetic task to investigate task overlap (Logie & Baddeley, 1987; Ashcraft, Donley, Halas, Vakali, 1992; for review, see Pashler, 1994). Our goal is to investigate the specificity of working memory functions in arithmetic through such a design and provide insight on the discrepancy between experimental and correlational findings. Thus, we review the literature on the role of working memory in arithmetic processing and provide meta-analytic evidence to characterize the causal relation between working memory and arithmetic as studied in dual-task experiments.

Causal Effects of Working Memory on Arithmetic?

Working memory has been conceptualized in a variety of ways. Some models suggest that it as distinct from executive functions while others argue that executive functions subsume working memory functions. Research on the cognitive processes that underlie arithmetical cognition has been largely influenced by the multicomponent model of working memory conceptualized by Baddeley and Hitch (1974). According to this model, working memory is a limited capacity system responsible for short-term storage and manipulation of elements within cognitive processes (Diamond, 2013; Miyake & Shah, 1999). The model has been refined over time but often separates working memory into three core subcomponents: the central executive (CE), phonological loop (PL), and visuospatial sketchpad (VSSP). Miyake's and colleagues' (2000) theory of executive functions proposes three aspects of executive functions: updating, shifting, inhibition. Updating involves the constantly monitoring and adding/deleting WM contents, shifting involves switching between tasks and mental sets, and inhibition involves the conscious overriding of predominant responses. Unlike the Baddeley model, there are no subcomponent systems (i.e. the visuospatial sketchpad and phonological loop), and WM is viewed separately as a passive-storage system that relies on executive functions. Diamond's model of EF is similar to Miyake's in that it includes inhibition and views WM as a separate construct, but it differs by including cognitive flexibility which involves task switching (analogous to shifting). Dual-tasks have been thought to involve executive functions such as those found in Miyake and Baddeley's models (specifically that of shifting), but no consensus has yet been found regarding the specificity of which executive function. Another perspective involves EFs as part of WM, that is, WM capacity is the ability to use attention to maintain or suppress information (Engle, 2002; Awh, Vogel, & Oh, 2006). Engle (2002) posits that a greater WM capacity is indicative of greater ability to control attention rather than a larger memory storage. While many alternative models to EF and WM have been proposed over the last few decades, the focus of this meta-analysis is on Baddeley's multicomponent model. The primary reason for this is that dual-task research including arithmetic tasks is largely predicated on this model, specifically predictions concerning its subcomponents and variations in arithmetic tasks. Thus, arithmetic tasks are hypothesized to be subject to interference to the extent that they overlap with specific arithmetic processing on Baddeley's subcomponents. More general processes, such as switching or inhibition, are likely required across most dual-task pairs, although perhaps in different amounts, with tasks hypothesized to require CE resources involving more switching and inhibition than tasks hypothesized to require only PL and VSSP. Thus, results can be interpreted with respect to all of these theories, but hypothesized examples of crosstalk pertain most directly to Baddeley's.

The central executive is the most important component of Baddeley's model. It acts in a supervisory role between the other two subsystems by coordinating visual and verbal information and between working memory and long-term memory. Compared to the PL and VSSP whose functions are more domain-specific and storage-based, the CE is amodal and facilitates processing. Beyond a supervisory position, the CE's other main functions include selective attention, inhibiting or suppressing automatic responses, updating working memory with new information, and shifting between tasks. Within Baddeley's model, the CE would appear to play a pivotal role across all types of single- and multi-digit arithmetic operations, because of the split in attentional resources and the necessity to maintain intermediate results (for reviews, see Raghubar, Barnes, & Hecht [2010] and DeStefano & LeFevre [2004]).

The phonological loop aids in temporarily storing and rehearsing verbal information. In the context of arithmetic cognition, the phonological loop seems to be primarily involved in verbally mediating calculation strategies, such as decomposing, transforming, and counting up/down in multi-digit arithmetic (Furst & Hitch, 2000; Imbo & Vandierendonck, 2007a).

The visuospatial sketchpad is responsible for the storage and processing of visual and spatial information of an element, such as its shape and position. The visuospatial sketchpad has been viewed as especially important to the development of mental arithmetic in young children whereby their use of the mental number line is reliant on visuospatial encoding (Hubbard et al. 2005; McKenzie, Bull, & Gray, 2003). While the role of the VSSP is less understood in mental arithmetic, some have found evidence suggesting that it is involved in strategy use (though to a much lesser extent than the PL and CE because these more sophisticated strategies take time to develop) and more difficult arithmetic problems in both children and adults, such as those requiring carrying operations or the encoding of intermediate results (Rasmussen & Bisanz, 2005; Xenidou-Dervou, van der Schoot & van Lieshot, 2015; Noël, Desert, Aubrun, & Seron, 2001; Logie, Gilhooly, & Wynn, 1994; Hubber, Gilmore, & Cragg, 2014).

Another component, the episodic buffer, was later added to the Baddeley model to explain how the central executive interacts with the other subsystems; however, there is little discussion or experimental manipulation of the episodic buffer within the dual-task literature for arithmetic (Ketelsen & Welsh, 2010).

The relations between working memory and arithmetic performance may also differ by participant age. The solving of simple arithmetic problems is often highly practiced in adults, but children solve problems more slowly and often use less sophisticated strategies (e.g., counting the smaller addend rather than retrieval) because they have yet to attain the same level of expertise (Ashcraft, 1992; Anderson, 1987, Siegler, 1988). Children tend to use more efficient strategies to solve arithmetic problems and rely less on working memory resources as they get older (Imbo & Vandierendonck, 2007b; McKenzie, Bull, & Gray, 2003, likely due to the acquisition of more efficient arithmetic strategies; Halford, Cowan, & Andrews, 2007).

Working memory's role in arithmetic processing has been studied using several different methodological approaches. We review these approaches, findings, and the costs and benefits associated with each approach. Here, we will examine correlational studies and two experimental approaches: working memory training and dual-task designs.

Correlations

Much of our understanding of the role of executive functions, specifically working memory, in mathematical and arithmetic performance comes from correlational designs. In general, all facets of working memory are known to be correlated with mathematical performance (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Bull & Lee, 2014), and arithmetic performance differences between children with and without mathematical difficulties are smaller after statistically controlling for differences in working memory capacity

(Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007). However, while these correlational designs provide insight into the possible cognitive processes influencing mathematical development, these designs have had inconsistent success at demonstrating the specificity of working memory contributions. In one meta-analysis, the average correlations for working memory and mathematical tasks were quite similar across the phonological and visuospatial working memory tasks (*r* = .34 for visuospatial updating, *r* =.38 for verbal updating, *r* =.34 for VSSP, and *r* =.31 for PL; Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013)[1]. Using a domain-specific model of working memory in which the domains are divided into verbal (mathematics tasks with verbal components like word problems), numerical (number related tasks like calculations), and visuospatial (mathematics tasks with visuospatial components like geometry) working memory, another meta-analysis of correlations between working memory and arithmetic tasks presented similar findings (*r* = .30 for verbal working memory, *r* =.34 for numerical working memory, and *r* =.31 for visuospatial; Peng, Barnes, Namkung, & Sun, 2015). Thus, the correlational literature has not identified substantial differences in the correlations between arithmetic performance and different working memory task types.

These findings imply that if there is specificity in the effects of different working memory components on arithmetic tasks, averaging correlations across studies does not reliably differentiate among them. Statistically controlling for facets of working memory simultaneously may not solve this problem, as the tasks used to measure each facet may differentially reflect

[1] Friso-van den Bos and colleagues (2013) used working memory components that were a combination of those posited by Baddeley and Hitch (1974 – the central executive, phonological loop, and visuospatial sketchpad) and Miyake et al. (2000 – updating, shifting, and inhibition). In our meta-analysis, we primarily use Baddeley and Hitch's model. We concluded this model was more closely aligned with the dual task literature, because it more naturally makes a distinction between domains (verbal and spatial), it makes a distinction between remembering and manipulating information that applies to many of the working memory tasks used in the dual task literature, and because almost all of the secondary tasks would be categorized as updating under Miyake and colleagues' model.

broader cognitive abilities (Schmidt, 2017). Taken together, this evidence is consistent with different aspects of working memory contributing similarly to arithmetic performance, but this method may also lack the required sensitivity to differentiate the contributions of different aspects of working memory to arithmetic performance.

Interventions

In order to better understand the unique contributions of working memory facets to arithmetic performance, we turn to evidence from experimental designs. One experimental approach to estimating the effects of working memory on mathematical cognition is cognitive training intervention. Typically, participants engage in an activity or game that targets either general or specific cognitive skills and are later measured on both cognitive abilities and school-related achievement tests (Diamond & Lee, 2011; Jaeggi, Buschkuehl, Jonides, & Shah, 2012; Loosli, Buschkuehl, Perrig, & Jaeggi, 2012). For example, a recent study by Ramani et al. (2017) trained three different groups of kindergarteners using a working memory game condition, math game condition, and a no-contact control condition and found improvements in numerical processing for both intervention groups. In this particular study, the working memory game involved remembering the orientation and sequence of cartoon characters on a tablet screen. The number of characters required of the children to remember would increase with successful responses.

This approach has produced a number of positive effects, and its design appears to be a straightforward approach for estimating causal links between working memory and arithmetic performance. However, a meta-analysis of training interventions by Melby-Lervåg and Hulme (2013) found short-term, near-transfer improvements in working memory ability, but inconsistent effects on substantively different cognitive tasks, including arithmetic. Subsequent meta-analyses and a systematic review reported similar findings of strong near transfer effects of cognitive training, with limited evidence that these improvements transfer to a variety of cognitive tasks, including verbal ability, reading comprehension, and arithmetic (Sala & Gobet, 2017; Simons et al., 2016; Melby-Lervåg, Redick, & Hulme, 2016). Notably, an evaluation of a school-based working memory training intervention in Australian first graders found persistent impacts on some working memory tasks at 6 and 12 months after training, but no evidence of transfer to math computation 12 or 24 months after training (Roberts et al., 2016).

Unfortunately, such findings are ambiguous with respect to the influence of working memory on arithmetic performance, primarily because of debates about the breadth of the constructs being trained. Though WM training is expected to transfer across WM tasks, this often may not be the case (Colom et al. 2013). If training fails to generalize to even other WM tasks, it is not clear if general cognitive processes like WM have been improved at all, and by extension, the mechanism for transfer is not clear either. If training generalizes to arithmetic or other tasks, the possibility of effects via mechanisms other than working memory improvement (especially in studies including a passive control condition; Shipstead, Redick, & Engle, 2012) also calls into question simple interpretations of such effects as evidence for effects of working memory on arithmetic performance.

Dual-Task Studies

Dual-task studies were developed and primarily used to investigate the role of WM and its components. In arithmetic cognition, dual-task experiments provide compelling evidence for some of the cognitive processes involved in a task, such as mental arithmetic, because they can be used to manipulate the roles of the different components of working memory across different arithmetic tasks (Logie & Baddeley, 1987). A dual-task design involves the completion of a primary cognitive task (in this case mental arithmetic) while simultaneously completing a secondary distractor task (in this case working memory tasks). Participants' accuracy and reaction time in a single-task condition are then compared to performance in various dual-task conditions.

Dual-task experiments vary on the types of tasks required of the participant (for review, see Pashler, 1994). For example, a participant may be required to remember a string of letters (z, h, d) while simultaneously completing an addition task. The interference effect of a concurrent memory load on the speeded task is attributed to either a decrease in shared resource capacity within WM or more likely, rehearsing the memory load causes interference in the stimulus-response mapping or preparation of the speeded task delays the retrieval process (Logan, 1978; 1979). These designs generally produce small or null effects on performance across different task modalities (Baddeley, 1986).

However, some studies have used secondary tasks with greater cognitive demands by instructing participants to randomly generate letters (Vreugdenburg, Bryan, & Kemps, 2003; Lemaire, 1996). As most commonly seen with central executive tasks, some designs have participants perform perceptual judgments concurrently with an arithmetic task. For example, a participant may perform an addition verification task (e.g., 5 + 6 = 11, indicating whether this is correct or not) while completing a task designed to load the central executive, such as pressing either a 1 or 2 key depending on whether they hear a high or low tone through a headset (e.g. Imbo & Vandierendonck, 2007a; Tronsky, McManus, & Anderson, 2008). With regards to dual-task interference in arithmetic, adult participants may rely on a small number of strategies when performing arithmetic, with direct retrieval being the most common. Cognitive load induced by dual-task experiments may interfere with the efficiency of using such strategies in calculations, especially those that require more steps (e.g. decomposition) or that rely on specific resources (Anderson, Reder, & Lebiere, 1996; Tronsky, 2005).

Two competing general theories have been posited to explain why these interference effects occur in dual-task studies: serial processing and parallel processing. In serial processing, people are believed to have some sort of structural limitation or a central bottleneck, whereby the cognitive demands of the first cognitive process delay performance on the second (Ruthruff, Pashler, & Klaasen, 2001). Ruthruff, Pashler, and Klaasen (2001) and Ruthruff, Pashler, and Hazeltine (2003) provided experimental evidence that a structural limitation, rather than a voluntary postponement, underlies slowed performance in dual-task studies. Assuming there is a central bottleneck in processing, the memory load imposed by these secondary tasks (especially difficult ones) are likely due to interference in the preparation of the arithmetic task rather than the actual processing of arithmetic regardless of the modality.

Under the alternative graded capacity sharing model, cognitive processes are thought to be performed in parallel, but interference effects occur due to capacity limitations rather than with a bottleneck (Ruthruff, Pashler, & Hazeltine, 2003). A prediction specific to this model is referred to as crosstalk, wherein if memory's limited capacity relies on different cognitive processes (e.g. visual and verbal processes), then completing two similar or within-modality tasks leads to a greater decrement in performance than completing two dissimilar tasks (Lien & Proctor, 2002; Miller, 2006; & Koch, 2009; Navon & Miller, 1987; Pashler, 1994). Crosstalk is a recurrent hypothesis in dual-task studies of arithmetic and working memory. For example, Lee and Kang (2002) hypothesized that arithmetic operations such as multiplication and subtraction relied on two separate encoding processes – verbal for multiplication and visuospatial for subtraction. Multiplication was more impaired by a verbal secondary task, while subtraction was more impaired by a spatial secondary task, consistent with the hypothesis that separate facets of working memory were required for different arithmetic tasks. While individuals may be able to encode stimuli from similar tasks simultaneously, crosstalk designs would predict that the processing of one of these tasks is harmful to the processing of the other (Treisman & Davies, 1973).

Although crosstalk is an influential hypothesis, specific findings have not always been reliably replicated. For example, the differential interference of PL and VSSP secondary tasks on multiplication and subtraction in the Lee and Kang (2002) study was partially replicated in another study in a Chinese, but not a Canadian sample (Imbo & LeFevre, 2010). Some of the apparently discrepant findings across studies may be related to confounding of working memory task type with the cognitive demands of the working memory task: A recent study found no selective interaction of working memory load type (PL and VSSP) with arithmetic conditions (subtraction and multiplication) once the task demands for arithmetic and working memory were matched on problem/set size and difficulty (Cavdaroglu & Knops, 2017). These findings do not falsify the crosstalk hypothesis, but they suggest that any such effects may be difficult to generalize across individuals and tasks. They also raise an alternative hypothesis: that the effects of hypothesized specific overlapping task demands on delays in performance may be small, especially in comparison to the general task demands inherent to the WM distractors themselves. Altogether, it is unclear exactly how specific mental arithmetic is in recruiting working memory resources among these dual-task designs.

Current Study

The purpose of the current study is to test theories of arithmetic cognition and dual-task performance by meta-analyzing a body of research that has used the dual-task paradigm to study the underlying cognitive processes involved in arithmetic. The aim of this meta-analysis is threefold. The first aim is to address the robustness of arithmetic performance’s reliance on working memory resources as predicted by dual-task studies by means of meta-analysis. The second aim is to address how the effects of working memory load on arithmetic performance might depend on factors typically manipulated in these dual-task studies: overlapping features characterized by the types of working memory load and arithmetic operations, task complexity characterized by the type of working memory task, level of expertise in arithmetic, as approximated by the age of participants, and author’s prediction about significant effects of WM load types as a proxy for the crosstalk hypothesis. The last aim is to in some way reconcile the discrepancy between previously reported correlational and experimental findings.

Methods

Inclusion criteria and screening

A flow-chart of the identification and screening process can be found in Figure 1. Keyword searches were used in Google Scholar and ProQuest (databases used were ERIC (1966-Current) and PsychINFO (1806-Current)) to obtain the sample of studies to be screened for this meta-analysis. The following search terms used included: ("working memory" OR WM OR "executive function" OR cognition OR visuospatial OR visu* OR spatial OR "phonological loop" OR "verbal" OR "slave system") AND (mathematics OR math OR arithmetic) AND ("dual task" OR paradigm OR interference OR suppression). In total, 1071 results were found from ProQuest and 5000 results were searched through Google Scholar. After removing duplicates, we were left with a total of n=4119 records. We then proceeded with pre-screening of titles and abstracts for relevance, which brought the number of records down to n=850 after title screening and then n=55 after abstract screening.

Studies were considered eligible for this meta-analysis if they met the following criteria. First, studies had to be of a dual-task design such as those described in Pashler (1994) and Ashcraft et al. (1992). Second, a primary arithmetic task must have been performed concurrently with a secondary working memory or cognitive load task such that accuracy or reaction time (RT) was measured. Third, the working memory or cognitive load task needed to be experimentally manipulated in either a within or randomized between subjects design, such that the same or comparable individuals also completed the math task under no (or different) cognitive load. Experimental manipulation of load allows for comparisons on arithmetic tasks with either participants' baseline performance or the performance of a randomized control condition. Fourth, to be included in the database, studies had to report sufficient statistical information to enable the computation of an unstandardized effect size (i.e., group RT means). When information was not directly presented in a published manuscript, we asked authors whose papers were published within the last ten years for the data (n=5) – of which we received data from 2. Fifth, studies must only include participants who are children above the age of 5 or adults below the age of 65. Dual-task literature generally excludes pre-school-aged children as well as adults over the age of 65, because of the difficulty in obtaining reliable estimates within these age groups (most data are collected from undergraduate or middle-adult samples). Lastly, we excluded studies that used only participants who had been identified as having a learning disability or special need prior to participating in the study.

The literature search yielded 55 records after abstract screening. Of these, 20 were excluded during the full text screening, because cognitive load was not experimentally manipulated. During final assessment for eligibility, 5 were excluded because studies included participants with either a math learning disability or autism spectrum disorder. Three studies were excluded because the primary task was a working memory task rather than an arithmetic task. Six studies were excluded due to lack of available reaction time data. In total, 21 eligible papers containing 400 effect sizes from 51 unique samples, obtained from 1,049 individuals, were included in these analyses.

Coding

The following variables were coded for the dataset. Study characteristics included: (a) whether the study design was within or between subjects, (b) a unique identifier for the experiment number, because some studies included multiple experiments with different samples or more than two conditions within the same experiment, (c) descriptions of the arithmetic and working memory tasks, (d) the arithmetic task type (i.e. exact addition, approximate addition, exact subtraction, exact multiplication, addition verification, and multiplication verification), (e) the number of items per condition, (f) the two conditions being compared (e.g. central executive cognitive load vs. control), (g) whether authors made a prediction about the significance of a certain condition, and (h) whether strategies were reported. Sample characteristics included (i) the mean age of the sample and (j) the type of sample (participants between the ages of 4 and 17 were considered children and participants between the ages of 18 and 65 were considered adults). Coding for the WM distractor tasks can be found in the online supplemental materials, Table S3. Statistical characteristics included: (l) control and experimental sample sizes and (m) RT means and standard deviations to calculate effect sizes. While accuracy data were coded, they were not included in effect size analyses for 2 reasons: (1) mean accuracies were very high (around .90, as reported in Table 1) and (2) some studies (n=4) did not report accuracy data by condition, but every study included RT means. Analyses lacked sufficient statistical power to detect statistical interactions between every arithmetic type and every WM load type, so we used author's predictions as a proxy to measure the predicted non-additive effects of WM distractor and arithmetic types. Authors' predictions were obtained by coding each of the 21 studies' introduction and methods sections for statements hypothesizing the effect of WM distractor on arithmetic performance. For example, "Given these arguments, we predict that the CRT-R task will interfere more with both non-retrieval-based and retrieval-based subtraction problem solving than the SRT-R task because of its response selection component" (Tronsky, McManus, & Anderson, 2008, p. 194). These predictions were then dichotomously coded as having hypothesized effect of WM load or not. The passages justifying each coded prediction are included in the online supplementary materials, Table S2.

Data Imputation and Transformations

Reaction time is a commonly used dependent variable used to draw inferences about cognitive processing, but studies often do not account for residual processes involved with completing RT tasks, such as encoding the item and producing a response (Rouder, 2005). For example, Geary, Widaman, and Little (1986) estimated that encoding single digits in complex addition and multiplication problems required approximately 170 ms. To try to account for such residual processes, 200 ms was subtracted from all RT means prior to analyses. Missing standard deviations for reaction times that were in studies older than ten years or those which we could not obtain from the authors were imputed by regressing RT standard deviations on RT means and the square of RT means. Imputing the missing standard deviations was deemed appropriate because of the high predictability of standard deviations from a two predictor model:

$$SD_t = b_0 + b_1Mean_t + b_2Mean_t^2 + e_t$$

$$SD_c = b_0 + b_1Mean_c + b_2Mean_c^2 + e_c$$

In these models, the subscript *t* indicates values from the experimental conditions across study, and the subscript *c* indicates values from the control conditions across studies. The regression equations for standard deviation used a second order polynomial term to account for the non-linear association between means and standard deviations across effects. The models explained the vast majority of variance in both the control and experimental conditions (control: $r^2 = .92$; experimental: $r^2 = .84$). Scatterplots with best fitting lines and polynomials appear in the online supplementary materials, Figure S1. In all, 168 (84 in the experimental conditions and 84 in the control conditions) standard deviations (or 21%) were missing from the data. Only those standard deviations with an available mean RT (n = 164 standard deviations: 82 in the experimental conditions and 82 in the control conditions) were imputed, while the remaining 4 standard deviations were left missing. Imputation led to a small number of predicted negative values of experimental standard deviations (n = 6). We set these values to the minimum observed (i.e., non-imputed) standard deviation of experimental conditions, which was 58.19 ms. As explained below, the meta-analysis was re-run with non-imputed values only to test the robustness of our main findings to these data processing decisions. Initial histograms revealed the RT data to be positively skewed (Figure S2). Further, there was a substantial relation between RT means and RT experimental effects (b = .090, se = .023, p < .001; Figure S5, left panel). In other words, secondary tasks resulted in more absolute slowing for arithmetic tasks that took longer to complete. In the dual-task arithmetic literature, we found little discussion of the functional relation between arithmetic speed and the presence of a secondary task. Specifically, did the secondary task slow each trial by a short but constant amount of time for participants to encode and rehearse the stimulus, or did the secondary task slow the rate of mental arithmetic proportionally, such that more difficult problems would be more slowed than easier problems? In the former case, one parameter of interest is a constant, which represents the average amount of time that each trial is slowed. However, if arithmetic slowing is hypothesized to be proportional to the amount of time the arithmetic task takes to perform in the absence of a secondary task, a more useful parameter of interest is the percentage by which the average trial is slowed in the presence of a secondary task. Based on our reading of the literature, authors seem to have the latter case in mind, whereby keeping information from the secondary task in working memory slows performance on the arithmetic task. Therefore, we used a logarithmic transformation in our main specification to better capture the underlying cognitive effects of secondary tasks. In addition, we also report an analysis of untransformed RT means in an alternative specification (Table S1, No Log). We performed logarithmic transformations of the RT data using Method 1 from Higgins, White, and Azures-Cabrera (2008). This method assumes log-normal distributions with different standard deviations. The approximate transformations were then converted to log base 10 for interpretability in our analyses by dividing these means and SD by the constant, ln(10). Following the transformation, we again checked the distributions of RT means, which were substantially less skewed (Figure S4). Average RTs and effect sizes were no longer substantially associated (*b* = .020, se = .019, *p* = .30; Figure S5, right panel).

Effect Size Calculation

Following Method 1 from Higgins and colleagues (2008), we computed the raw and log mean differences in the experimental and control conditions. The equation for calculating the standard error of these differences is shown below for between-subjects designs:

$$\text{Between: } SE_{diff} = \sqrt{}(SD_c^2/n_c + SD_t^2/n_t)$$

However, 17 out of the 21 eligible studies were conducted within participants, requiring a measure of covariance between control and experimental scores to compute the standard error of these effects. Because this information was not directly reported in most studies, we computed within-subjects standard errors under three different assumptions of correlated performance across conditions: *r* = .2, .5, and .8. These correlations were then each multiplied with the control and experimental RT standard deviations to create 3 measures of covariance, using the following equation:

$$\text{Within: } SE_{diff} = \sqrt{}(SD_c^2/n + SD_t^2/n - 2*cov_{tc}/n)$$

Analyses

The metafor package in R was used to conduct this meta-analysis (Viechtbauer, 2010). A multilevel random-effects meta-analysis of these data was used to estimate and account for the amount of heterogeneity between papers and between different samples included in the same paper. Effect sizes were modeled as nested within samples, which were nested within papers across all of these specifications. We performed several sensitivity analyses, including an alternative estimation strategy of robust variance estimation adjustment of the standard errors using the robumeta package in R (Park & Beretvas, 2018; Fisher & Tipton, 2015). All specifications included some of the same characteristics: Nine specifications (1 main, 3 alternatives, 1 PEESE (precision-effect estimate with standard errors) adjustment, 4 robustness checks) were used in this analyses and are as follows: (1) Main=assumed within subject correlation=.5, (2) Alt 1=assumed within subject correlation=.2, (3) Alt 2=assumed within subject correlation=.8, (4) Alt 3=assumed within subject correlation=.5 and RT data restricted to ≤ 5000 ms. Four other alternative specifications used for sensitivity analyses included: (6) Nonimputed=analyses without SD imputations, (7) RVE=standard errors adjusted using robust variance estimation with small sample correction, (8) Adult=analyses included only studies with adult participants, (9) No Log=same as Main model but without log transformations, and (10)

PEESE=PEESE adjusted model, assumed within subject correlation=.5.. Specifications Main, Alt 1, and Alt 2 reflected the three assumed correlations used to calculate the standard errors for within-subject studies. Alt 3 used a subset of the RT data, such that only RTs ≤ 5000 ms were included for analyses, because several child effect sizes had much larger RT means (Figure S4), and because of the possibility that cognitive processing lasting longer than 5000 ms could plausibly be differentially affected by a secondary task. We ran the latter 4 alternatives specifications using the assumed correlation structure of the Main model as robustness checks. The Nonimputed models used data without the SD imputations to examine any potential bias from these estimates. Robust variance estimation (RVE) models were used as an alternative to multilevel modeling to account for the non-independence of observations within samples and papers. Finally, Adult models that only contained adult participants were included, because of the small sample of child effect sizes (n=36) and some effect sizes with unreported participant age (n=7).

To address the second aim of this meta-analysis, 5 potential moderators of the overall effect were tested to determine differences between subgroups of samples. The first moderator examined was participant age (adult vs. child), as a proxy for expertise in arithmetic. The second moderator was working memory load type: central executive, verbal, visuospatial, or spatial. The third moderator was arithmetic problem type: addition verification, exact addition, approximate addition, exact multiplication, exact subtraction, or multiplication verification. The last moderator was the authors’ predictions of whether load had significant effects on arithmetic performance (given the very large number of possible interactions between arithmetic problem type and working memory load type, this was the method we chose for examining the strength of evidence for crosstalk). Although hypothesized cases of crosstalk make more complex predictions, such as interactions among secondary task modality (e.g., visual vs. verbal), type of arithmetic (e.g., multiplication vs. subtraction), and presentation format (e.g., vertical vs. horizontal), a complex set of such instances was predicted by authors corresponding to a large number of parameters to test. Thus, we chose a broader definition of crosstalk captured through the authors' predictions instead. The moderators were examined across the nine specifications.

Results

Publication Bias

A funnel plot of the distribution of effect sizes was used as a visual aid to detect publication bias (Figure 2) (Egger et al. 1997). Effect estimates from studies are plotted against a precision measure from those studies (e.g. standard error). Estimates of effects from smaller studies are more variable than those from larger studies leading to larger amount of scatter towards the base of the plot. In the absence of bias, a symmetrical funnel shape is observed. However, asymmetrical distribution of points around the average RT effect indicate possible publication bias. The main specification (assuming a within subject correlation= .5) was used. The effect estimates (log difference in RT means) were plotted along the x-axis while the standard errors of the effect estimates were plotted along the y-axis. The vertical line in the middle of Figure 2 represents the location of the estimated effect of a working memory task on performance (*b* = .074). Examination of the dispersion of effect sizes in the funnel plot revealed some asymmetry – specifically, for smaller studies – suggesting the possibility of some publication bias.

The PEESE test is a meta-analytic approach to detecting publication bias using metaregression. PEESE (precision effect estimate of standard error) uses a weighted-least-squares regression model where the variance (squared standard errors) of each sample is used to predict the distribution of effect sizes (Stanley & Doucouliagos, 2014). Assuming a true effect, publication bias is stronger for studies when the standard error is large and weaker when the standard error is small. The PEESE model revealed a smaller, non-significant effect of a working memory task on performance (b_1 = .031 se = .017, p = .070), and a moderate degree of publication bias (QM = 24.11, p < .001) (Table S1, PEESE).

Summary Effect Size

The main model specification (assumed within subject correlation=.5) produced a summary effect of b = .074, z = 5.21, p < .001 across 400 comparisons, which suggests a .074 difference in log RT. An effect of this size corresponds to an 18.7%[2] slowing of performance for participants in experimental conditions where they are performing dual-tasks compared to control conditions where arithmetic was tested by itself. This effect remained stable across the six other specifications including the Nonimputed, RVE, and Adult models (Table 2). However, as described above the PEESE adjustment decreased the estimated effect to b = .031, z = 1.81, p = .070, which equates to a non-significant 7.3% decrease in speed. However, this model obscures important heterogeneity in the estimates, which was explored in the following moderation analyses.

Moderators

Participant age, working memory task type, arithmetic task type, and authors' prediction of whether load would be significant were entered separately as moderators in different models (Table 3). Results of sensitivity analyses for moderators can be found in the online supplementary materials (see Table S1). Across most specifications, working memory task type, arithmetic task type, and authors' predictions statistically moderated differences in log reaction

[2] 18.7% and subsequent percentage effects were calculated by taking 10^b where b (e.g. .074) is the coefficient estimate.

times, suggesting that differences in RT performance are dependent on the type of working memory load (main specification: Q(2) = 796.04, *p* < .001), the type of arithmetic problem (main specification: Q(5) = 12.11, *p* = .033), and authors' predictions (main specification: Q(1) = 265.14, *p* < .001). Compared to other moderators, working memory load type explained the most heterogeneity (22% of the total heterogeneity) across effect sizes. This finding was largely robust across specifications.

Furthermore, dual-tasks involving the central executive appear to have the most impact out of all load types (CE: *b* = .146, *z* = 12.35, *p* < .001, CI [0.123, 0.169]) (Table 3, Column 1, WM load type). Indeed, across all specifications working memory load reflected a strong effect of the central executive (a 40% decrease in speed) with much smaller effects for the other load types (Table 3, Column 1, WM load type). For example, verbal tasks generated an effect of *b* = -.115, *z* = -27.64, *p* < .001, CI [-0.123, -0.107] in relation to the intercept (CE); meaning an effect of verbal distractors on performance of .031 or a 7.4% decrease in speed for dual-tasks with verbal distractors (Table 3, Column 1, WM load type). The strong effect of CE was robust, showing similar estimates across the sensitivity analyses.

Of the various arithmetic problem types, addition and multiplication verification tasks had consistent effects on arithmetic performance across most specifications (reference task, addition verification: *b* = .077, *z* = 2.53, *p* = .011, CI [0.018, 0.137], with more slowing for multiplication verification: *b* = .046, *z* = 2.99, *p* = .003, CI[0.016, 0.076]) (Table 3, Column 1, Arithmetic problem type)[3]. On average, participants performing multiplication verification tasks were slowed 13.3% more than participants performing addition verification tasks. Although

[3] However, it should be noted that both verification tasks were statistically significant only after they were log-transformed (Table S1, Column 6, Arithmetic problem type).

these differences were smaller than the differences among working memory load types, the effect of arithmetic problem type was robust to the inclusion of controls for working memory load type (addition verification: *b* = .131, *z* = 5.50, *p* < .001, CI [0.084, 0.178], multiplication verification: *b* = .044, *z* = 2.99, *p* = .003, CI [0.015, 0.074]; Table 3, Column 1, WM load + arithtype).

Another metaregression model was analyzed to test whether differences in RT performance were driven by authors' predictions for significance, which measured the predicted non-additive effects of WM distractor and arithmetic types. This analysis revealed a highly significant effect of authors' predictions across most specifications (*b* = .072, *z* = 16.28, *p* < .001, CI [.063, .081], Q(1) = 265.14; Table 3, Column 1, sig predict). On average, authors' predictions predicted an 18.0% decrease in performance between experimental and control conditions in dual-task studies. To further test whether predicted effects were driven by main effects of WM load and arithmetic task types, another metaregression model was estimated using authors' predictions while controlling for WM load and arithmetic task type. These analyses revealed a smaller but usually still significant effect of authors' prediction of significance controlling for WM load type and arithmetic task type (*b* = .023, *z* = 4.58, *p* < .001, Q(8) = 820.41, p < .001; Table 3, Column 1, sig predict + WM + arithtype). Expressed differently, the authors' predictions were associated with a 18.0% decrease in RT performance, but this dropped to 5.3% after controlling for WM and arithmetic task types, suggesting that authors' predictions about the effect of load likely contribute somewhat to slower arithmetic performance, but this is somewhat confounded with the type of working memory and arithmetic tasks. Importantly, the effect of *unpredicted* effects of CE distractors on addition verification tasks (*b* = .107, *z* = 4.33, *p* < .001, CI [.058, .156]) was still significant and larger than the remaining effect of authors' predictions with a 27.9% slowing in arithmetic performance (Table 3, Column 1, sig predict + WM+ arith type).

Discussion

We conducted a meta-analysis for the purpose of assessing how robust the speed of solving arithmetic problems is affected by changes in WM resources as predicted by dual-task studies. Consistent with our predictions, we found strong evidence for the influence of performing a secondary WM task on arithmetic performance. The main effect of WM distractors on arithmetic performance in dual-task studies was robust across all 9 of our specifications. These findings suggest that arithmetic performance relies heavily on WM, specifically central executive resources.

Among the moderators, working memory load type was the most substantial moderator of performance decrements, followed by the type of arithmetic operation, and authors' predictions. Tasks taxing the central executive specifically incurred greater decrements in performance than any other WM load type, indicating the importance of considering the general cognitive complexity of the secondary task when predicting its influence on primary task performance. This finding was consistent with previous literature citing the overall importance of the central executive/executive functions to working memory (Engle, 2002; Engle & Kane, 2004). Assuming the central bottleneck theory to be correct, these more difficult cognitive loads may not be competing for shared WM resources but rather there is delay in preparation or switch to arithmetic processing. In comparison, the PL and VSSP tasks showed much smaller impacts than CE tasks. Consequently, the parallel processing theory is not entirely ruled out, as the impact of PL and VSSP tasks would imply processing interference within-modality. Of course, the larger effects of CE tasks may also be indicative of a cognitive bottleneck. The null effect of age was somewhat surprising given its prevalence in the literature. However, it should be noted that the number of child participants was quite small compared to adults (see Table 1). Furthermore, both distractor and arithmetic tasks did vary in the number of observations across our models which led some categories to have higher standard errors (see Table 3), so these estimates are less precise. Variability in precision across estimates would imply that greater scrutiny is required for the categories with fewer observations. For example, VSSP secondary tasks had been used in fewer studies, and therefore estimates are less precise for this category. However, despite a higher standard error in our analyses, the VSSP load estimate was still 2 standard errors below the estimate for CE load (Table 3, Column 1, WM load type).

Importantly, this meta-analysis's finding that CE tasks have a much greater impact on arithmetic processing than other WM load types appears to contrast with findings from prior meta-analyses based on correlations between working memory and math tasks. These prior meta-analyses reported very similar correlations across working memory facets (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Peng, Barnes, Namkung, & Sun, 2015).A possible explanation for this discrepancy is that factors common to working memory tasks, such as maintaining a constant memory of a single element (number or letter) or an element's position in space (carrying values or grids) inflate the correlations between arithmetic performance and visuospatial and phonological working memory. It may also be possible that the specific encoding behind similar modality tasks, especially subcomponent WM tasks, are being suppressed by their relation to the central executive. For example, performance on a subtraction task following a VSSP matrix task may be impacted by need to switch between tasks in addition to requiring similar resources. Dual-task measures can be viewed as forms of inhibition or shifting tasks, which may indicate that the CE is being used in addition to the PL or VSSP. Future work might apply methods that attempt to make the magnitudes of effects from correlational and experimental studies more directly comparable by using regression-adjusted estimates and intervention effect sizes (e.g., Bailey, Duncan, Watts, Clements, & Sarama, 2018) to studies of working memory and arithmetic to test this prediction directly.

We attempted to estimate crosstalk or the non-additive effects of WM distractors on arithmetic performance by coding ' predictions of the effect of specific WM secondary tasks on arithmetic performance. The effects of these predictions were nonzero, but after accounting for the main effects of WM and arithmetic load type, the author-predicted effects were not substantially larger than the non-predicted effects. Even these may be somewhat over-estimated, because almost all studies were conducted before study preregistration was encouraged in psychology. Thus, it is possible that these predictions could have been changed over the course of data-collection or through pilot testing, inflating the estimated effect of crosstalk when operationalized as authors' predictions. We realize that research teams with more specific hypotheses about crosstalk may prefer to code predicted effects in a different way. Our hope is that, by publishing the meta-analytic database, this will allow others to test these hypotheses in future work. Taken together, while having to process similar modalities in mental arithmetic may explain some of the underlying processes behind mental arithmetic, general structural limitations brought on by the general demands of each distracting task may deserve additional scrutiny, relative to the crosstalk effects often predicted in this literature.

Notably, our preliminary analyses indicated that the detrimental effects of secondary WM tasks on performance were strongly associated with the average RT on the arithmetic task. In other words, RT in arithmetic tasks increased more under more difficult secondary WM tasks. This finding mirrors Salthouse's (1988) findings that aging-related performance deficits are higher on more complex cognitive tasks. Our findings also provide convergent evidence for

Salthouse's theory that changes in general cognitive resources exert the largest effects on complex cognitive tasks. Specifically, in this meta-analysis, for participants of the same age taking the same tests on the same day, the proportionality of WM demands is larger for more complex arithmetic tasks when using an experimental research design.

Limitations & Future Directions

One of the key limitations was that our analyses lacked sufficient statistical power to detect smaller effects of the interactions between arithmetic and WM load types, prompting us to use authors' predictions as a proxy for crosstalk. This may hide important heterogeneity in crosstalk effects, with some being substantial and others null. Further, as noted, a moderate degree of publication bias was present in the eligible studies; this may have inflated the apparent effect of authors' predictions.

As stated previously, accuracy data were not always available and mean accuracy scores were consistently high leading us to solely use reaction time data in our analyses. We recognize that there is a speed-accuracy trade-off meaning participants will sacrifice time in order to correctly answer questions or vice-versa. For example, Kalaman & LeFevre (2007) reported more errors in two-digit plus two-digit addition with carrying vs. no carrying but found no significant differences in speed. These findings suggest that such speed-accuracy trade-offs in dual-task studies could potentially reduce an estimated effect of working memory on mental arithmetic based on RT data alone.

We recognize that Baddeley's multicomponent model of WM is one of several models used to describe the relations between memory processes. This model is commonly used in the arithmetic dual-task literature; thus, we chose to use similar terminology as it more closely aligned with those designs. However, more recent research (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Christophel et al., 2017) as well as our own findings suggest that the use of this model in the dual-task literature lacks appropriate discussion of the central executive's role interacting with the more specialized domains. Moreover, other models of WM have proposed promising alternative perspectives on the role of WM and executive functioning (see Miyake et al., 2000; Diamond & Lee, 2011; Engle, 2002). These models were introduced previously and most if not all would point to the strong influence of general cognitive ability within EF or central executive functions driving dual-task performance rather than specific modalities highlighted in Baddeley's model. Indeed, most would argue that dual-tasks involve quickly switching between mental tasks. Results are also consistent with Engle's (2002) model which posits attention rather than capacity as the limiting factor in performance. This model would predict dual-task performance to be reliant on more domain-general processes rather than domain-specific components. Altogether, many other perspectives of EF and working memory are in line with our conclusions about the importance of central executive functions in mental arithmetic.

Most of the studies in our sample did not report data on participant strategy use. Aggregating reaction time data across different arithmetic strategies obscures information about the cognitive processes underlying performance (Siegler, 1987). Because both the frequency and efficiency of arithmetic strategy use are correlated with working memory capacity (Bailey, Littlefield, & Geary, 2012; Geary et al., 2004, 2007), understanding whether working memory distractors influence performance via strategy changes or slowing within strategies would be theoretically useful.

Because identical tasks were generally not used across age groups, our analyses may have been unable to detect a moderating effect of age, despite the strong support for this prediction in the literature. Larger dual-task studies with child participants will be required to test for hypothesized developmental differences in the nature of working memory effects on arithmetic performance (Meyer, Salimpoor, Wu, Geary, & Menon, 2010; for review, see Anderson, 1987).

Because cognitive load was not matched across primary arithmetic or secondary WM tasks types, the moderating effects of these task types may reflect some combination of the kinds of demands and the magnitudes of demands of different tasks. Based on descriptions of the different WM tasks, we suspect the CE tasks differed in their magnitudes of general cognitive demands, but future work that experimentally manipulates the magnitude of demands in the secondary tasks would be useful for testing the importance of this construct directly. For example, a recently used approach of systematically equating demands of WM and arithmetic tasks through the use of an adaptive psychophysical staircase to determine appropriate span sizes per individual (Cavdaroglu & Knops, 2017) is a useful model for separating the effects of crosstalk from general task demands.

Finally, like the previous correlational work on arithmetic and WM, the exact constructs being measured or manipulated in the dual-task literature are not wholly clear and warrant further attention. For example, it is not clear whether variations in working memory task type or complexity *within* individuals are qualitatively similar approximations of between-individual differences in working memory capacity. Fully reconciling these two literatures would require more precise models of the processes and parameters underlying both the correlations and dual-task effects.

Conclusion

Our meta-analysis indicates that the dual-task literature provides strong evidence that mental arithmetic relies on working memory resources. We hope that further work will attempt to build on these findings by both attempting to quantify the underlying domain-specific and domain-general effects of working memory on arithmetic and by building quantitative models that will reconcile the discrepancy between correlational and experimental literature.

Appendix

Table 1. Descriptive statistics of analysis variables

	N	Count
Age	400	
Adult		359
Child		36
NA[1]		7
Within vs. Between Subjects	400	
Within		328
Between		72
Working Memory task type	400	
CE		169
Verbal		178
VSSP		53
Arithmetic task type	400	
Add verification		83
Approximate addition		8
Exact addition		211
Exact multiplication		22
Exact subtraction		22
Mult verification		48
NA		6
	mean	sd
Sample Size	20.8	13.1
Accuracy		
Experimental	.89	.10
Control	.94	.05
RT		
Experimental	3014	2894
Control	2360	2657
RT SD		
Experimental	2038	2261
Control	1263	1577
Log RT		
Experimental	3.28	0.32
Control	3.19	0.31
Log RT SD		
Experimental	0.19	0.08
Control	0.17	0.06
Number of observations	N=400	

Note. N is number of effects. Frequencies calculated from non-missing RT data. [1] One study did not report age of participants

Table 2. Main effects by specifications

	k	b	10^b	Q	QE	QM
Main	400	.074***	1.187	3847.37***		
		(.014)				
Nonimputed	318	.073***	1.183	3586.29***		
		(.016)				
RVE	400	.091***	1.232			
		(.006)				
Adult	357	.077***	1.193	3425.04***		
		(.014)				
Alt1	400	.074***	1.187	2544.01***		
		(.014)				
Alt2	400	.074***	1.185	8667.91***		
		(.015)				
Alt3(subset)	339	.077***	1.193	2809.28***		
		(0.013)				
No Log	400	318.609***		2909.20***		
		(74.451)				
PEESE	400	.031	1.073		3606.48***	24.11***
		(0.017)				

Note. SE in parentheses. QE = test for residual heterogeneity. QM = test of moderator. Main: assumes within subject correlation = .5, Nonimputed: Without SD imputations, assumes within subject correlation= .5, RVE: Robust Variance Estimation for small samples (used with Main specification), Adult: subset of data that excludes child and no age data, Alt1: assumes within subject correlation = .2, Alt2: assumes within subject correlation = .8, Alt3: Subset of data where only RT's <=5000 ms are kept, PEESE: corrects for publication bias, assumes within subject correlation = 0.5. No Log: Analyses conducted without log transformation, assumes within subject correlation =.5 * p<.05 ** p<.01 *** p<.001

Table 3. Moderator analyses of working memory and arithmetic performance

	Main model		Alt 1		Alt 2		Alt 3	
	k/		k/		k/		k/	
	Qm	b	Qm	b	Qm	b	Qm	b
Age	393		393		393		339	
Intercept(adult)	0.38	.076***	0.34	.076***	0.51	.077***	2.89	.071***
		(.016)		(.016)		(.016)		(.012)
child		-.027		-.026		-.031		.085
		(.044)		(.044)		(.048)		(.050)
WM load type	400		400		400		339	
Intercept(ce)	796.04***	.146***	528.81***	.146***	1775.14***	.145***	586.04***	.140***
		(.012)		(.012)		(.012)		(.012)
verbal		-.115***		-.116***		-.113***		-.102***
		(.004)		(.005)		(.003)		(.004)
vssp		-.128***		-.129***		-.127***		-.122***
		(.007)		(0.008)		(.004)		(.007)
Arithmetic problem type	394		394		394		333	
Intercept(add verification)	12.11*	.077*	11.31*	.077*	14.61*	.077*	13.18*	.074**
		(.031)		(.030)		(.031)		(.023)
approx addition		-.041		-.038		-.049		N/A
		(.080)		(.078)		(.086)		N/A
exact addition		-.001		-.001		-.002		-.012
		(.037)		(.036)		(.037)		(.029)
exact multiplication		-.027		-.028		-.026		-.038
		(.043)		(.043)		(.044)		(.034)
exact subtraction		-.010		-.010		-.011		-.022
		(.043)		(.042)		(.044)		(.034)
multiplication verification		.046**		.046***		.046***		.045**
		(.015)		(.015)		(.015)		(.015)
WM+arith type	394		394		394		333	
Intercept(ce+add ver)	802.31***	.131***	533.95***	.131***	1783.45***	.130***	592.49***	.125***
		(.024)		(.024)		(.024)		(.023)
verbal		-.115***		-.116***		-.113***		-.102***
		(.004)		(.005)		(.003)		(.004)
vssp		-.129***		-.129***		-.128***		-.125***
		(.007)		(.009)		(.005)		(.008)
approx addition		.013		.016		.005		N/A
		(.053)		(.051)		(.059)		N/A
exact addition		.024		.026		.022		.029
		(.029)		(.029)		(.029)		(.029)
exact multiplication		-.010		-.013		-.007		-.012
		(.033)		(.033)		(.033)		(.032)
exact subtraction		.006		.004		.008		.004

		(.032)		(.032)		(.033)		(.031)
multiplication verification		.044**		.044**		.046**		.044**
		(.015)		(.015)		(.015)		(.015)
Sig Predict	400		400		400		339	
Intercept(no)	265.14***	.021	170.10***	.021	611.36***	.022	152.63***	.013*
		(.014)		(.014)		(.014)		(.013)
yes		.072***		.072***		.071***		.061***
		(.004)		(.006)		(.003)		(.005)
Sig + WM + arith type	394		394		394		333	
Intercept (no+ce+add ver)	820.41***	.107***	542.70***	.108***	1838.28***	.104***	596.10***	.112***
		(.025)		(.025)		(.025)		(.024)
sig(yes)		.023***		.021***		.024***		.012*
		(.005)		(.006)		(.003)		(.005)
verbal		-.106***		-.108***		-.104***		-.097***
		(.005)		(.006)		(.003)		(.005)
vssp		-.117***		-.118***		-.114***		-.118***
		(.008)		(.009)		(.005)		(.008)
approx addition		.013		.017		.006		N/A
		(.056)		(.053)		(.062)		N/A
exact addition		.026		.028		.024		.030
		(.029)		(.029)		(.030)		(.029)
exact multiplication		-.007		-.010		-.004		-.010
		(.033)		(.033)		(.034)		(.032)
exact subtraction		.009		.007		.011		.005
		(.033)		(.033)		(.034)		(.031)
multiplication verification		.045**		.044**		.046**		.044**
		(.015)		(.015)		(.015)		(.015)

Note. Model names are bolded. The variable names following each model name are the moderators included in that model. The first row of each model represents the estimated effect at the reference group, which is given in parentheses after Intercept, for each model. Standard errors are in parentheses. k = number of effect sizes included in the model; Qm = Q test for heterogeneity explained by the predictors (total heterogeneity for each specification appears in Table 2), ce=central executive, vssp=visuospatial sketchpad, approx=approximate, add ver=addition verification, sig(yes)=author made a prediction on significance. Main: assumes within subject correlation = 0.5; Alt1: assumes within subject correlation = 0.2; Alt2: assumes within subject correlation = 0.8; Alt3: Subset of data where only RT's <=5000ms are kept, assumes within subject correlation=.5

Figure 1. Flowchart of Literature Search and Screening Process.

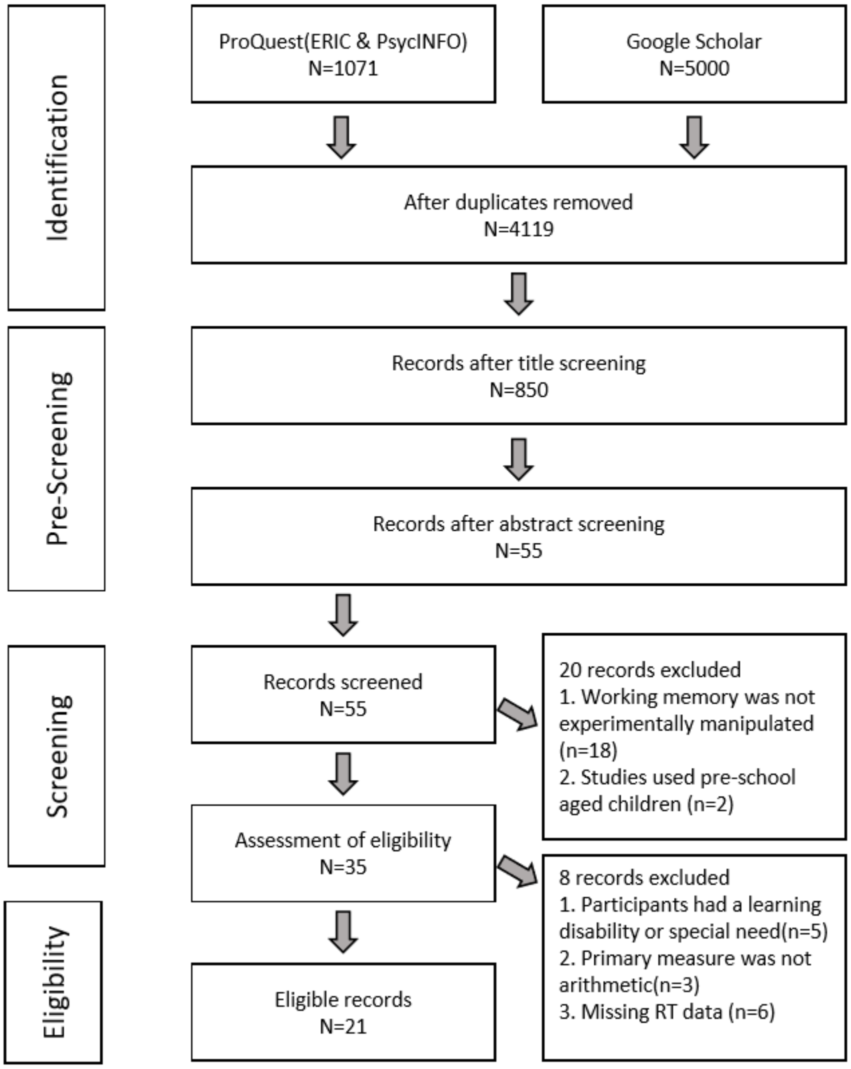
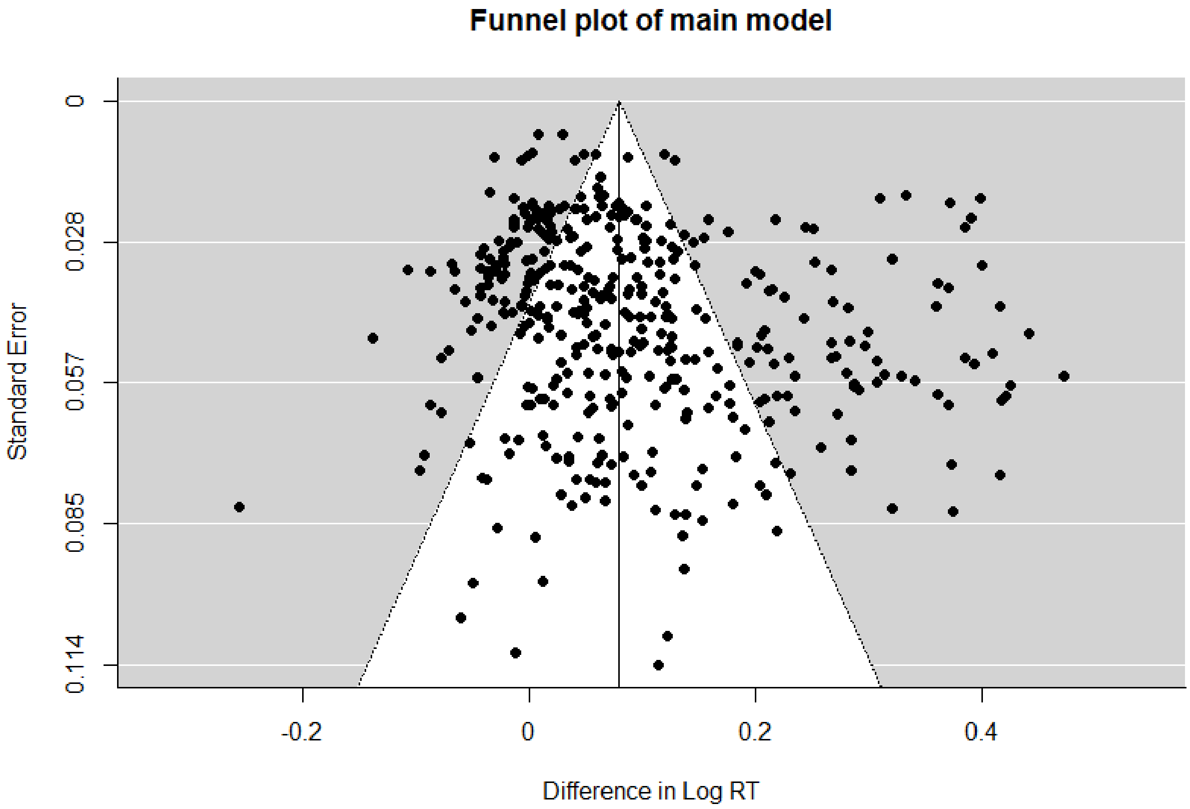

Identification
Pre-Screening
Screening
Eligibility
ProQuest(ERIC & PsycINFO)
N=1071
Google Scholar
N=5000
After duplicates removed
N=4119
Records after title screening
N=850
Records after abstract screening
N=55
Records screened
N=55
20 records excluded
1. Working memory was not experimentally manipulated (n=18)
2. Studies used pre-school aged children (n=2)
Assessment of eligibility
N=35
8 records excluded
1. Participants had a learning disability or special need(n=5)
2. Primary measure was not arithmetic(n=3)
3. Missing RT data (n=6)
Eligible records
N=21

Figure 2. Funnel Plot of Log RT Differences.



Note: Standard errors were calculated using the assumptions from the main specification, assuming a within subject correlatio

Study 2

No clear support for differential influences of visuospatial and phonological resources on mental arithmetic: A Registered Report

Accepted for publication at *Journal of Numerical Cognition*

Differential influences of visuospatial and phonological resources on mental arithmetic

Evidence from cognitive psychology and neuroscience suggests domain-specific components of working memory may contribute to differences in mental arithmetic performance, but several important questions remain unanswered. A number of imaging and lesion studies suggest the parietal regions are heavily involved with the process of mental arithmetic, specifically addition and subtraction as well as with visuospatial processes (Dehaene & Cohen, 1997; Dehaene et al., 2003; Prado et al., 2011). Meanwhile, additional evidence suggests that another arithmetic operation, namely multiplication, relies on different neural substrates found within the perisylvian areas which have been found to modulate phonological and language processes (Dehaene & Cohen, 1997; Dehaene et al., 2003; Kawashima et al., 2004, Prado et al., 2011). These would suggest that visuospatial processes are involved with subtraction while phonological resources are involved in multiplication; however, behavioral experiments do not paint this exact picture.

The current study will review these approaches and their findings and describe our current approach to investigate the unique contributions of working memory within mental arithmetic. An influential study by Lee & Kang (2002) investigated a differential effect of working memory resources on arithmetic operation type. In their study, participants were given single-digit multiplication and subtraction trials where answers were typed in using a number pad. Participants performed arithmetic in three conditions: with no secondary task, while repeating a non-word string (phonological (PL) load), or while remembering the shape and position of an object (visuospatial (VSSP) load). They reported very large effects indicating that Korean undergraduates' multiplication performance was worse than subtraction under phonological load (Cohen's d = 2.39[4]; Table 1). A similarly large but opposite effect was reported where subtraction performance was worse than multiplication under visuospatial load (d = 3.35). Interestingly, the effect of PL load on subtraction relative to subtraction alone was almost 0, as was the effect of VSSP load on multiplication relative to multiplication alone. They predicted arithmetic operations to be facilitated through specific modular representations; that is, multiplication is enacted through an auditory-phonological encoding while subtraction is enacted through an analog magnitude system like a mental number line. This line of reasoning is consistent with parallel processing theories of dual-task performance which ascribe differences in reaction time and accuracy performance to domain-specific resources competing for space within the working memory (Navon & Miller, 1987; Pashler, 1994). In other words, the more similar two tasks appear with regards to the overlap between the demands of the primary task and the modality imposed demands of the secondary task, such as a visuospatial span task with a visual imagery task, the more interference we should observe.

Several studies have used similar methods; although some have replicated the direction of these effects, none have produced the pattern of opposite effects with magnitudes approaching the size in the original Lee & Kang (2002) study. Strikingly, while there is variation in the kinds of tasks and samples among others, the original study does not seem to be sufficiently different in its design that would lead to the discrepancy in effect sizes (Table 1). Neither of the two partial replications used an entirely within-subject design like Lee & Kang (2002), which could have potentially led to the discrepancy in effect sizes. The current paper will go beyond Lee &

[4] Cohen's *d* was calculated by hand. RT means were taken from reported values within Lee & Kang while SD were calculated from reported standard errors (SD = SE*$\sqrt{n}$). Thus, we used the following values: multiplication under phonological load (M = 1169.5, SD = 82.85); subtraction under phonological load (M = 993, SD = 63.56). Values were then input into the classic Cohen's *d* formula: d = $\frac{M_1 - M_2}{\sigma}$, where σ is the pooled standard deviation of the two means: pooled SD = $\sqrt{\frac{SD_1^2 + SD_2^2}{2}}$. The same method was used to calculate the effect size for visuospatial load.

Kang (2002) and the previous replication attempts by using an entirely within-subjects design and by using a larger sample size than any of the previous studies. Imbo & LeFevre (2010) attempted to replicate the findings using a mix of native Chinese and Canadian participants to perform arithmetic problems under load (see Table 1 for details). They found differential impacts of phonological and visuospatial loads in Chinese students attending a Canadian university but not in other Canadian students. However, the interaction was only found in multiplication errors such that multiplication was less accurate in the Chinese students compared to Canadian students when under phonological cognitive load. While the effect of visuospatial load was not found in subtraction, Chinese students exhibited decreased performance compared to Canadian students on the secondary visuospatial task when arithmetic was presented vertically. While multiplication was affected by PL load, subtraction should have been impaired by VSSP load due to students having abacus training. Differences in performance were attributed to cultural differences in education, such as the use of the rhyming song many Chinese students use to learn multiplication which requires phonological resources (Imbo & LeFevre, 2010, p. 183). Meanwhile, authors hypothesized that learning addition and subtraction on an abacus, a more common practice in China than Canada, causes students to use strategies that require greater visuospatial resources.

Considering the variation in design features, the inconsistent results from previous attempts to replicate Lee & Kang (2002) have been attributed to a number of possible reasons. First, a lack of balancing the cognitive demands of the working memory and arithmetic tasks within and across participants raises uncertainty over whether it was the difficulty or specific modality of secondary tasks that led to the interaction reported in Lee & Kang (2002). The use of different multiplication and subtraction tasks as well as WM tasks mask the extent to which modality effects are separate from the inherent demands of the tasks themselves. Cavdaroglu & Knops (2016) attempted to resolve this issue by having German participants perform arithmetic under similar load conditions to Lee & Kang (2002). Importantly, they created two difficulty conditions that were individually determined through psychometric functions to ensure participants were performing symmetrically difficult secondary tasks. In addition, their calculation tasks attempted to minimize central executive resources by controlling for problem size and difficulty. Under these conditions, their results yielded no differential impact of working memory resources on multiplication and subtraction. Despite claims that the most prominent dissociations exist between multiplication and subtraction (see Lee & Kang, 2002; Lee 2000), these results suggest the validity of the domain-specific working memory influences on mental arithmetic is not as clear. However, difficulty alone may not fully explain the disparity in effect sizes. These previous replication attempts have used different working memory tasks to load the PL and VSSP, so it is not clear either whether the tasks in Lee & Kang (2002) happened to load the WM components more than the replication attempts. Third, the original study included Korean participants, whose math education differs from U.S. and Canadian samples. As evident from Imbo & LeFevre (2010), the Chinese participants who share similarities with Koreans in number system and arithmetic education (e.g., favoring rote memorization through drilling and songs and some mental-abacus training) were the only population that saw a selective interaction effect while the Canadian participants did not. The automaticity gained through extensive practice using specific representational strategies (i.e. phonologically-based rhyming songs and visuospatially-based mental-abacus) in Chinese students was believed to cause a stronger connection between arithmetic operations and specific working memory components. In comparison, Imbo & LeFevre (2010) argued that western math education caused students to use more variable strategies suggesting a weaker link between specific components and arithmetic but a stronger link to executive resources.

Moreover, current meta-analytic evidence of dual-task experiments also suggest that the influence of specific working memory components on arithmetic performance may not be as robust as other findings related to dual-task performance, such as the effect of domain-general demands of the secondary task on performance (Chen & Bailey, 2021). Specifically, it appears that larger effect sizes between different combinations of WM load and arithmetic may be partly driven by researchers predicting larger effects for more demanding secondary tasks (e.g., those that require more central executive processing). Given that there are several ways to probe potential interactions in dual task arithmetic experiments, the robustness of these findings warrants further testing. In summary, it is unclear whether the results from replication attempts reflected important insights regarding arithmetic cognition or if they reflected idiosyncratic aspects of Lee & Kang's (2002) study, specific to a combination of the tasks and sample. Thus, it is imperative to establish better practice towards registering planned analyses in the future.

Current Study

While Cavdaroglu & Knops (2016) improved upon the original design of Lee & Kang (2002), some remaining issues need to be experimentally investigated. The current design will go beyond Cavdaroglu & Knops (2016) in a number of ways. First, the arithmetic condition will be a within- rather than between-subject factor design. It should be noted that this is only fully true when there are no differential sequence effects to be expected, thus we have carefully randomized and counterbalanced the order of conditions and will perform additional analyses to follow-up our main analysis. Specifically, we tested for the key interaction (i.e. load type × arithmetic operation) for the first arithmetic under load condition within each participant.

Second, it was unclear from Imbo & LeFevre (2010) whether cultural differences in arithmetic performance were confounded by the particular tasks used, so this study will re-examine cultural differences in arithmetic cognition by recruiting students who received their primary math education in China as well as participants who received their primary math education in the U.S.

In the current study, a dual-task paradigm was used to test the involvement of phonological and visuospatial resources within mental subtraction and multiplication. The aim of this study is to test whether the findings reported in Lee & Kang (2002) can be replicated using similar procedures and tasks as used by Cavdaroglu & Knops (2016). Participants solved either multiplication or subtraction problems under phonological (i.e. remembering a string of letters or numbers) and visuospatial load (i.e. remembering the positions of dots in an array). The interaction between these memory load types and operation types was most prominent in Lee & Kang (2002). However, attempts to replicate this large dissociation since have not been wholly successful (see Table 1). Task difficulty (i.e. span size) was balanced and varied within and across participants through an adaptive staircase procedure. Two different difficulty thresholds (80% and 99%) were determined in blocks at the beginning of the experiment in session 1. These difficulty thresholds were used to investigate how task difficulty affects performance. Altogether, this study will attempt to reconcile debates over the differential contributions of working memory in mental arithmetic and provide insight with respect to potential underlying mechanisms related to mathematical cognition.

Methods

Participants

Power analysis

We used the software program G*Power to conduct an a priori power analysis (Faul et al., 2009). F statistics or η_p^2 values for the interaction between WM load and arithmetic operation could not be derived from Lee & Kang (2002) nor Cavdaroglu & Knops (2016). However, other 2- and 3-way interactions were provided from Imbo & LeFevre (2010) (e.g. culture × problem difficulty; culture × problem difficulty × presentation format) to approximate values for the power analysis. Our goal was to obtain .90 power to detect a partial eta-squared (η_p^2) of .07 for a 3-way interaction at alpha = .05. We used the η_p^2 reported for the 3-way interaction between culture x problem difficulty x presentation format in Imbo & LeFevre experiment 2 (2010), as this was the most conservative effect size reported relating to arithmetic performance. For the statistical test, we chose “ANOVA: Repeated measures, within-between interaction” because the interaction from Imbo & LeFevre (2010) contained within factors (problem difficulty & and presentation format) and a between factor (culture). We inputted the reported $\eta_p^2 = .07$ after clicking “Determine =>”. Calculating this provided an effect size of 0.27. The assumed correlation between repeated measures was left at the default of 0.5 because we had no other underlying assumptions about the repeated measures. In addition, we specified that there were 2 groups (Chinese and U.S. math educated students) and 16 measurements (i.e. 2 arithmetic operation × 2 difficulty × 4 WM load types). While four factors are present in our design, our main focus was the 2-way interaction between operation and WM load. The additional factors used in the G*Power analysis helped derive a more conservative estimate for the number of participants needed and will be used in subgroup analyses explained further below. Following these specifications, a minimum of 14 participants was required to be powered to detect an interaction similar to that in Imbo & LeFevre (2010) and our design had an estimated power of 0.94. Prior meta-analytic data also suggests the average sample size among dual-task arithmetic experiments (containing both within and between designs) consists of around 20 participants with a range from 10-60. Following prior literature and our power analyses, we planned on collecting data from a sample larger than any other study before. As such, we determined that 100 participants would be sufficiently powered to detect our key interaction within our main model and secondary analyses.

Following this plan, we recruited and ran 100 total participants from the University of California, Irvine (Female = 64, age range = 18 – 25 years old, mean = 20.1 (*SD* = 1.3). 22 of the final analysis sample received the majority of math education in China prior to entering university studies in the US. All participants had normal or corrected-to-normal vision. All research was performed in accordance with the ethical standards of the Institutional Review Board. Written informed consent was obtained from all participants and were given course credit through the Human Subjects Lab Pool or were reimbursed $30 for their participation.

Stimuli

All tasks used in these experiments were created through PsychoPy 3 (Peirce et al., 2019). Performance on the span tasks and arithmetic will be measured by reaction time (RTs in ms) and accuracy (ACCs in percentage correct). For examples, see Figure 1. Arithmetic problems used in this experiment are the same as in Cavdaroglu & Knops (2016). Working memory staircase tasks are based on the descriptions used in Cavadaroglu & Knops (2016). Strategy report is a one-item survey question asking about strategy use. All materials including experimental tasks and protocol used will be available online as supplementary materials (https://doi.org/10.23668/psycharchives.6881).

Subtraction

Subtraction problems were presented using a 2-alternative forced choice (2AFC) paradigm. Participants were presented with simple two-digit – two-digit problems for 2 s. There were no borrowing or crossing of decade boundaries to minimize central executive involvement. Participants then chose from two answer choices which were displayed for 3 s or until participants respond. Three different sets of subtraction problems were used across three rounds (round 1: subtraction only; round 2: subtraction under phonological load; round 3: subtraction under visuospatial load) with easy and hard working memory loads split across 2 blocks. The order of the three sets as well as the difficulty blocks were counterbalanced across participants. Each set contained 28 different subtraction problems where each was displayed twice in total with a different answer pair each time. The order of the three sets was counterbalanced across all participants. In half of the answer pairs, the correct and alternative answers had a distance of 2; whereas the other half had a distance of 10. This was done in order to encourage participants to take into account both decades and units and to discourage the strategy of paying attention only to the units or decades. Distance from correct response were either in the positive or negative direction. For example, for the problem 36-14, the two answer pairs were 22 vs. 20 (difference = -2) or 12 vs. 22 (difference = +10). Problems with a decade in one of the operands or in the result were excluded. Eleven was not used as an operand.

Multiplication

Multiplication problems were presented using a 2AFC paradigm. Participants were presented with simple one-digit by one-digit and two-digit by one-digit multiplication problems. Participants then chose from two answer alternatives which were displayed for 3 s or until participants responded. Three different sets of multiplication problems were used across three rounds of tasks (round 1: multiplication only; round 2: multiplication under phonological load; round 3: multiplication under visuospatial load) with easy and hard working memory loads split across 2 blocks. The order of the three sets as well as the difficulty blocks was counterbalanced across participants. Each set contained 28 different subtraction problems where each was displayed four times in total with a different answer pair each time. Among the four answer pairs, one contained a response alternative from the multiplication table of the first operand, another contained an alternative from the multiplication table of the second operand (table-related response alternatives) and the other two pairs contained response alternatives that were not from either operand's multiplication table (non-table-related response alternatives). For example, for the problem 12 × 7, the four different answer pairs were 84 vs 98 (98 from 7's table), 84 vs 72 (72 from 12's table), 84 vs 64, and 84 vs 94. Half of the problems were two-digit by one-digit and the other half were one-digit multiplication. In one-digit multiplication trials, the smaller operand preceded the larger operand. In two-digit by one-digit trials, the two-digit operand preceded the one-digit. The two-digit number was smaller than twenty. The one-digit number was larger than two. Tie problems (e.g. 6 × 6) and problems with a decade in the operand or result were excluded. Products were all below 100 to restrict responses to be two-digits at most like in the subtraction task.

Phonological staircase

Following the same task designs as those outlined in Cavdaroglu & Knops (2016), participants' phonological processing span was measured using an adaptive staircase procedure of letter sequences. Participants were instructed to keep a sequence of letters – in original order – in mind and decide whether a second set of letters (shown 7s after onset of the first sequence) contained the exact same order of letters or not. Letter sequences were displayed for a duration of 0.4 s * n – n indicating number of letters – followed by 3 s on a fixation screen before participants are given 4 s to respond. Participants were presented upper case letters in the first sequence and tested using lowercase letters (B C D vs. b c d) in order to encourage participants to use their phonological rather than visual memory. In half of the trials, the test sequence had the same letters in the exact order as the first sequence (e.g., ‘B C D’ and ‘b c d’); whereas in the other half of the trials the position of two letters were swapped (e.g., ‘B C D’ and ‘b d c’). The ‘F’ and ‘J’ keys were used for responding to allow for natural hand placement on the keyboard. The task will start with 3 letters and reach a maximum of 9 letters and a minimum of 1 letter. After three correct responses in a row, the difficulty of the task increased by 1 letter otherwise, if there were three consecutive incorrect responses, the difficulty of task decreased by 1 letter until the minimum number of letters are reached or until a correct response is given. 30 trials were conducted to measure phonological span. In addition, a Weibull function was fit on the data where the inverse of the Weibull function was used to determine the number of letters corresponding to 80 and 99% accuracy. The two threshold levels were chosen to examine the effect of task difficulty (low vs high) on arithmetic performance in both single- and dual-task conditions. In each trial, the string of letters was randomly chosen from this set of 10 consonants [B, C, D, F, G, H, J, K, L, M]. Vowels were excluded to prevent use of semantic strategies and other consonants were excluded to maintain the same number of digits to letters. In total, the staircase contained 30 trials.

Visuospatial staircase

The visuospatial span task also followed similar procedures to those used in Cavdaroglu & Knops (2016), where span was measured using an adaptive staircase procedure on dot-matrices. Participants were instructed to keep the locations of dots within a 5×5 grid in mind and decide if a second grid (shown 7s after onset of the first grid) contained the exact same locations of dots. Dot-arrays were displayed for a duration of 0.4 s * n – n indicating number of dots – followed by 3 s on a fixation screen before participants are given 4 s to respond. In half of the trials, the positions of the test dots were in the same position; whereas in the other half of the trials, the positions of two dots were replaced somewhere else on the grid. The ‘F’ and ‘J’ keys were used for responding. The task started with 3 dots and reached a maximum of 9 dots and a minimum of 1 dot. After three correct responses in a row, the difficulty of the task increased by 1 dot otherwise, if there were three consecutive incorrect responses, the difficulty of the task decreased by 1 dot until the minimum number of dots was reached or until a correct response was given. 30 trials were conducted to measure visuospatial span. Finally, a Weibull function was used to determine the 80 and 99% accuracy thresholds for the dual-task condition.

Procedure

The study used a 2×3 factorial design using within-subject factors. The within-subject factors were arithmetic operation type (subtraction or multiplication) and WM load type (no load, PL load, and VSSP load). No-load (i.e. arithmetic alone) conditions served as controls against dual-task conditions. While culture and difficulty were part of the analysis, these were only considered in the subgroup analyses and not for additional interactions, because our main focus was on the operation × load interaction. The entire experiment was conducted online through video conferencing in which an experimenter guided the participant in downloading the required materials and protocol for completing experimental tasks. The experiment was administered within two sessions that were scheduled to be around the same time and spread apart by one week. Participants were also instructed to abstain from taking any alcohol or drugs prior to either session. Participants completed the experiment using their own devices. To ensure that reactions times were sufficiently accurate and consistent across different devices and operating systems, participants were instructed to use either a home desktop or laptop rather than a tablet or mobile phone. No information related to the participants’ devices, such as IP address, were maintained except for the operating system (e.g. Windows 10, Mac-OS) in order to ensure proper installation of PsychoPy and the experiment itself. Recordings were also not taken to respect the privacy of the participants.

In session 1, participants were given a brief questionnaire to capture their demographic information and math education background before being introduced to the PsychoPy environment and to downloading the experimental tasks. These questions included asking about their current major and the number of math courses they have taken since entering university. In addition, we asked specific math background questions including, “Prior to coming to university, in which country did you receive the majority of your math education?”, “If you were taught how to use an abacus or mental abacus strategy for doing math, how often have you used it? (Never taught; Never used; Rarely; Sometimes; Often; Very often)”, and “Do you consider yourself an A, B, C, D, or F student compared to your peers?”. Altogether, these questions allowed us to potentially examine differences in math proficiency among our sample, especially in our comparison between the Chinese-educated student group and the non-Chinese-educated student group. From here, participants were given the adaptive phonological and visuospatial staircase tasks. Prior to the staircase, 10 practice trials were administered to familiarize the participant with the stimuli and testing environment. Discounting the practice trials, there were 30 trials per staircase for a total of 60 trials to determine difficulty thresholds. The order of these tasks were randomized and counterbalanced for all participants. Staircase performance from session 1 were used to determine easy and hard span levels for the dual-task conditions used in session 2. In total, the first session took approximately 60 minutes.

In session 2, participants started the dual-task experiment. Participants downloaded their PsychoPy tasks that were modified to fit the appropriate difficulty levels as determined in session

1. Participants then completed arithmetic alone and under load over 4 experimental blocks (multiplication-easy load, multiplication-hard load, subtraction-easy load, subtraction-hard load). The order of these tasks followed a block-randomization wherein the single-arithmetic task was always administered first in the block followed by either the visuospatial or phonological loads. Half of the participants received the visuospatial load before the phonological load, while the other half received the phonological load first. The order of the four blocks was also randomized and counterbalanced for each participant such that each of the possible sequences as well as their reverse orders appeared an equal number of times. 10 practice trials were given before the start of the first block to familiarize participants with the dual-task procedure. Participants then completed each block which contained 28 arithmetic problems for each condition (arithmetic alone, with PL load, with VSSP load) for a total of 336 trials. The order of conditions was also randomized and counterbalanced. At the end of each block, participants were be given up to a 5-minute break. Participants finished after completing the 4th block. In total, the second session took no more than 2 hours to complete.

Analysis plan

In this experiment, we focused on the key interaction predicted by Lee & Kang (2002). Specifically, we tested the following hypotheses:

Hypothesis 1: As predicted by Lee & Kang (2002), we expected an interaction between arithmetic operation type and WM load type; specifically:

Hypothesis 1a: Multiplication performance is slower and less accurate under PL load compared to VSSP load

Hypothesis 1b: Subtraction performance is slower and less accurate under VSSP load compared to PL load.

In addition to these, we tested secondary hypotheses regarding the differences between single-task arithmetic conditions vs each of the dual-task conditions as they were reported in Lee & Kang (2002) such that:

Hypothesis 1c: Multiplication performance alone is significantly faster than under PL load but not VSSP load.

Hypothesis 1d: Subtraction performance alone is significantly faster than under VSSP load but not PL load.

According to Imbo & LeFevre (2010), the crossover effect may be found within Chinese-educated samples; but not US-educated samples, thus we tested the following hypotheses:

Hypothesis 2: Receiving primary math education from China but not the US is associated with differences in load type by arithmetic operation performance, specifically:

Hypothesis 2a: Multiplication performance is slower and less accurate under PL load compared to VSSP load only in Chinese-educated samples.

Hypothesis 2b: Subtraction performance is slower and less accurate under VSSP load compared to PL load only in Chinese-educated samples.

Hypothesis 2c: Multiplication performance alone is significantly faster than under PL load but not VSSP load only in Chinese-educated samples.

Hypothesis 2d: Subtraction performance alone is significantly faster than under VSSP load but not PL load only in Chinese-educated samples.

In order to test hypotheses 1a-1d, we conducted multiple 2 × 2 ANOVAs under four model specifications (for summary of planned analyses, see Table 2). The first model included all participants and both difficulty levels. We then tested the robustness of this interaction effect by restricting the data in the following three ANOVA models: easy load condition trials only, hard load condition trials only, and first block trials only. The first block model tested whether the crossover interaction was observed for the first presented arithmetic operation under load (Table 2: last column), for which performance was assumed to be less prone to order effects. To test hypotheses 2a-2d, we restricted the sample to only those students who reported having received the majority of their math education in China prior to entering university. We conducted both a 2 × 2 ANOVA of the restricted sample and compared the Chinese-educated students to the rest of our sample using a 2 × 2 × 2 ANOVA with country of primary math education as a between subject factor. While we investigated this possible group difference, the crossover interaction was our main interest. Given the unequal sample sizes in the Chinese vs non-Chinese model, we ran a Tukey-Kramer test as a post-hoc adjustment. If any of the above models produced a significant interaction effect, we conducted post hoc analyses to test whether results aligned with hypotheses 1a -2d.

Even though we acknowledge that testing these multiple hypotheses inflates the probability of type-1 errors, we chose not to adjust error levels for each statistical test, because a statistically significant interaction does not guarantee any of the more specific hypotheses to be supported. Instead, we reported on the level of support for the theorized crossover effect and predicted simple effects based on how closely our reported findings aligned with our predictions. For hypotheses 1a-1d, we concluded that there was strong support for the underlying theory if we detected an interaction and main load effects in directions consistent with Lee & Kang (2002) within our main specifications containing all participants. We concluded there was mixed evidence for the crossover effect if only one of the main load effects was consistent with predictions within the main model (i.e., a) if VSSP affects subtraction but not multiplication or b) PL affects multiplication but not subtraction, but not both a and b) or if we only found the interaction in one or more of the subgroup analyses; for example, if the crossover effect was only present in the Chinese-educated sample but not the US sample or only in hard but not easy load conditions. If results were fully null, we concluded that we were unable to find evidence for an interaction. Results of analyses will be reported regardless of whether our hypotheses were supported or not.

As a complement to the frequentist analyses of the interaction effect, we also report a Bayesian analysis for the main model (whole group) to examine the relative support for both our hypotheses of interest and the null hypothesis. We conducted a Bayesian repeated measures ANOVA, dependent on the 2 × 2 factors in the main model. Following Morey & Rouder (2011), we set a non-informative Jeffreys prior width of 0.5 to correspond to a small effect. Such analyses result in a Bayes factor (BF_{10}), which can be interpreted as the likelihood ratio for the alternative hypothesis over the null. Given that the Bayes factor (BF_{10}) is a ratio of the likelihood for the alternative hypothesis over the null hypothesis, the inverse of the Bayes factor (BF_{01}) can be interpreted as the likelihood ratio for evidence of the null hypothesis over the alternative hypothesis. Following Jeffreys (1961) we used the following designations to interpret the strength of the Bayes factors: 0–3 offer anecdotal support for H1, 3–10 moderate support for the H1, 10–30 strong support for H1, 30–100 very strong evidence for H1, and values greater than 100 offer decisive evidence for H1. Conversely, we use the inverse of these ranges to interpret support for the null hypothesis (BF_{01} anecdotal 0.33–0, moderate 0.10–0.33, strong 0.10–0.03, very strong 0.03–0.01).

Data were analyzed primarily in JASP using its frequentist and Bayesian repeated measures ANOVA and paired-sample t-test functions (JASP Team, 2020). Data were organized for JASP usingRStudio (RStudio Team, 2020), specifically tidyverse for data visualization and formatting (Wickham et al., 2019). The RMarkdown is available as supplementary material to reproduce data created for JASP (https://doi.org/10.23668/psycharchives.6880). Where appropriate, Holm-Bonferroni correction was used to correct for multiple comparisons in post-hoc testing (Holm, 1979). Huynh–Feldt correction was used when sphericity was violated. Bayesian analyses were conducted using the Bayesian repeated measures ANOVA function in JASP (JASP Team, 2020). All reaction time (RT) analyses were based on correct trials only. Accuracy or response times outside the range of a participant's mean ± 3 SDs were discarded from further analyses. Responses faster than 200 ms were also discarded. Based on that criterion, 1.02 % of trials in single arithmetic blocks and 3.56 % of the trials in dual-task blocks were eliminated. In addition, 3 participants were excluded from data analyses for not responding in a majority of trials during the second session. All data are publicly available in PsychArchives (https://doi.org/10.23668/psycharchives.6882). Of note, even though our participants were tested at home on their own devices, average reaction times per WM load condition within our study were comparable to those found in Lee & Kang (2002) and Cavdaroglu & Knops (2016) (Table 1).

Deviations in pre-registration analyses

The following analyses were either changed or added from the pre-registration. Full documentation of all deviations can be found in a document within the supplementary files (https://doi.org/10.23668/psycharchives.6881). The 2 × 2 × 2 ANOVA investigating the differential effect of WM load on arithmetic operation between the samples receiving education from the US and China was included in the pre-registration, but we also included the 2 × 2 ANOVA analyses which only looked at the Chinese-educated subsample as an additional robustness test. We also conducted additional Bayesian paired samples t-tests in addition to the

Bayesian repeated measures ANOVA to investigate post-hoc differences in reaction time and accuracy for hypotheses 1a, 1b, 2a, & 2b. Conclusions did not vary across methods.

Results

Hypothesis 1a: Multiplication performance is slower and less accurate under PL load compared to VSSP load

In contrast to our hypothesis, in the full sample, multiplication performance was not significantly slower (Figure 2) nor was it less accurate (Figure 3) under PL load compared to VSSP load. ANOVA results from Tables 4 and 5 yielded no significant difference in multiplication reaction time [RT: $F(1, 96) = 1.20$, $p = 0.28$, $\eta_p^2 = 0.01$] nor accuracy [ACC: $F(1, 96) = 0.49$, $p = 0.49$, $\eta_p^2 = 0.01$] between verbal and visuospatial dual-task load. We ran complementary Bayesian t-tests of WM load on multiplication RT and ACC for the full sample. We found stronger evidence for the null hypothesis than for Hypothesis 1a such that there was no difference in multiplication RT and ACC by WM load type [RT: BF_{01}: 4.99; ACC: BF_{01}: 7.03]. Higher BF_{01} indicate greater support for the null hypothesis over the alternative. In addition to our Bayesian *t*-tests, we also ran a Bayesian repeated measures ANOVA of all of the full sample focusing on the 2 (verbal and visuospatial WM load) × 2 (multiplication and subtraction) interaction. Comparison of model Bayes factors (BF_{10}) can be found in Tables 8 and 9. While the tables use the null model as a reference, it is more useful to compare Bayes factor between an additive model (WM task + arithmetic) and the interaction-included model (WM task + arithmetic + WM task × arithmetic). Comparing model fit between the two can be accomplished by taking the ratio of the Bayes factor of the additive model to the interaction-included model. The inverse of the ratio would provide a Bayes factor of the interaction alone compared to the null. The Bayesian ANOVA indicated anecdotal to moderate support for the additive model over the interaction-included model. The Bayesian ANOVA indicated that Bayes factors for additive model of WM task and arithmetic fit both reaction time and accuracy data better across the whole sample than with the additive and WM task × arithmetic operation interaction term included [RT BF_{10} ratio: 6.29; ACC BF_{10} ratio: 3.69]. Higher BF_{10} ratios indicate greater support for the additive model over the interaction model.

As a preregistered robustness check, we estimated the same models for three subsamples of the data: easier secondary task blocks, more difficult secondary task blocks, and the first arithmetic block under cognitive load only. Similar patterns of results for both frequentist and Bayesian analyses can be found in our secondary analyses of the easier load, harder load, and first block conditions (Figure 2 & 3, Tables S1, S2, S5, & S6 in the Appendix); nearly all Bayesian estimates provide support for the null hypothesis. Only in the harder difficulty load condition[5] was there an effect on reaction time consistent with Hypothesis 1a [$F(1, 96) = 6.57$, $p < 0.05$, $\eta_p^2 = 0.07$]. A post-hoc pairwise t-test of the hard load condition revealed a small but significant slowing in multiplication RT of 39 ms when under verbal load compared to visuospatial load [$t(96) = 2.67$, $p = 0.017$, $d = 0.16$, Holm-Bonferroni corrected].

Hypothesis 1b: Subtraction performance is slower and less accurate under VSSP load compared to PL load.

Again, in contrast to our hypothesis, in the full sample, subtraction performance was not significantly slower (Figure 2) nor was it was less accurate (Figure 3) under VSSP load compared to PL load. The ANOVA results shown in Tables 4 yielded no significant difference in

[5] The staircase procedure used during the first session of each experiment to estimate each participant's subjective 80th and 99th percentile threshold for their verbal and visuospatial cognitive loads provided reasonable estimates. On average, the 99th percentile (easy load) threshold for participants was 5.52 (*sd* = 1.28) for their verbal WM load and 6.52 (*sd* = 0.95) for their visuospatial WM load. For the 80th percentile (hard load), the threshold for participants' verbal WM load was 7.52 (*sd* = 1.28) and 8.52 (*sd* = 0.95) for their visuospatial WM load.

subtraction reaction time [RT: $F(1, 96) = 0.15$, $p = 0.70$ $\eta_p^2 = 0.002$]. Results shown in Table 5 yielded a significant difference in accuracy [ACC: $F(1, 96) = 6.31$, $p = 0.01$, $\eta_p^2 = 0.06$] between verbal and visuospatial dual-task load. However, this effect was in the opposite direction as predicted: Post-hoc pairwise t-test of the whole sample yielded a statistically significant decrease in subtraction accuracy of about 2 percentage points when under verbal load compared to visuospatial load [Whole: $t(96) = 2.57$, $p = 0.04$, $d = 0.19$, Holm-Bonferroni corrected]. Bayesian t-tests of WM load on subtraction RT and ACC found stronger evidence for the null hypothesis than for Hypothesis 1b such that there was no difference in subtraction RT by WM load type [RT BF_{01}: 8.29], but there was a difference in accuracy in favor of verbal load [ACC BF_{01}: 0.46]. $BF_{01} > 1$ indicate more support for the null than the alternative while $0 \leq BF_{01} \leq 1$ indicate greater support for the alternative. Our Bayesian repeated measures ANOVA from the previous section included subtraction in the model, thus they can be applied here as well (also see Tables 8 and 9).

As a preregistered robustness check, we estimated the same models for the easy, hard, and first-arithmetic block under load subsamples of the data. Similar patterns of reaction time results for both frequentist and Bayesian analyses can be found in our secondary analyses of the easier load, harder load and first block conditions (Figures 2 & 3, Tables S1, S2, S5, & S6 in the Appendix). In accuracy, we found a significant effect of WM load type within the easy load and first cognitive load block [$F(1, 96) = 7.44$ & 5.77, $p = 0.01$ & 0.02, $\eta_p^2 = 0.07$ & 0.06, respectively]. However, this effect was consistent with what was found for the whole group, such that verbal load *lowered* accuracy more than visuospatial load [Easy: $t(96) = 2.73$, $p = 0.01$, $d = 0.23$, Holm-Bonferroni corrected; First: $t(96) = 2.40$, $p = .02$, $d = 0.22$, Holm-Bonferroni corrected], opposite of theoretical predictions.

Secondary hypotheses:

Hypothesis 1c: Multiplication performance alone is significantly faster than under PL load but not VSSP load.

To test Hypothesis 1c we included the single multiplication task condition into the 2-way ANOVA and performed pairwise t-test comparisons with Holm-Bonferroni corrections as needed. We did not find support for Hypothesis 1c: multiplication performance under no load was significantly faster (Figure 2) and more accurate (Figure 3) than both load conditions across most subsamples. There was a significant main effect of load vs. no load on multiplication reaction time for the whole sample[$F(1, 96) = 74.90$p $< .001$, $\eta_p^2 = 0.44$]. WM load yielded an average slowing of 143 ms or 18% [$d = 0.64$] in the whole sample. Mean comparisons and post-hoc pairwise *t*-test results are shown in Tables 6. Reaction times under both verbal and visuospatial load were significantly slower than multiplication alone [RT: both $t(96) > 9.70$, $p <$ 0.001, d = [vs Verbal: 0.68; vs Visuospatial: 0.60]]. There was also a significant effect of load on multiplication accuracy for the whole sample [$F(1, 96) = 12.92$p < 0.001, $\eta_p^2 = 0.12$] with accuracy being reduced by about 3 % [$d = 0.30$]. Mean comparisons and post-hoc pairwise *t*-test results are shown in Tables 7. Accuracy comparisons were significant for the whole sample [ACC: both $t(96) > 3.99$, $p < 0.001$, d = [vs Verbal: 0.32; vs Visuospatial: 0.28]]. We included the no load level into our Bayesian repeated measures ANOVA. Our Bayesian ANOVA indicated moderate to strong support for the additive model over the interaction-included model. Tables 10 and 11 of our Bayesian ANOVA indicated that even with the inclusion of single task arithmetic, the combination of WM task and arithmetic operation fit both reaction time and accuracy data better across the whole sample than with the inclusion of the WM task × arithmetic operation interaction term [RT BF_{10} ratio: 19.12, ACC BF_{10} ratio: 13.90]. Higher BF_{10} ratios indicate greater support for the additive model over the interaction model.

For our preregistered robustness check, we estimated the same models for the easy, hard, and first-arithmetic block under load subsamples of the data. Our frequentist and Bayesian analyses for our subsample analyses yielded similar patterns of results to our whole sample analyses (Tables S3, S4, S7, & S8 in the Appendix). Only in the easier load condition was there no significant difference in accuracy between single multiplication and multiplication under visuospatial load, $p = 0.62$.

Hypothesis 1d: Subtraction performance alone is significantly faster than under VSSP load but not PL load.

We found no support for Hypothesis 1d either. Subtraction performance under no load was significantly faster than either load condition (Figure 2) and more accurate than either load condition (Figure 3) across all subsamples. There was a significant main effect of load vs. no load on subtraction reaction time for the whole sample[$F(1, 96) = 62.28$, $p < .001$, $\eta_p^2 = 0.39$]. WM load yielded an average slowing of 139 ms or 19% [$d = 0.59$] in the whole sample. Mean comparisons and post-hoc pairwise *t*-test results are shown in Tables 6. Reaction times under both verbal and visuospatial load were significantly slower than subtraction alone[RT: both $t(96) > 9.76$, $p < 0.001$, d = [vs Verbal: 0.61; vs Visuospatial: 0.57]]. There was a significant effect of WM load on accuracy as well [ACC: $F(1, 96) = 13.54$, p < 0.001, $\eta_p^2 = 0.12$] with about a 3 % [$d = 0.31$] reduction in accuracy under load in the whole sample. Mean comparisons and post-hoc pairwise *t*-test results are shown in Tables 7. Subtraction accuracy was weaker under verbal load compared to no load [$t(96) = 5.23$, $p < 0.001$, $d = 0.37$] but not between no load and visuospatial load [$p = 0.07$]. Our Bayesian repeated measures ANOVA results are the same as reported for Hypothesis 1c.

For our preregistered robustness check, we estimated the same models for the easy, hard, and first-arithmetic block under load subsamples of the data. Both frequentist and Bayesian analyses for the subsample analyses yielded similar patterns of results to our whole sample analyses (Tables S3, S4, S7, & S8 in the Appendix).

Comparing US- v Chinese-educated participants: Hypotheses 2a-2d

To test whether the differential influence of working memory depends on where students received their primary math education, we computed a 2 (country; US- vs. Chinese-educated) × 2 (WM load) × 2 (arithmetic) ANOVA in order to test whether the differential impact of WM load type on arithmetic operation is dependent on where participants received the majority of their math education. The 3-way ANOVA did not yield a significant main effect for country [$F(1, 91) = 0.56$, $p = 0.46$, $\eta_p^2 = 0.01$], but it did yield a significant main effect for arithmetic operation [$F(1, 91) = 8.31$, $p = .005$, $\eta_p^2 = 0.08$]. Furthermore, the ANOVA did not yield a significant 3-way interaction for country × WM task × arithmetic [$F(1, 91) = 1.57$, $p = 0.21$, $\eta_p^2 = 0.02$] nor 2-way interactions for WM task × country [$F(1,91) = 0.27$, $p = 0.60$, $\eta_p^2 = 0.003$] or WM task × arithmetic [$\mathrm{F}(1, 91) = 2.73$, $p = 0.10$, $\eta_p^2 = 0.03$]. However, there was a significant 2-way interaction for and country × arithmetic [$\mathrm{F}(1, 91) = 4.25$, $p = 0.04$, $\eta_p^2 = 0.05$]. Post-hoc pairwise t-test comparisons revealed that the US-educated participants were generally slower in multiplication than in subtraction by about 100 ms [$t(91) = 5.08$, $p < .001$, $d = 0.43$, Holm-Bonferroni corrected] while no such difference in reaction times were present in the Chinese-educated participants.

In accuracy, there were no significant effects for country [F(1, 91) = 3.30, $p = 0.07$, $\eta_p^2 = 0.04$], or any 3-way or 2-way interactions. The 3-way ANOVA yielded only main effects for WM task [F(1, 91) = 6.38, $p = 0.01$, $\eta_p^2 = 0.07$] and arithmetic [F(1, 91) = 4.29, $p = 0.04$, $\eta_p^2 = 0.05$]. Taken together, these findings provide some support for the validity of two of the sources of variation in our population: First, the WM load manipulations were sufficiently difficult to impair arithmetic performance. Second, Chinese-educated students showed a different pattern of performance on arithmetic tasks, being approximately equally fast and accurate at multiplication and subtraction, relative to the US-educated participants, which were consistently faster and more accurate at subtraction than multiplication.

Hypothesis 2a: Multiplication performance is slower and less accurate under PL load compared to VSSP load only in Chinese-educated participants.

Following the lack of a 3-way interaction, we examined the Chinese-educated subgroup directly. Overall, we did not find evidence to support Hypothesis 2a. While there appeared to be a moderate effect of verbal vs visuospatial load on multiplication reaction times (see Table 1, row 5, column 5), this effect was not statistically significant, $d = 0.28$, $p = 0.51$. Our ANOVA results in Tables 4 and 5 also suggest that WM load type did not differentially impact multiplication performance [RT: $F(1, 21) = 1.69$, $p = 0.21$, $\eta_p^2 = 0.07$; ACC: $F(1, 21) = 3.59$, $p = 0.07$, $\eta_p^2 = 0.15$]. Bayesian pairwise t-tests for reaction times and accuracy produced $BF_{01} = 2.14$ and 0.99, respectively, suggesting anecdotal evidence in favor of the null hypothesis. $BF_{01} > 1$ indicate more support for the null than the alternative while BF_{01} approaching 1 suggest no evidence for either null or alternative. Bayesian repeated measures ANOVA models in Tables 8 and 9 indicated a better fit for the additive (WM load type + arithmetic operation) model over the additive + interaction model as well [RT BF_{10} ratio = 2.48; ACC BF_{10} ratio = 3.27]. Higher BF_{10} ratios indicate greater support for the additive model over the interaction model.

Hypothesis 2b: Subtraction performance is slower and less accurate under VSSP load compared to PL load only in Chinese-educated participants.

Overall, we did not find evidence to support Hypothesis 2b. The effect of visuospatial load on subtraction reaction times had a much smaller effect size than in Lee & Kang (2002), (see Table 1, row 5, column 6), but was not statistically significant either, $d = -0.09$, $p = 0.98$. Our ANOVA results in Tables 4 and 5 also suggest that WM load type did not differentially impact subtraction performance [RT: $F(1, 21) = 0.17$, $p = 0.67$, $\eta_p^2 = 0.01$; ACC: $F(1, 21) = 3.41$, $p = 0.08$, $\eta_p^2 = 0.14$]. Bayesian pairwise t-tests for reaction time and accuracy produced BF_{01} = 4.15 and 1.06, suggesting anecdotal evidence in favor of the null hypothesis. $BF_{01} > 1$ indicate more support for the null over the alternative. Similarly, our Bayesian repeated measures ANOVA models in Tables 8 and 9 found better fit for the additive (WM load type + arithmetic operation) model over the additive + interaction model as well [RT BF_{10} ratio = 2; ACC BF_{10} ratio = 3.27].

Hypothesis 2c: Multiplication performance alone is significantly faster than under PL load but not VSSP load only in Chinese-educated participants.

We did not find evidence to support Hypothesis 2c. Multiplication performance under no load was significantly faster than both load conditions (Figure 2 & 3). There was a significant main effect of load on multiplication reaction time but not accuracy for the Chinese-educated participants [RT: $F(2, 42) = 13.41$, $p < 0.001$, $\eta_p^2 = 0.39$; ACC: $F(2, 42) = 2.84$, $p = 0.07$, $\eta_p^2 = 0.12$]. Mean comparisons and post-hoc pairwise *t*-test results are shown in Tables 6 and 7. Overall, multiplication reaction time was impacted by both load types [RT: vs. Verbal: $d = 0.79$,

BF_{10} = 417; vs. Visuospatial: *d* = 0.67, BF_{10} = 31]. BF_{10} > 1 indicate greater support for the alternative that there was a difference in performance between no load and either secondary task load. The Bayesian repeated measures ANOVA reported similar patterns of results as those from Hypothesis 2a and 2b with the additive (WM load type + arithmetic operation) model fitting the data better than the additive + interaction model [RT BF_{10} ratio = 4.71; ACC BF_{10} ratio = 4.4] (Tables 10 & 11).

Hypothesis 2d: Subtraction performance alone is significantly faster than under VSSP load but not PL load only in Chinese-educated *samples.*

We did not find evidence to support Hypothesis 2d. Subtraction performance under no load was significantly faster than both load conditions (Figure 2 & 3). There was a significant main effect of load on multiplication reaction time but not accuracy for the Chinese-educated sample [RT: F(2, 42) = 5.02, p = 0.01, ηp^2 = 0.19; ACC: F(2, 42) = 0.97, p = 0.39, ηp^2 = 0.04]. Mean comparisons and post-hoc pairwise t-test results are shown in Tables 6 and 7. Overall, subtraction reaction time was impacted by both load types [RT: vs. Verbal: d = 0.36, BF10 = 1.36; vs. Visuospatial: d = 0.41, BF10 = 7.19]. BF10 > 1 indicate greater support for the alternative that there was a difference in performance between no load and either secondary task load. The Bayesian repeated measures ANOVA from Tables 10 & 11 are the same as those reported for Hypothesis 2c.

Discussion

In this registered report, we tested several pre-registered predictions based on previous findings from the dual-task literature with respect to the differential effects of secondary WM task load on arithmetic performance. That is, we tested whether verbal secondary tasks reduce multiplication performance but not subtraction performance, whether visuospatial secondary tasks reduce subtraction performance but not multiplication performance, if these differential effects can be observed relative to each other or a no load control. These predictions have implications for theories of mathematical cognition and working memory, along with dual-task performance specifically. Building upon previous work in the field, we identified potential moderators that could explain contradictory findings from the literature - specifically, secondary task difficulty and having learned mathematics primarily in China – and tested whether hypothesized effects emerge under these conditions.

Consistent with all previous literature, we found arithmetic performance to be generally slower and less accurate under cognitive load. However, contrary to our pre-registered predictions, we found no evidence for the moderating effect of secondary WM task load on arithmetic operations performance across the whole sample (Hypotheses 1a-1d) or any of our preregistered subgroup analyses, nor did we find evidence for these effects in the Chinese-educated participants (Hypotheses 2a-2d). In our follow-up Bayesian analyses, our results generally provided support for the null hypothesis that there was no moderation by secondary WM load on arithmetic operation over the additive effects of secondary WM task and arithmetic operation. However, the majority of Bayes factors suggested only anecdotal evidence for the null over the alternative, suggesting that these differential effects could still be real but that our current experiment was unable to find sufficient evidence otherwise. At best, we found anecdotal evidence of a verbal WM load effect on subtraction accuracy across some of our subsample analyses; however, the direction of this effect was the opposite of what was predicted by previous work. In sum, we did not find any evidence for the large strong crossover interaction reported by Lee & Kang (2002).

Interactions between secondary task types and arithmetic play a prominent role in the dual-task literature. These interactions have been interpreted as providing evidence that domain-specific pathways, such as verbal or visuospatial pathways have differential effects on numerical cognition (e.g. Ashcraft, 1992; Dehaene 1992; Dehaene & Cohen, 1995, 1997; see Chen and Bailey, 2021, for review). However, we argue that there are strong theoretical and empirical reasons to reexamine the robustness of these interactions. Several theoretical accounts of working memory argue against multiple domain-specific influences in favor of a more centralized executive processing system (Barrouillet & Camos, 2001; Cowan, 1999; Engle, 2002; Oberauer, 2009). Further, two other studies since Lee & Kang (2002) were also unable to replicate the key crossover interaction reported in the original paper (Cavdaroglu & Knops, 2016; Imbo & LeFevre, 2010). Imbo & LeFevre (2010) reported a differential effect in accuracy among Chinese students, such that there were more multiplication errors under verbal load compared to visuospatial load, but no differential effects of visuospatial vs. verbal load on subtraction performance. Both prior studies used a mix of within and between subject factors in their design. In comparison, our fully within-subjects study did not find any differential effects of WM load in our Chinese-educated participants nor across difficulty levels. To our knowledge, Lee & Kang (2002) remain the only study to have reported this crossover effect. Given the small sample size of the previous study (n = 10) and lack of subsequent replication, we propose that the field should consider the possibility that such crosstalk effects may be idiosyncratic to particular combinations of primary and secondary tasks and/or the particular population – irrespective of where they received their primary math education. Thus, crosstalk effects may be difficult to predict a priori. This view could certainly be replaced by a theory that can 1) account for when crossover interactions occur or not in the previous literature, and 2) make predictions about replicable effects in future work.

Additionally, our results conflict with parallel processing models of dual-task theory that attribute differences in dual-task performance to the amount of overlap in cognitive resources between two tasks (Navon & Miller, 2002; Tombu & Jolicœur, 2003; Wickens, 2008). While Pashler (1994) notes that crossover effects are still possible in the absence of parallel processing and may explain similar effects seen in sequential processing theories, the reasons for this may be specific to the combination of primary and secondary tasks and difficult to predict a priori (for another review see, Fischer & Plessow, 2015). For example, Hubber, Gilmore, & Cragg (2014) had participants complete addition tasks alone and with a visuospatial task (i.e. remembering patterns of colored blocks) and initially found that visuospatial memory moderated the types of strategies used in addition. The visuospatial task included an n-back component in which they had to remember if the target block was the same as the one presented before the previous block, so a follow up experiment was done in which a more static visuospatial task was used (i.e. without n-back component) and a separate central executive task was used (i.e. random letter generation). The follow-up found no difference in arithmetic performance or strategy selection between the single task condition and the dual-task with visuospatial load, but a major difference between the single task and dual-task with central executive load, suggesting that evidence of a parallel processing effect was confounded by the complexity of the secondary task. If dual-task performance relies on sequential processing, the current study still provides evidence for the effect of working memory on arithmetic performance, but the cost of performance caused by a potential cognitive bottleneck is likely more domain-general in nature than what is commonly assumed in the literature (for review see, Doherty et al., 2018).

To conclude, the current study investigated the differential effect of WM task loads (verbal and visuospatial) on arithmetic operations (multiplication and subtraction). Consistent with prior meta-analytic work on correlations between WM tasks and arithmetic performance and the dual-task literature on WM and arithmetic performance, the current study found consistent effects on arithmetic performance when under load of more complex secondary tasks, but no clear pattern for domain-specific interference. Despite investigating whether the crossover effect would emerge under conditions previously hypothesized to moderate the effect (difficulty and the system in which participants were educated), we did not find evidence for the predicted interaction in any of our analyses. Although multiplication and subtraction seemed to operate exclusively through verbal and visuospatial pathways, respectively, in the original study, this interaction has not been subsequently observed. We interpret these findings as evidence for a more domain-general pathway for WM secondary tasks' influence on numerical cognition, although we encourage future work that continues to carefully consider how theories of working memory and dual-task performance could explain previous domain-specific effects within numerical cognition.

Appendix

Table 1. Studies that tested arithmetic operation × WM load type interaction

Author	sample size	WM tasks	Arithmetic tasks	Multiplication effect (PL vs VSSP) *d*	Subtraction effect (PL vs VSSP) *d*	PL vs VSSP in Multiplication; Subtraction (ms)
Lee, K. M., & Kang, S. Y. (2002)	10	Repeat nonword string (PL), Matching abstract shapes and location (VSSP)	exact subtraction, exact multiplication	2.42	-3.31	1170 vs 996 993 vs 1271
Imbo, I., & LeFevre, J.A. (2010) – Canadian sample	57	Repeat nonword string (PL), 4×4 grid location task (VSSP)	two-digit subtraction, one × two-digit multiplication	0.04	0.04	5103 vs 5018 4823 vs 4738
Imbo, I., & LeFevre, J.A. (2010) – Chinese sample	73	Repeat nonword string (PL), 4×4 grid location task (VSSP)	two-digit subtraction, one × two-digit multiplication	-0.02	0.07	3015 vs 3038 3068 vs 2998
Cavdaroglu, S., & Knops, A. (2016)	32	Letter span (PL), 5×5 grid location task (VSSP)	2AFC multiplication (one × one; two × one digit), 2 AFC subtraction (two – one digit)	0.10	0.00	1015 vs 989 864 vs 863
Chen, E.H., Jaeggi, S.M., & Bailey, D.H. – Chinese-educated sample	22	Letter span (PL), 5×5 grid location task (VSSP)	2AFC multiplication (one × one; two × one digit), 2 AFC subtraction (two – one digit)	0.28	-0.09	883 vs 841 840 vs 851
Chen, E.H., Jaeggi, S.M., & Bailey, D.H. – other-educated sample	71	Letter span (PL), 5×5 grid location task (VSSP)	2AFC multiplication (one × one; two × one digit), 2 AFC subtraction (two – one digit)	0.05	-0.02	946 vs 939 841 vs 842

Note. Cohen’s *d* were calculated for the columns 5 and 6. Cohen’s *d* represents effect size between multiplication and subtraction RT performance under PL or VSSP load, respectively. PL – Phonological/Verbal. VSSP – Visuospatial. Reaction times rounded to nearest ms.

Table 2. Means and standard deviations for reaction time and accuracy as a function of WM task × arithmetic operation

WM task	**Arithmetic**	**RT Mean**	**SD**	**ACC Mean**	**SD**	**N**
No load	Multiplication	787	222	93%	11%	97
	Subtraction	713	213	95%	10%	97
Verbal	Multiplication	938	221	90%	11%	97
	Subtraction	850	234	91%	8%	97
Visuospatial	Multiplication	923	231	90%	9%	97
	Subtraction	854	276	93%	8%	97

Note. WM: working memory. Reaction times (RT) in nearest ms and accuracy (ACC) in nearest percentage

Table 3. Demographic information

Variable	**n=97**	**M(SD)**	**%**
Gender			
Male	33		66
Female	64		34
Age		20.1(1.3)	
Country of primary math education			
US	71		73.2
China	22		22.7
Other	4		4.1
Math grade compared to peer			
A	35		36.08
B	49		50.52
C	12		12.37
D	1		1.03
F	0		0
Abacus use			
Never Taught	69		71.13
Never Used	16		16.49
Rarely	11		11.34
Sometimes	1		1.03

Table 4. Planned ANOVA on arithmetic reaction time by model specification

	Whole sample				Chinese-educated			
Factor	F	df	p	η_p^2	F	df	p	η_p^2
PL vs VSSP × Multiplication	1.20	(1, 96)	.28	.01	1.69	(1, 21)	.21	.07
PL vs VSSP × Subtraction	.15	(1, 96)	.70	.002	.17	(1, 21)	.67	.01

Note. PL – Phonological/Verbal load, VSSP – Visuospatial load. Whole sample – no restriction on participants, Chinese-educated – only participants that reported their primary math education came from China.

Table 5. Planned ANOVA on arithmetic accuracy by model specification

	Whole sample				Chinese-educated			
Factor	F	df	p	η_p^2	F	df	p	η_p^2
PL vs VSSP × Multiplication	.49	(1, 96)	.49	.01	3.59	(1, 21)	.07	.15
PL vs VSSP × Subtraction	6.31*	(1, 96)	.01	.06	3.41	(1, 21)	.08	.14

Note. PL – Phonological/Verbal load, VSSP – Visuospatial load. Whole sample – no restriction on participants, Chinese-educated – only participants that reported their primary math education came from China. * $p < 0.05$

Table 6. Comparison of reaction time with no load arithmetic

Model			**Mean difference**	**SE**	**t**	**df**	$\mathbf{p_{holm}}$
Multiplication	Whole	Verbal	151	13.98	10.77	96	< .001
		Visuospatial	136	13.98	9.71	96	< .001
	Chinese	Verbal	160	34.45	4.65	21	< .001
		Visuospatial	118	34.45	3.42	21	0.014
Subtraction	Whole	Verbal	137	13.98	9.77	96	< .001
		Visuospatial	141	13.98	10.12	96	< .001
	Chinese	Verbal	94	34.45	2.74	21	< .001
		Visuospatial	106	34.45	3.07	21	< .001

Note**.** Mean difference is reaction time in nearest ms. p_{holm} – p value after Holm-Bonferroni correction

Table 7. Comparison of accuracy with no load arithmetic

Model			**Mean difference**	**SE**	**t**	**df**	**p_{holm}**
Multiplication	Whole	Verbal	-3%	1	-4.68	96	< .001
		Visuospatial	-3%	1	-3.99	96	< .001
	Chinese	Verbal	-4%	2	-2.67	21	0.14
		Visuospatial	-2%	2	-1.39	21	1
Subtraction	Whole	Verbal	-4%	1	-5.23	96	< .001
		Visuospatial	-2%	1	-2.64	96	0.07
	Chinese	Verbal	-2%	2	-1.22	21	1
		Visuospatial	0%	2	0.22	21	1

Note**.** Mean difference is accuracy in nearest percentage. p_{holm} – p value after Holm-Bonferroni correction

Table 8. Bayesian model comparisons of 2 × 2 ANOVA on reaction time

	Whole		**Chinese**	
Models	BF$_{10}$	error %	BF$_{10}$	error %
Arithmetic	3.80e +6	0.91	0.25	1.33
WM task + Arithmetic	4.50e +5	1.78	0.06	1.42
WM task + Arithmetic + WM task ✻ Arithmetic	9.60e +4	4.67	0.03	2.38
WM task	0.12	0.79	0.26	1.92

Note. All models include subject. Null model is used as reference.

Table 9. Bayesian model comparisons of 2 × 2 ANOVA on accuracy

	Whole		**Chinese**	
Models	BF_{10}	error %	BF_{10}	error %
Arithmetic	159.70	1.11	0.31	0.85
WM task + Arithmetic	106.89	3.19	0.49	1.73
WM task + Arithmetic + WM task ✻ Arithmetic	28.05	2.00	0.15	2.86
WM task	0.59	0.90	1.58	1.68

Note. All models include subject. Null model is used as reference.

Table 10. Bayesian model comparisons of 3 × 2 ANOVA on reaction time

	Whole		**Chinese**	
Models	BF_{10}	error %	BF_{10}	error %
WM task	4.0e +26	0.87	808.66	0.81
Arithmetic	4.19e +7	0.73	0.19	2.02
WM task + Arithmetic	1.3e +37	1.02	153.86	2.02
WM task + Arithmetic + WM task ✻ Arithmetic	6.8e +35	3.51	32.60	3.00

Note. All models include subject. Null model is used as reference.

Table 11. Bayesian model comparisons of 3 × 2 ANOVA on accuracy

	Whole		**Chinese**	
Models	BF_{10}	error %	BF_{10}	error %
WM task	1.61e +5	0.92	1.21	0.86
Arithmetic	1868.54	1.21	0.18	1.26
WM task + Arithmetic	6.13e +8	2.00	0.22	1.34
WM task + Arithmetic + WM task ✻ Arithmetic	4.41e +7	0.90	0.05	2.12

Note. All models include subject. Null model is used as reference.

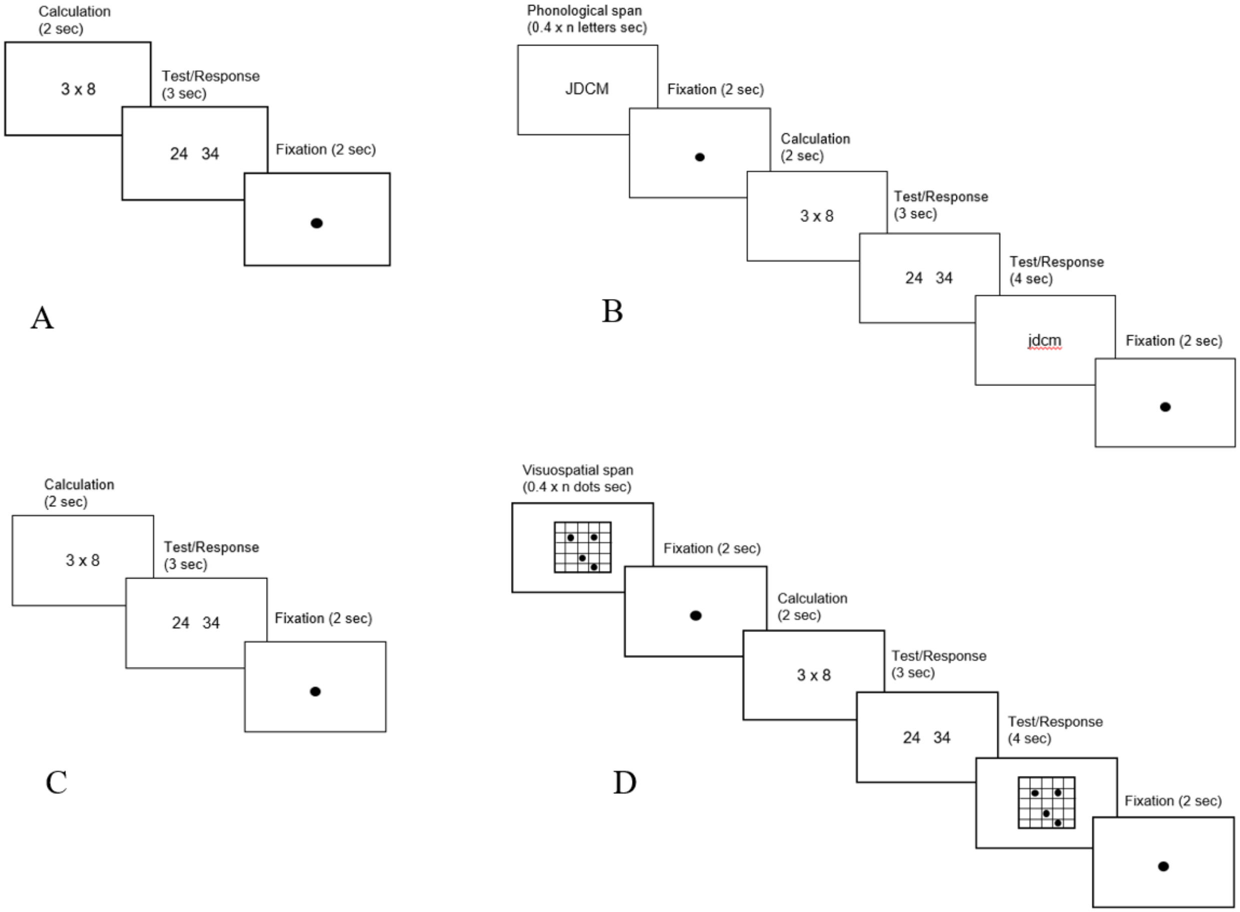


Figure 1. Single multiplication task (A & C). Dual-task multiplication with phonological letter WM load (B) and visuospatial WM load (D).

Figure 2. Comparison of dual-task effects in reaction times across subsample analyses. A = Whole group, B = Received majority of math education in China, C = Easy load conditions only, D = Hard load conditions only, E = First under cognitive load conditions only. Error bars represent standard errors of the mean.

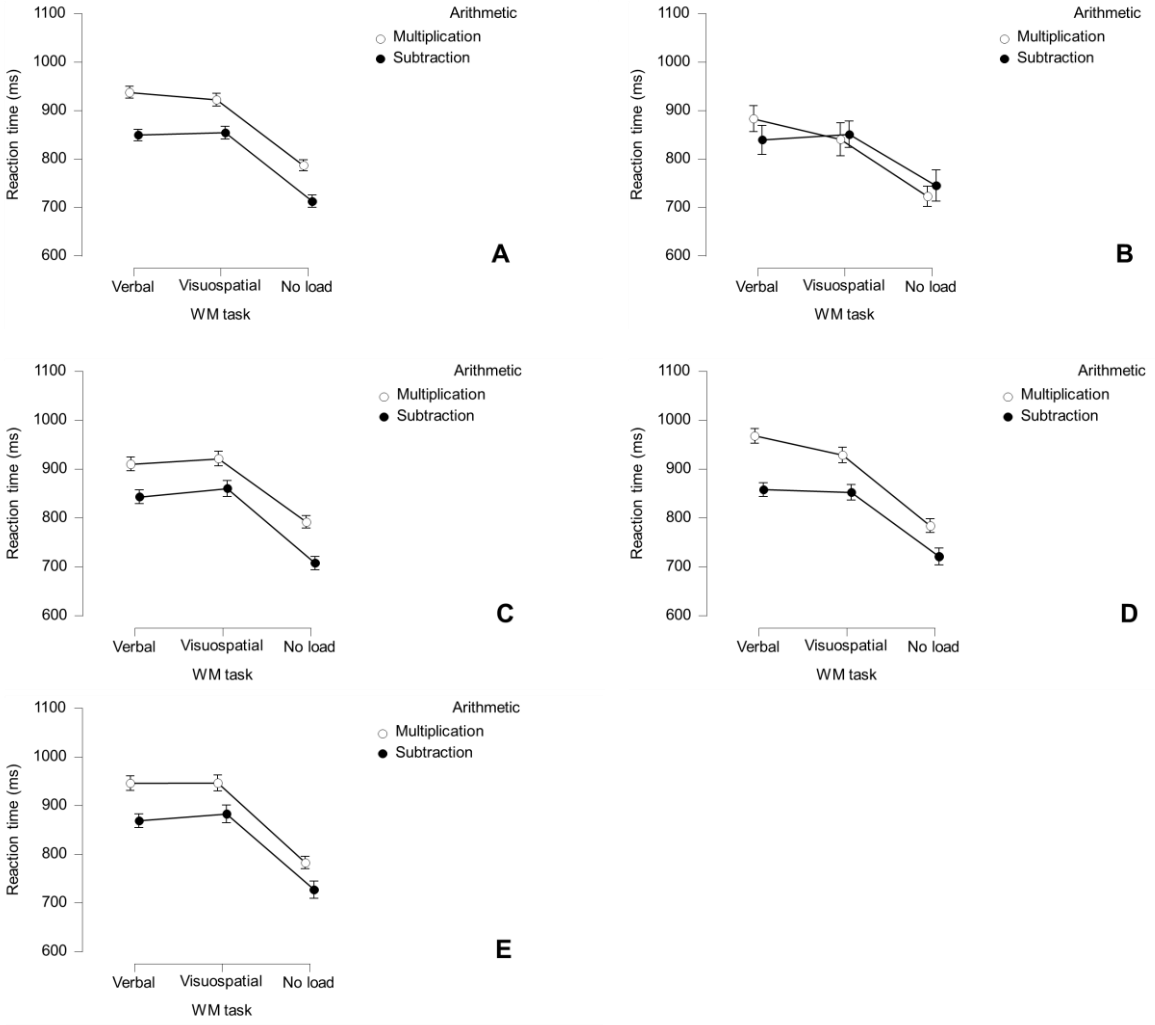
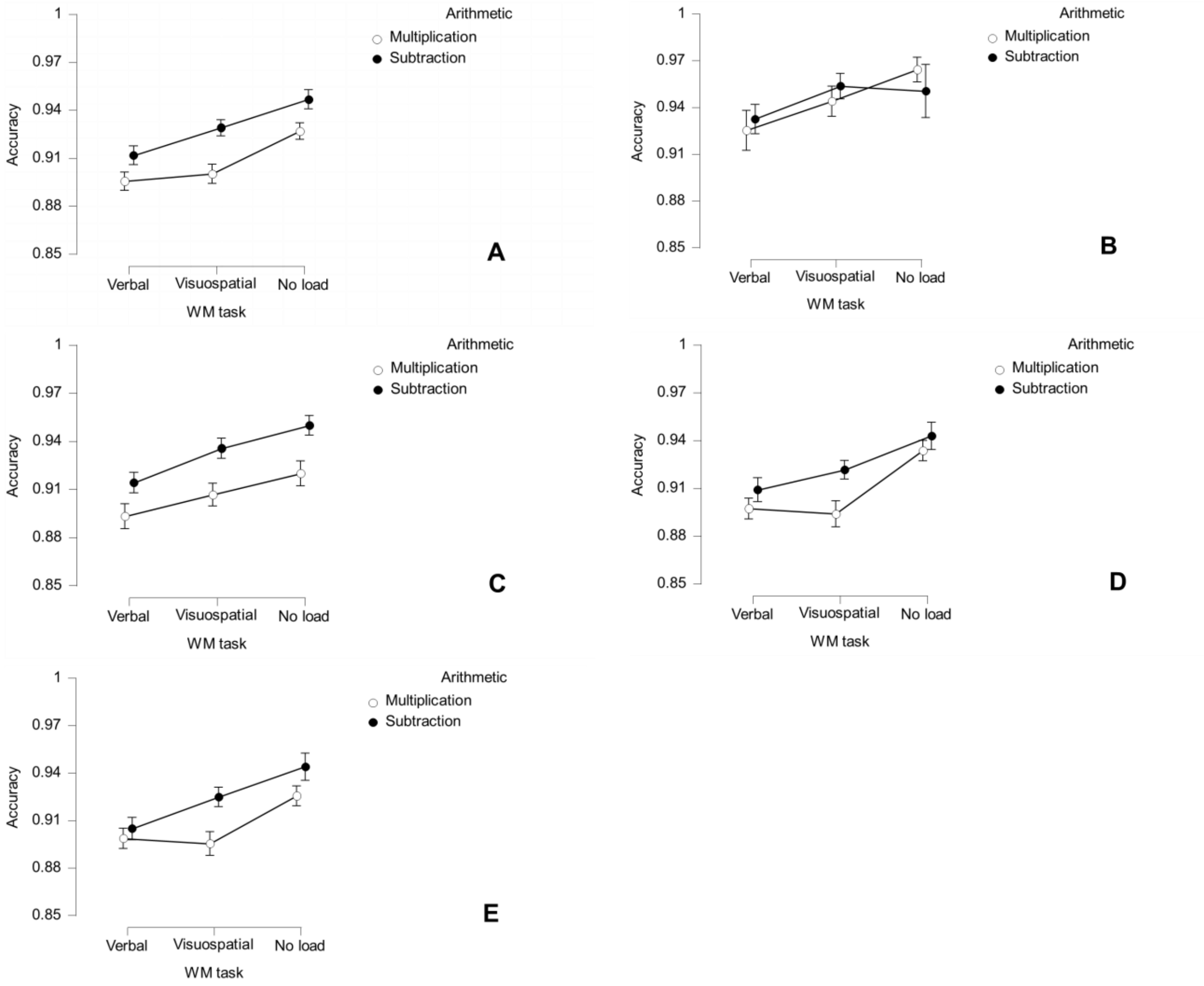


Figure 3. Comparison of dual-task effects in accuracy across subsample analyses. A = Whole group, B = Received majority of math education in China, C = Easy load conditions only, D = Hard load conditions only, E = First under cognitive load conditions only. Error bars represent standard errors of the mean.



Study 3

Domain-general or domain-specific? Working memory influences in a dual-task arithmetic experiment

Domain-general or domain-specific? Working memory influences in a dual-task arithmetic experiment

An important question within cognitive psychology concerns how cognitive processes are employed to carry out multiple tasks simultaneously. Extensive research finds that performance under multiple tasks declines in a range of seemingly easy laboratory tasks compared to when tasks are performed alone (e.g. Pashler, Johnston, & Ruthruff, 2001). Two competing hypotheses have been raised in discussion of how our cognitive architecture is utilized in these dual-task situations. The more widespread and well-accepted cognitive bottleneck model (Pashler, 1984, 1994; Welford 1952) assumes that only information from one task can be processed at a time. Thus, while Task 1 is processing, Task 2 must wait until space is available. Others have suggested a parallel processing model in which different tasks require the same limited resource, which can be shared between tasks to allow for parallel – but less efficient – processing (Navon & Miller, 2002; Tombu & Jolicoeur, 2003). A central argument of this hypothesis is that content-dependent characteristics of Task 2 can influence performance in Task 1; although others have suggested that interference effects on Task 1 can also be explained through a cognitive bottleneck (Strobach, Schütz, & Schubert, 2015). In addition, while performance is often hypothesized to slow down in dual-task designs, some cases in which pairs of primary and secondary tasks may also reduce the interference effect by removing task-switching costs (Park, Kim, & Chun, 2007).

While both models have intuitive arguments to explain how working memory resources are used to complete cognitive tasks, research on working memory and arithmetic has used the paradigm as a method for testing test specific hypotheses about variation in the cognitive demands of different arithmetic tasks, with less focus on the implications of these findings for theories of human cognition more broadly. These hypotheses largely rely on the latter parallel processing hypothesis, and specifically the assumption that when primary and secondary tasks require the same specific cognitive demands (e.g., verbal or visuospatial working memory capacity), arithmetic performance will be impaired above and beyond what would be expected based on the general cognitive demands of the secondary task.

Domain-specific theories in dual-tasks

The dual-task arithmetic literature has been largely predicated on the influence of domain-specific storage systems within working memory models, such as a visuospatial and verbal memory storage (Baddeley & Hitch, 1974; Baddeley, 2000; Logie, 2016). These multicomponent or "modular" models have contributed to specific predictions regarding the use of content-dependent memory stores within various arithmetic operations (e.g., Dehaene & Cohen, 1997). Principal of these predictions is that characteristics from one task (namely a working memory task) may influence performance on another (e.g. mental arithmetic) through interference in encoding and processing stages of retrieval. Specifically, this has been referred to as a crosstalk effect, wherein cognitive tasks that share similar features impede performance on dual-task measures more compared to tasks that are more dissimilar or require different encoding mechanisms (Pashler, 1994). Conversely, there should also be conditions where storage and processing can run in parallel, with little conflict between the content demands. In support of this hypothesis are studies which find no significant conflict between storage and processing tasks. For example, phonological secondary tasks have been found to not impair performance on single-digit multiplication problems, particularly for easy multiplication problems (i.e., both operands less than five) (De Rammelaere et al., 2001; Seitz & Schumann-Hengsteler, 2000, 2002), which has been interpreted as evidence that access to permanent information in long-term memory does not require intervention by the working memory subsystems. However, phonological load has been found to interfere with the performance of single-digit multiplication in Korean speaking university students (Lee & Kang, 2002), who may rely on phonological codes when storing and accessing multiplication facts compared to individuals taught in other languages (e.g. English and German) in which the counting systems vary significantly from some Asian languages (DeStefano & LeFevre, 2004). Related research suggests that Chinese-speaking individuals store and access multiplication facts using phonological codes of their math education's strong emphasis on rote memorization and verbal repetition (Imbo & LeFevre, 2010; LeFevre, Lei, Smith-Chant, & Mullins, 2001; LeFevre & Liu, 1997).

According to this position, an important condition for minimizing the conflict between storage and processing is that the working memory capacity of the components should not be overloaded. As suggested by Logie (2011), a reduction in dual-task performance compared with single-task performance may be observed when domain specific components are pushed beyond their capacity. However, a domain-general executive or attentional system may support the performance of domain-specific resources leading interference effects to be interpreted as stemming from the central executive resources rather than the subcomponents of working memory. Therefore, when assessing the overlap of storage and processing in working memory, it is important to ensure that individual task demand is controlled (Doherty & Logie, 2016; Cavdaroglu & Knops, 2017).

General capacity theories in dual-tasks

Juxtaposed with the influence of subsystems within domain-specific theories of working memory are explanations that focus on domain-general influences of completing a secondary task on primary task performance. These explanations have been a major focus within the dual-task literature generally, but interestingly not in dual-task arithmetic literature specifically. When considering dual-task performance, executive functioning or embedded models of working memory place more emphasis on domain-general processes, such as inhibition and task-switching over the influence of any one or multiple content-dependent storage system (Miyake et al., 2000; Engle, 2002; Cowan, 1999; Barrouillet & Camos, 2020; Oberauer, 2009). That is, storage and processing functions of working memory rely in part on a shared, general purpose, limited capacity attentional resource. Dual-task performance is thus constrained by a bottleneck in which cognitive operations take place one at a time. Processing is therefore limited by attention more so than cognitive overlap, and in some cases, may help explain why individual influences of domain-specific storage components on arithmetic may be small. Indeed, it may be the case that these crosstalk effects are much more idiosyncratic to the design of the experiment than previously thought. A replication attempt at a previously large crosstalk effect was unable to detect any differential effects of phonological and visuospatial memory on multiplication and subtraction across a number of different analytical methods (Chen, Jaeggi, Bailey, 2022).

There is a substantial dual-task literature separate from the dual-task arithmetic literature that focuses on domain-general focused predictions and explanations within dual-task performance, especially the roles of attention and task demand. For example, when attention is occupied by processing, it is no longer available for maintaining memory traces and so these traces suffer from temporal decay and interference. However, decayed memory traces may be restored through attentional refreshing when attention is available during pauses in processing. While temporary verbal memory can be bolstered by subvocal rehearsal in a phonological loop, performance is highly dependent on access to attention. A number of empirical observations have demonstrated how the demand of a secondary processing task is inversely correlated with memory performance in a dual-task complex span paradigm (Barrouillet et al., 2015; Uittenhove et al., 2019). This secondary task demand or "cognitive load" might be measured as the proportion of time the processing task captures attention and therefore diverts the focus away from maintenance of temporary memory traces. Further, dual-task studies of arithmetic have consistently found the central executive crucial to the ability to perform mental arithmetic (Chen & Bailey, 2020; Imbo, Vandierendonck, & De Rammelaere, 2007; Seitz & Schumann-Hengsteler, 2000; 2002).

Attentional refreshing, the specific process that is interrupted by high cognitive load tasks, distinguishes itself from phonological rehearsal in one major aspect; it is dependent on central executive resources and therefore amodal (Camos, Lagner, & Barrouillet, 2009; Camos, Lagner, & Loaiza, 2017). That is, while phonological rehearsal relies on verbal resources, refreshing can be done with any form of information. The serial processing model further states that refreshing can either be actively or passively engaged depending on whether subvocal rehearsal is available or more effective - given task parameters - or indeed whether participants are instructed to rehearse or refresh (Camos, Mora, & Oberauer, 2011). In the same way as processing prevents refreshing, refreshing activities postpone processing, as observed in Vergauwe, Camos, and Barrouillet (2014) when the slowing of processing task responses increased with memory loads. It is important to note that this effect occurs only when the phonological loop is unavailable or when its capacity is exceeded. The same study by Vergauwe et al. (2014) provided evidence contrary to the existence of a storage system for visuospatial information where visuospatial information was not maintained by any domain-specific storage system and so its maintenance relied entirely on attention (Morey & Bieler, 2013; Morey, Morey, van der Reijden, & Holweg, 2013). It should be noted, however, that participants are likely to employ different strategies across cognitive tasks, such as with the use of the attentional refreshing for letters and digits rather than subvocal rehearsal (Logie, 2018).

Given these findings, secondary WM tasks may not interfere with arithmetic performance via some content-dependent pathway, but rather through an attentional pathway, especially when the cognitive demands of the secondary task are large and take up "too much space" or too many cognitive resources in working memory. In such cases where cognitive demands from both tasks are high, we are unlikely to see any kind of selective interference in arithmetic by domain-specific storage systems. That is to say that there is no evidence to suggest domain-specific systems do not influence arithmetic but that their influence may not be as strong as domain-general resources in much of the of dual-task arithmetic literature.

Strategy choice in arithmetic

Working memory is believed to influence arithmetic solving via the selection and execution of solving strategies (Hecht, 2002; Hubber, Gilmore, & Cragg, 2014; Imbo & Vandierendonck, 2007). There are a number of ways in which the solution to an arithmetic problem can be reached. For example, if given the problem 5 + 7 = ?, an individual could select from several different strategies including a) directly retrieve the answer from long-term memory (*retrieval*), b) decompose the problem into a series of simpler problems, e.g. 5 + 5 = 10, 10 + 2 = 12 (*decomposition*) or c) count on seven times from 5 (*counting*) (Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Siegler & Shrager, 1998). The latter two strategies are termed as procedural strategies. Working memory is likely to be required to a greater extent for procedural as compared to retrieval strategies. This characterizes procedural strategies for arithmetic as a way for individuals to store interim solutions or the number of count steps performed so far, while carrying out other procedural steps. This is not required for retrieval strategies, which only involve the single step of retrieving the solution from long-term memory.

Drawing on the Baddeley & Hitch multicomponent model, dual-task experiments have investigated the extent to which domain-general (i.e. central executive) and domain-specific (i.e. PL and VSSP) resources contribute to strategy choice and execution in adults and children (Cragg et al., 2017; Hecht, 2002; Hubber, Gilmore, & Cragg, 2014; Imbo & Vandierendonck, 2007). Across these studies, loads from domain-general resources via central executive secondary tasks are consistently involved in strategy choice and execution, particularly in the use of procedural strategies. On the other hand, while some studies suggest phonological and visuospatial resources may contribute to strategy use (Hubber, Cragg, & Gilmore 2014; Imbo & Vandierendonck, 2007), these may be confounded by executive resources. Thus, it remains an open question about whether the effect of a secondary task on strategy choice and execution in arithmetic operations are a function of secondary task general demands only or also primary and secondary task similarity.

Current Study

The current study is part of a larger experiment that includes a registered report examining differential dual-task interference effects on arithmetic operations, and thus includes language and details from the registered report. Empirical evidence may not fit either parallel processing or cognitive bottleneck predictions perfectly, so the purpose of this study was to reconcile the findings from Chen & Bailey (2021) and prior findings of a differential working memory effect on arithmetic operations. We do this by investigating the combined effects of task demands and task similarity in a controlled experiment using span tasks assessing the phonological loop, visuospatial sketchpad, and direct arithmetic processing to influence the performance of simple multiplication and subtraction. Each of the proposed theoretical models previously discussed predict different dual-task performance, so these are broken down as follows: (1) parallel processing (crosstalk) effects (2) cognitive bottlenecking effects and (3) integrated interaction effects (i.e. task similarity and task demands) (for summary see, Table S1 in the Appendix). In a second exploratory analysis, we investigated the role of working memory loads on strategy choice. Specifically, we will explore which strategies young adults select for simple arithmetic when under varying kinds of cognitive loads and difficulties.

Predictions following from domain-specific theories

Previous research has hypothesized separate domain-specific encoding processes – one verbal and one visuospatial – for multiplication and subtraction, respectively (Lee & Kang, 2002; Dehaene et al. 2003). In order to understand the extent to which crosstalk affects arithmetic performance, this study will also test arithmetic procedures as secondary tasks. Specifically, we include a repeated addition task that is likely to negatively affect the performance of both multiplication and subtraction due to the nature of the task sharing similar arithmetic processing. Following the crosstalk hypothesis, multiplication performance should be negatively affected by the storage and processing of phonological secondary tasks with no effect of visuospatial secondary tasks. Contrastingly, subtraction performance will be negatively affected by visuospatial secondary tasks but not letter span tasks. With the inclusion of the repeated addition task, we predict greater dual-task interference in multiplication than subtraction, because the repeated addition task constrains a phonological component, and repeated addition is commonly employed as a subroutine of multiplication. Thus, we predict the following for the domain-specific model:

Multiplication and subtraction are differentially affected by crosstalk from tasks that share similar cognitive resources.

Specific model prediction A: Multiplication performance is most impacted by the repeated addition task followed by the PL task and then the VSSP task.

Specific model prediction B: Subtraction is impacted most by the repeated addition and VSSP tasks than the PL task with there being no difference between the repeated addition and VSSP tasks.

Predictions following from domain-general theories

Compared to the hypothesis that only domain-specific overlap results in reduced primary task performance, general capacity theories predict that multiplication and subtraction will not be differentially affected by the type of secondary memory task; rather, performance on both arithmetic types decreases as the difficulty (cognitive load) of the secondary task increases. We will include two different performance conditions in which participants perform arithmetic under an easy and hard cognitive load. When task difficulty is equated across conditions, we should expect no difference in arithmetic performance across the different secondary tasks. However, when the required span increases in size, we should expect decreased performance in the harder span trials than in the easier span trials independent of any overlapping resources. Importantly, the easier span conditions will still incur deficits in performance but not as much as the hard conditions. Thus, we predict the following for the domain-general model:

Multiplication and subtraction performance will not be differentially affected by the type of secondary memory task; rather, performance on both arithmetic types decreases as the difficulty (cognitive load) of the secondary task increases.

General model prediction A: Multiplication and subtraction under difficult loads lead to worse performance compared to arithmetic under easy loads.

General model prediction B: The interactions between PL, Repeated addition, and VSSP dual-task loads with arithmetic type (multiplication or subtraction) are null.

Predictions from an integrated model

Both domain-general and domain-specific models draw from diametrically separate working memory structures leading to conflicting predictions. In the former case, working memory is viewed amodally with little to no influence of subsystem involvement (e.g. Engle, 2002 or Cowan, 1999). In the contrast, the latter proposes a more modular account of working memory with greater emphasis to the influence of subsystems (e.g. Logie, 2016). Both accounts may fit the data to the extent that there is some influence of cognitive overlap as well as general task demands derived from working memory tasks that influence arithmetic operations. To this end, we predict a more integrated model for dual-task interference where the influence of task demands and cognitive overlap is additive. The interaction between task demands and cognitive overlap are still in line with the modular, domain-specific model proposed by Baddeley & Hitch (1974) in which the central executive largely controls the processing of information between subsystems. Thus, we predict the following for the Integrated 1 model:

Integrated 1 model prediction A: Multiplication is impacted most by the hard repeated addition task, because the secondary task is difficult and the cognitive demands of the tasks overlap; followed next by easy repeated addition and hard PL tasks, because the repeated addition task overlaps more than the PL task despite its difficulty; easy PL and hard VSSP because the VSSP task is difficult but is not predicted to overlap with multiplication; and least impacted by the easy

VSSP task because it is neither as difficult as other tasks nor does it share any overlap in resources.

Integrated 1 model prediction B: subtraction is impacted most by hard repeated addition and hard VSSP because the secondary task is difficult and the cognitive demands of the tasks overlap; followed by easy repeated addition, easy VSSP, and hard PL because the PL task is difficult but is not predicted to overlap with subtraction as much as the other secondary tasks; and least impacted by the easy PL task because it is neither as difficult as other tasks nor does it share any overlap in resources.

Alternatively, perhaps the effects of secondary tasks depend on *interplay* between the difficulty of the secondary task and the overlapping cognitive demands between the primary and secondary tasks. Specifically, perhaps difficulty effects are more pronounced when there is also cognitive overlap. For example, easy verbal load may not differentially influence multiplication and subtraction problems but multiplication may be much slower than subtraction under hard verbal load. Thus, we predict the following for the Integrated 2 model:

Integrated 2 model prediction A: multiplication is most impacted by hard repeated addition because this secondary task has the most overlap under difficult conditions; followed by hard PL and then by hard VSSP because PL tasks have more overlap than VSSP with multiplication; and least impacted by easy PL, easy repeated addition, and easy VSSP because the easier tasks are not producing a large enough cognitive load to see meaningful differences.

Integrated 2 model prediction B: subtraction is most impacted by hard repeated addition and hard VSSP because these secondary tasks have the most overlap under difficult conditions; followed by hard PL because PL tasks have less overlap with subtraction; and least impacted by easy because the easier tasks are not producing a large enough cognitive load to see meaningful differences.

The stronger influence of central executive loads over phonological and visuospatial loads found by Chen & Bailey (2021) suggests another set of alternative predictions where the effect of working memory load types may only be present when load difficulty does not overly constrain executive resources. Perhaps interplay between difficulty and cognitive overlap affects performance in the opposite way, such that crosstalk effects are more pronounced at lower difficulty levels, because higher difficulties constrain central executive resources instead of domain-specific resources. Higher difficulty tasks will more negatively affect performance than easy secondary tasks, but crosstalk effects are only visible under easier loads. That is, an effect of visuospatial load on subtraction over multiplication may only be present in easy load conditions, whereas the effect of specific working memory loads disappear in hard loads due to a shift in using more executive functioning resources. Thus, we predict the following for the Integrated 3 model:

Integrated model 3 prediction A: multiplication performance is impacted more by easy PL and easy repeated addition compared to VSSP load because these secondary tasks share more overlap with multiplication, but all are negatively impacted similarly by harder secondary tasks because the secondary task load is too high to see meaningful differences.

Integrated model 3 prediction B: subtraction performance is impacted more by easy VSSP compared to both easy PL and easy repeated addition because the VSSP task shares more overlap with subtraction, but all are negatively impacted similarly by harder secondary tasks because the secondary task load is too high to see meaningful differences.

Methods

Participants

Power analysis

We used the software program G*Power to conduct an a priori power analysis (Faul et al., 2007, 2009). The outline for the power analysis can be found in the Supplementary Files. Following our power analysis, we recruited and ran 100 total undergraduate participants from a large Western university (Female = 64, age range = 18 – 25 years old, mean = 20.1 (*SD* = 1.3). 22 participants in the final analysis sample reported receiving the majority of their math education in China prior to enrolling in a US university. All participants had normal or corrected-to-normal vision. All research was performed in accordance with the ethical standards of the Institutional Review Board. Written informed consent was obtained from all participants and were given course credit through the Human Subjects Lab Pool or were reimbursed $30 for their participation. Participants from this study are the same as those from Chen, Jaeggi, & Bailey (2022). To our knowledge, this is the largest study with a fully within-subjects design of a dual - task arithmetic experiment.

Stimuli

All tasks used in these experiments were created through PsychoPy 3 (Peirce et al., 2019). Performance on the span tasks and arithmetic were measured by reaction time (RTs in ms) and accuracy (ACCs in percentage correct). For examples, see Figures 1 & 2. Arithmetic problems used in this experiment are the same as in Cavdaroglu & Knops (2016) and Chen, Jaeggi, & Bailey (2022). Phonological and visuospatial secondary tasks were based on the descriptions used in Cavadaroglu & Knops (2016). Strategy reports were measured with a one-item survey question at the end of each block. All materials including experimental tasks and protocol used are available online as supplementary materials.

Arithmetic

Multiplication and subtraction problems were the same as those used in Chen, Jaeggi, & Bailey (2022). Technical details for tasks can be found in the supplementary materials.

General staircase procedure

Each working memory task (i.e., phonological, visuospatial, numeric) followed a similar staircase procedure. After three correct responses in a row, the difficulty of the task increased by 1 letter/dot/addend; otherwise, if there were three consecutive incorrect responses, the difficulty of task decreased by 1 letter/dot/addend until the minimum number of stimuli were reached or until a correct response was given. 30 trials were conducted to measure each working memory task's span. In addition, a Weibull function was fit to the data where the inverse of the Weibull function was used to determine the number of letters corresponding to 80% and 99% accuracy. The two threshold levels were chosen to examine the effect of task difficulty (low vs high) on arithmetic performance in both single- and dual-task conditions. In total, each staircase contained 30 trials for a total of 90 trials. The phonological and visuospatial tasks were the same as those in Chen, Jaeggi, & Bailey (2022), so their technical details were moved to the supplementary materials.

Repeated addition staircase

We created a repeated addition task to measure the crosstalk effects of a secondary arithmetic task on the primary arithmetic operations (i.e. multiplication and subtraction). Participants' repeated addition ability was measured using an adaptive staircase procedure of addition problems. Following a 2AFC choice design like our primary measures, participants were shown a short addition problem and then later instructed to choose between two answer choices (shown 7s after onset of problem). Problems were displayed for a duration of 0.4 s × n – n indicating number of addends – followed by 3 s on a fixation screen before participants are given 4 s to respond. No response trials were counted as incorrect. In half of the trials, the correct response was on the left and on the right in the other half. The ‘F’ and ‘J’ keys were used for responding. The task started with 3 addends and reach a maximum of 9 and minimum of 2. The adaptive staircase procedures were the same as those in the previous two tasks. In order to minimize extraneous cognitive load, the possible addends were kept between numbers 1-5 (e.g. 2 + 2 + 3 + 1; 5 + 2 + 1 + 1 + 4). Alternative answers were randomized to be between 1-5 units away from the answer to maintain a consistent distance effect (e.g. 8 vs [3,4,5,6,7,9,10,11,12,13]). Trials in which only a single number is repeated were removed to avoid multiplication strategies.

Strategy reports

The strategy report was a one-item survey asking participants to select from four strategy choices (Retrieval, Counting, Decomposition, Mixed/Other). The question included a description of each strategy as well as an example. It is presented at the end of each block. Participants who select Mixed/Other are given an opportunity to type a description of their strategy. Although it is not ideal to ask participants to reflect on their strategies after having completed several problems instead of immediately after a particular problem, our interest in the effects on strategy use were secondary, and we wanted to avoid the possibility of strategy reports influencing task difficulty.

Procedure

The study used a 2 × 2 × 4 factorial design using within-subject factors. The within-subject factors were arithmetic operation type (subtraction or multiplication), load difficulty (easy or hard), and WM load type (no load, PL load, VSSP load, repeated addition load). No load (i.e. arithmetic alone) conditions served as controls against dual-task conditions. The entire experiment was conducted online through video conferencing in which an experimenter guided the participant in downloading the requisite materials and protocol for completing experimental tasks. The experiment was administered within two sessions that were scheduled to be around the same time and spread apart by 1 week. In session 1, participants completed the staircase trials to determine difficulty levels for dual-task conditions and answered demographic questions. In session 2, participants completed the full dual-task experiment including strategy reports at the end of each block. Each block consisted of a combination of single and dual-task conditions (e.g., multiplication-no load, multiplication-easy phonological load, subtraction-easy phonological load, etc.). Additional details for procedures can be found in the supplementary materials.

Analysis Plan

In order to test each set of models and their respective hypotheses against each other, we computed the difference between dual-task and single-task performance for both multiplication and subtraction under each load and difficulty condition using a 2 × 4 × 2 repeated-measures ANOVA with all factors within-subjects. As a complement to the frequentist analyses and to address our testing of a null hypothesis, we also reported Bayesian analyses of our repeated measures ANOVA to examine the relative support for both the different model predictions against the null hypothesis. Details for our Bayesian analyses can be found in the supplementary materials. For the manipulation check: Single arithmetic performance was compared to arithmetic under load using a one-tailed paired-samples t-test. It is expected that primary arithmetic task performance (multiplication and subtraction) is negatively impacted while under cognitive load of a secondary task compared to performing arithmetic alone.

The domain-general model argues that differences in dual-task performance come from the demands of the secondary tasks rather than overlapping resources. If both hypotheses from this model are supported by our frequentist and Bayesian analyses instead of any other model (i.e. Domain-specific, Integrated 12,3), there would be strong support for the domain-general model. That is, if ANOVA model only yields significant main effects for task features (i.e., difficulty, arithmetic, and/or secondary task) but contained no 2- or 3-way interactions, our findings would support a domain-general model. If only one or parts of the model's hypotheses are supported by our findings, we could only conclude that there is mixed support for this model. If none of the hypotheses are supported in any way, we would conclude that there is little to no support for the domain-general model in dual-task arithmetic experiments.

On the other hand, the domain-specific model predicts slower, less accurate arithmetic performance under cognitive loads from secondary tasks that share similar cognitive resources. If both hypotheses from this model are supported by our frequentist and Bayesian analyses instead of any hypotheses from the other models, there would be strong support for the domain-specific model. That is, if the ANOVA model detected a significant 2-way interaction specifically between secondary task types and arithmetic operation but no 3-way interaction with difficulty, our findings would support a domain-specific model. If only one or parts of the model's hypotheses are supported by our findings, we could only conclude that there is mixed support for this model. If none of the hypotheses are supported in any way, we would conclude that there is little to no support for the domain-specific model in dual-task arithmetic experiments.

The integrated models predict interactions between task demands and task similarity. However, each set of predictions are distinct from one another. The Integrated 1 model implies an additive effect of difficulty and crosstalk. The Integrated 2 model suggests a moderation effect wherein difficulty matters only when there is also crosstalk. Finally, the Integrated 3 model suggests performance at higher difficulties are more constrained by task demands while crosstalk is more likely in lower difficulties. The same interpretations used for the domain-general and domain-specific models apply for the three integrated models as well. That is, if there was a significant 3-way interaction between secondary tasks, arithmetic operation, and difficulty in addition to significant main and 2-way effects, our findings would suggest evidence for one of the integrated models. If only one or parts of an integrated model's hypotheses are supported by our findings, we could only conclude that there is mixed support for that model. If none of the hypotheses are supported in any way, we would conclude that there is little to no support for an integrated model in dual-task arithmetic experiments.

Data handling

Data were analyzed primarily in JASP using its frequentist and Bayesian repeated measures ANOVA and paired-sample t-test functions (JASP Team, 2020). Data were organized for JASP using RStudio (RStudio Team, 2020), specifically tidyverse for data visualization and formatting (Wickham et al., 2019). The RMarkdown is available as supplementary material to reproduce data created for JASP. Where appropriate, Holm-Bonferroni correction was used to correct for multiple comparisons in post-hoc testing (Holm, 1979). Huynh–Feldt correction was used when sphericity was violated. Bayesian analyses were conducted using the Bayesian repeated measures ANOVA function in JASP (JASP Team, 2020). Following Morey & Rouder (2011), we will set a non-informative Jeffreys prior width of 0.5 to correspond to a small effect.

All reaction time (RT) analyses were based on correct trials only. Accuracy or response times outside the range of a participant's mean ± 3 SDs were discarded from further analyses. Responses faster than 200 ms were also discarded. Based on that criterion, 1.02 % of trials in single arithmetic blocks and 3.56 % of the trials in dual-task blocks were eliminated. In addition, 3 participants were excluded from data analyses for not responding in a majority of trials during the second session. 2 more participants were excluded from our primary and secondary analyses because they were missing data for one of the load conditions. Demographic information can be found in Table 1. All data are publicly available on the OSF page (https://osf.io/6egt5/).

Results

In order to test each model's predictions against each other, we evaluated the overall 2 × 4 × 2 repeated measures ANOVA on the basis of whether main effects and/or interactions were present. If the ANOVA model only yielded significant main effects for difficulty, arithmetic, and/or secondary task but contained no 2- or 3-way interactions, our findings would support a domain-general model. If the ANOVA model detected a significant 2-way interaction specifically between secondary task types and arithmetic operation but no main effects or 3-way interaction with difficulty, our findings would support a domain-specific model. Finally, if there was a significant 3-way interaction between secondary tasks, arithmetic operation, and difficulty, our findings would suggest evidence for one of the integrated models.

Overall, we found strong support for a domain-general model for reaction times and a bit more mixed support for accuracies such that performance was generally worse under all dual-task conditions compared to single-task arithmetic (Figure 1 & Table 2). In addition to the main effect of secondary task load [RT: F = 111.36, $p < .001$, $\eta_p^2 = 0.54$; ACC: F = 66.34, $p < .001$, $\eta_p^2 = 0.41$], the ANOVA yielded significant main effects for arithmetic operation [RT: F = 18.72 $p <$

.001, $\eta_p^2 = 0.17$; ACC: F = 4.07, $p = .047$, $\eta_p^2 = 0.04$] and difficulty for reaction times but not accuracy [RT: F = 4.94, $p = .03$, $\eta_p^2 = 0.05$; ACC: F = 2.99, $p = .09$, $\eta_p^2 = 0.03$].

Arithmetic performance was generally slower and less accurate under greater secondary load conditions or when paired with addition as a secondary task compared to arithmetic alone (Table 3). No significant 2- or 3-way interactions [i.e., WM × Arithmetic Operation; WM × Difficulty; Difficulty × Arithmetic Operation; WM × Arithmetic Operation × Difficulty] were detected for reaction times [*p*-values range: .06 - .89]. However, a significant 2-way interaction for WM × Arithmetic Operation was found in accuracies only [F = 5.37, $p = .002$, $\eta_p^2 = 0.05$] but no other 2-way nor the 3-way interaction was significant [*p*-values range: .13 - .48; see Table S3 in the Appendix for post-hoc analyses]. Despite the 2-way interaction being found, main effects were still significant for accuracies and post hoc comparisons were mainly driven by the addition secondary task, which did not provide strong support for the domain-specific or any of the integrated models we described a priori.

We compared the fit of different regression models that included different combinations of main effects and interactions. We started with a full model including the 2- and 3-way interactions and subsequently dropped each interaction term to test whether a simpler model fit the data better than the more complex one. Reaction times overall were best explained by a model that only included the main effects of working memory task load, arithmetic, and difficulty (Table S4). Our Bayesian repeated measures ANOVA yielded results consistent with our frequentist model comparisons (Table S5). Models including interactions along with the main effects (e.g. WM + Arithmetic + Difficulty + WM × Arithmetic; WM + Arithmetic + Difficulty + WM × Arithmetic + WM × Difficulty) yielded Bayes factors (B_{10}) < .001, suggesting strong evidence against the interaction terms. The added complexity did not explain our data more than the simpler additive model with main effects only. The simpler models tended to fit the data better than the more complex model, but not always (Tables S6 & S7). The Bayesian repeated measures ANOVA suggested the best fitting model for accuracy was one with WM + Arithmetic + WM × Arithmetic which appears to be primarily driven by the addition secondary task as can be seen in Table S3. We investigated these potential interactions further in our exploratory analyses.

Exploratory analysis 1: Probing interactions

Thus far, our results suggest stronger evidence for a domain-general model over a domain-specific or any of the specific integrated models we considered a priori. However, such results do not necessarily discount the other two models entirely. Therefore, we further probed whether there were changes in arithmetic performance specific to different combinations of factors often hypothesized to moderate performance (i.e., different working memory secondary tasks, receiving math education from a different country, and difficulty). We performed exploratory analyses by subjecting the data to multiple 2 × 2 (Secondary task load × Arithmetic operation) ANOVAs to identify if and where interactions could occur and compiled their F-values into a F-curve distribution (Figure 2). Overall, we detected 12 statistically significant interaction effects out of the 48 comparisons[6] all driven by the addition task compared to all other secondary tasks (Table S8). Across all significant interaction effects, the addition secondary task slowed and/or induced more errors than any other secondary task load with no other combination of factors (e.g., verbal/visuospatial × multiplication/subtraction) producing a significant interaction, suggesting that main effects alone may not accurately describe this

[6] 6 load comparisons (add/no load; add/verbal; add/visuospatial; verbal/no load; verbal/visuospatial; visuospatial no load) × 2 outcomes (RT/ACC) × 4 models (whole sample, Chinese-educated sample, easy load condition only, hard load condition only

pattern of performance. Addition secondary tasks produced interaction effects across the subgroup analyses, sometimes on reaction time but not accuracy and sometimes vice-versa (see F-value columns in Table S8).

These exploratory analyses suggest that chance alone could not adequately explain the interactions that were detected, but these effects may be somewhat idiosyncratic to the experiment's design. Importantly, *none* of the theoretical models described in the introduction reliably predicted the pattern of response across these interactions. Rather, the 2-way interactions here can be entirely attributed to the addition secondary task. Furthermore, subtraction performance was impacted more under addition than multiplication, contrary to our hypotheses. A possible reason for this pattern of results could be that participants are retrieving and processing repeated addition facts in both the addition and subtraction task. While retrieving facts would also suggest phonological involvement similar to multiplication, there may be greater switching costs between subtraction and addition. This account is speculative, however.

Exploratory analysis 2: Strategy use

In addition, we explored how strategy selection was related to performance and whether task features would impact strategy selection overall. Overall, strategy choice was not very informative in understanding the impact of task features or explaining our pattern of null results. Descriptive information on strategy selection can be found in Table S2. Participants favored retrieval based strategies (i.e., retrieval and decomposition) over procedural ones (i.e., counting and mixed/other) (58% vs 42%). Consequently, procedural strategies was associated with a significant slowing of about 66 ms (*se* = 21.8, *p* = .002) but no significant association with arithmetic accuracy. Task features (i.e., working memory tasks, arithmetic operation, difficulty) did not predict strategy choice outside of participants preferring retrieval based strategies (Table S9 Appendix).

Discussion

In this study, we tested several pre-registered predictions on dual-task performance in mental arithmetic with respect to several types of working memory models. That is, we tested whether the effects of different secondary tasks and difficulty on arithmetic could be attributed to three general categories of theories linking cognition to dual task performance:

1. A domain-general model whereby dual-task effects can be primarily attributed to task complexity or difficulty such that there is little to no influence of cognitive overlap between tasks;
2. A domain-specific model whereby dual-task effects can be primarily attributed to cognitive overlap such that tasks sharing similar cognitive resources cause more interference than tasks that do not share resources;
3. Integrated models whereby task complexity and cognitive overlap could have either an additive effect or that there is some level of interplay between them such that dual-task effects depend on both how complex and how much overlap a secondary task had with the arithmetic.

Overall, we found multiplication and subtraction performance to be slower and less accurate under dual-task conditions. Specifically, our results were largely in favor of a domain-general model of working memory (General model predictions A & B) such that a majority of our statistical models supported the main effects of task complexity via difficulty and the effect of the repeated addition task. Contrary to previous studies in the past that found different interactions between different types of secondary tasks and arithmetic operations (e.g., Lee &

Kang, 2002; Imbo & LeFevre, 2010), we were unable to reproduce any similar interactions that would support either a domain-specific or integrated model. We did not find differential effects of verbal and visuospatial secondary tasks on multiplication and subtraction nor did we find the addition task reducing multiplication performance more than subtraction. Instead, we found the addition secondary task to impair performance more than any other secondary task especially with subtraction. While our initial prediction that multiplication performance would be impaired from its reliance on repeated addition processes, the more prominent effect in subtraction may imply a greater amount of similarity between the inverse operations or that task switching between inverse operations may be particularly cognitively demanding.

Despite a major focus on domain-specific interactions in the arithmetic dual task literature, this study's results are more consistent with a number of non-arithmetic dual-task studies, which have found that slowing effects of secondary tasks primarily rely on domain-general or attentional control resources over domain-specific resources (Barrouillet et al., 2015; Uittenhove et al., 2019). A latent variable approach also suggests that while short-term memory span tasks alone could measure the domain-specific effects of constructs like verbal or visuospatial working memory, complex span tasks that reflect the design of a dual-task arithmetic experiment largely reflect a domain-general factor instead (Kane et al., 2004).

Taken together, the largely null interactions between primary and secondary tasks, with the exception of the larger slowing effect of the addition secondary task on subtraction performance arithmetic, raise the possibility that interactions may be real but elusive and idiosyncratic to highly specific task demands. If so, then perhaps broad conclusions about human cognition should not be drawn from interactions between primary and secondary task types in the absence of replications across versions of the primary and secondary tasks that vary surface features but hold constant the hypothesized core features hypothesized to overlap between them.

In conclusion, our results mostly supported the domain-general influence of working memory in dual-task arithmetic. However, to say our results discount any domain-specific or integrated models would be an oversimplification. In particular, we found evidence for a specific effect of a secondary addition task on subtraction (compared to multiplication) performance. Still, this is a different kind of specific effect than typically hypothesized in dual task arithmetic experiments. Unfortunately, we were unable to reach meaningful conclusions with our use of strategy reports. This may be due to the retrospective nature of reporting strategy at the end of each experimental block. To our knowledge, dual-task arithmetic studies have only looked at strategy reports but have not experimentally manipulated strategy choice, so future work may want to control this factor to limit the effects of individual differences in dual-task arithmetic performance.

We conclude that selective effects of secondary tasks on performance on different primary tasks may be real but difficult to predict in dual-task experiments and generally smaller than the effects on performance via domain-general pathways. There are likely specific ways that working memory can affect dual-task arithmetic performance, such as through individual differences in strategy choice for arithmetic or dual-task performance. We encourage future work that attempts to build and test theories of arithmetic performance under cognitive load that can make predictions across a variety of primary-secondary task combinations.

Appendix

Table 1. Demographic information

Variable	**n=97**	**M(SD)**	**%**
Gender			
Male	33		66
Female	64		34
Age		20.1(1.3)	
Country of primary math education			
US	71		73.2
China	22		22.7
Other	4		4.1
Math grade compared to peer			
A	35		36.08
B	49		50.52
C	12		12.37
D	1		1.03
F	0		0
Abacus use			
Never Taught	69		71.13
Never Used	16		16.49
Rarely	11		11.34
Sometimes	1		1.03

Table 2. Means and standard deviations for reaction time and accuracy as a function of WM task × arithmetic operation × difficulty

WM	Arithmetic	Difficulty	RT Mean	SD	ACC Mean	SD	N
No load	Multiplication		789	236	0.93	0.09	95
	Subtraction		716	231	0.95	0.10	95
Verbal	Multiplication	Hard	973	246	0.90	0.11	95
		Easy	910	227	0.89	0.13	95
	Subtraction	Hard	864	252	0.91	0.11	95
		Easy	843	243	0.91	0.11	95
Visuospatial	Multiplication	Hard	934	256	0.89	0.13	95
		Easy	921	236	0.91	0.11	95
	Subtraction	Hard	857	284	0.92	0.08	95
		Easy	860	299	0.94	0.08	95
Repeated Add	Multiplication	Hard	982	220	0.86	0.12	95
		Easy	975	242	0.87	0.13	95
	Subtraction	Hard	963	296	0.85	0.14	95
		Easy	913	251	0.86	0.14	95

Note. WM: working memory. Reaction times (RT) in nearest ms and accuracy (ACC) in nearest percentage

Table 3. Effect Size Comparisons between Difficulty x Load vs No load by Arithmetic

		Verbal		Addition		Visuospatial	
	Arithmetic	easy	hard	easy	hard	easy	hard
RT	multiplication	0.52	0.76	0.78	0.85	0.56	0.59
	subtraction	0.53	0.61	0.82	0.93	0.54	0.54
ACC	multiplication	-0.31	-0.30	-0.49	-0.66	-0.20	-0.29
	subtraction	-0.32	-0.37	-0.71	-0.80	-0.11	-0.28

Figure 1. Comparison of dual-task effects in reaction times and accuracies. A and B (top row) represent reaction time. C and D (bottom row) represent accuracy. Error bars represent standard errors of the mean.

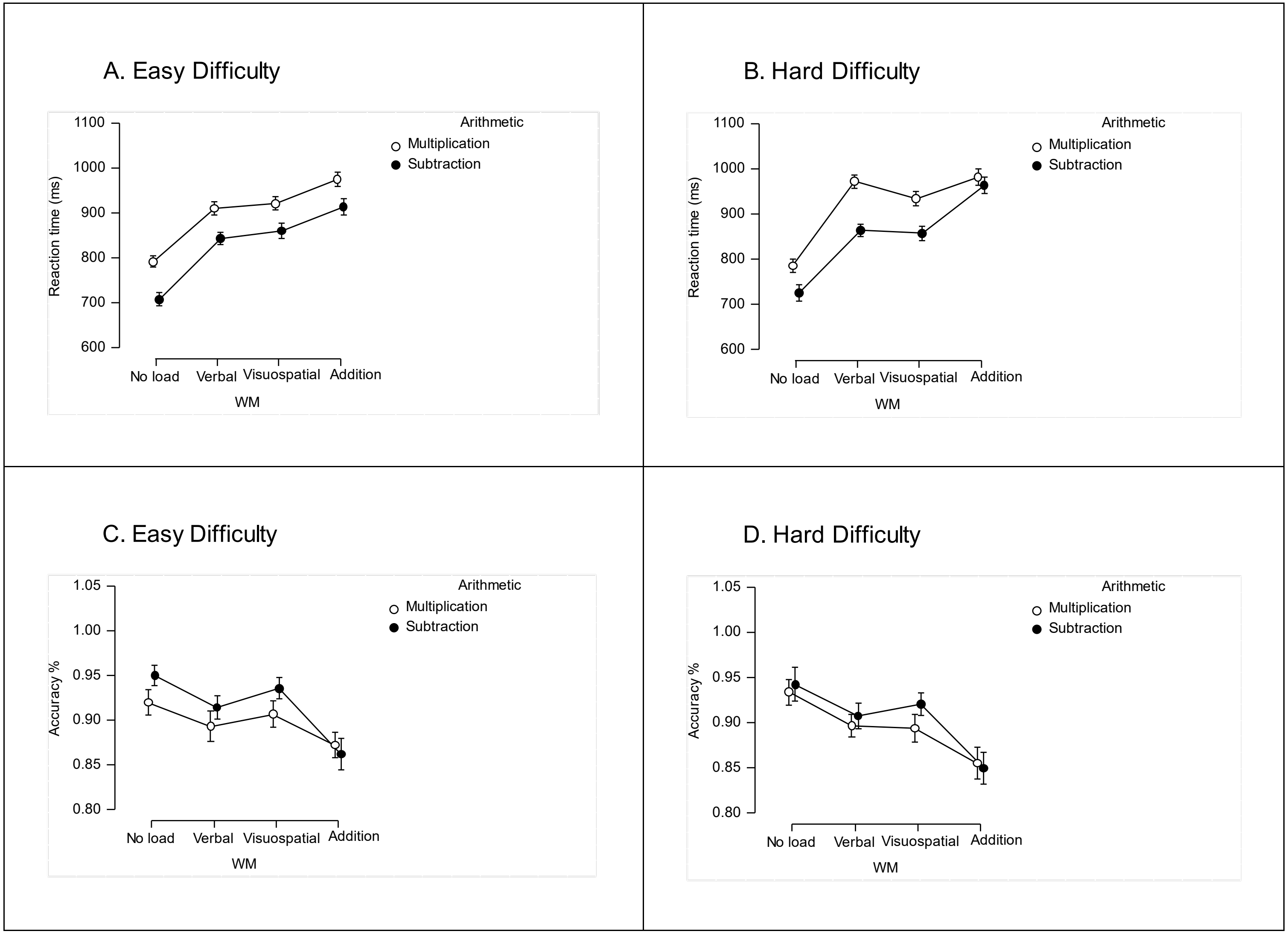
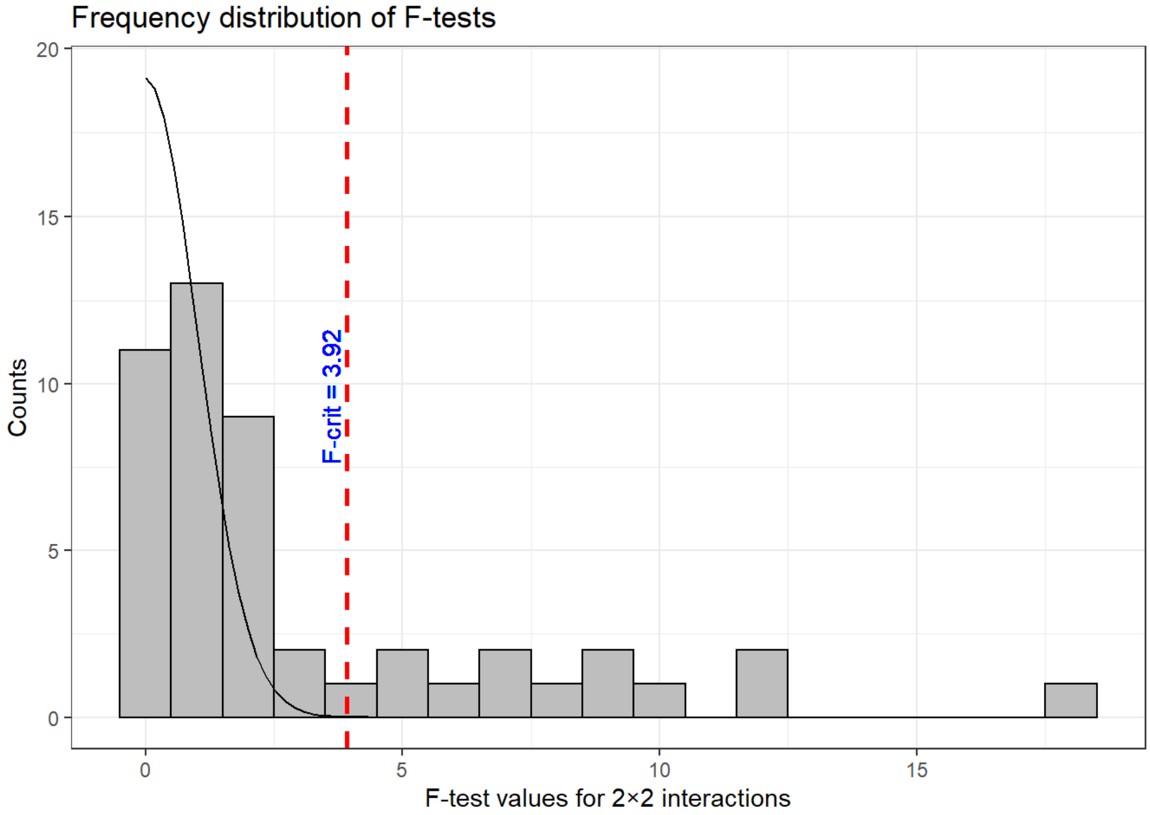


Figure 2. Distribution of F-test values comparing secondary task type by arithmetic operations interactions.



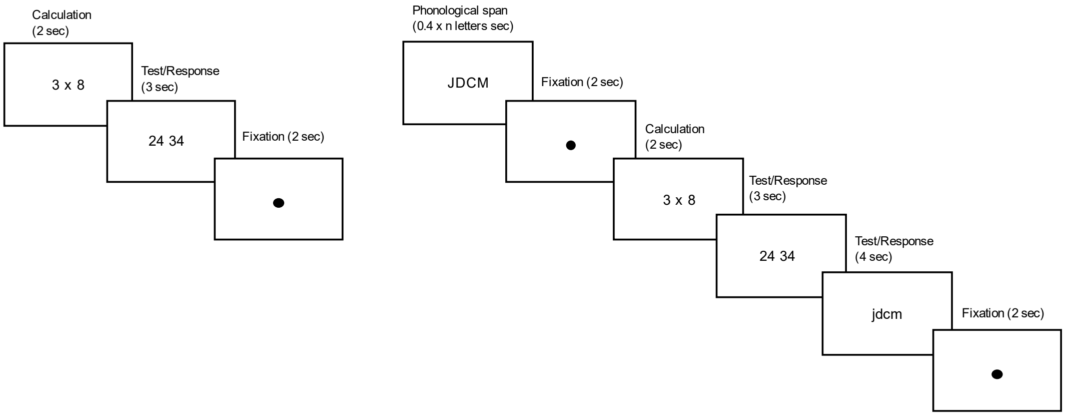


Figure 3. Single multiplication task (left). Dual-task multiplication with verbal letter WM load (right)

General Discussion

The aim of the present dissertation was to investigate whether and how working memory influences mental arithmetic performance within dual-task situations. More broadly, I sought to reconcile differences across the math cognition literature in which developmental research pointed to a consistent relation between working memory and math where experimental research had not. To this end, I approached this problem in two ways: a meta-analysis of dual-task experiments (Study 1) and a within-subject dual-task experiment testing the predictions of different working memory theories on arithmetic operations using a variety of secondary distractor tasks (Studies 2 and 3).

In Study 1, I conducted a meta-analysis of dual-task experiments to test the extent to which mental arithmetic performance relied on working memory resources and to what extent features of the dual-task design moderated performance. Through the meta-analysis, I found robust evidence in support of a causal relationship between working memory and arithmetic. More so, the type of working memory task – namely demanding domain-general tasks – contributed the most to the overall change in performance between single and dual-task conditions. Evidence for more specific secondary and primary task combinations was much more limited in comparison, suggesting a few explanations. One would be that – in and of itself – the dual-task paradigm is a complex cognitive task that draws on domain-general processes like task switching or inhibition, thereby reducing the overall impacts of domain-specific working memory resources drawn from secondary tasks. That is, the general demands of a secondary task outweigh any domain-specific effects from similarly coded tasks. Alternatively, the domain-specific effects themselves may be smaller than what has been predicted by the literature and have less of an impact than previously thought as well.

In Study 2, I probed possible domain-specific effects of working memory in mental arithmetic further in a large within-subject experiment. In addition, I attempted to replicate one such interaction between secondary task type (verbal and visuospatial memory) × arithmetic operation (subtraction and multiplication) and compared the extent to which the interaction effect in my experiment aligned with prior research. Despite testing for the interaction within numerous analytical models based on prior results from the dual-task previous studies, I was unable to replicate the interaction. These types of primary and secondary task interactions are often interpreted as evidence for domain-specific pathways of working memory influence, but these findings along with those of Study 1 would otherwise suggest a greater role of domain-general influence in dual-task arithmetic. Dual-task arithmetic literature has often focused on these domain-specific effects following predictions based on the multicomponent model of working memory from Baddeley & Hitch (1974). However, other theoretical accounts which emphasize domain-general influences that focus more on attentional control may be more useful in making predictions for future dual-task experiments (Barrouillet & Camos, 2001; Cowan, 1999; Engle, 2002; Oberauer, 2009). That is not to say domain-specific effects do not exist, just that instances in which these effects can be reliably predicted a priori and tested upon may require highly specific combinations of primary and secondary tasks.

Finally in Study 3, I compared predictions from domain-general, domain-specific, and integrated theories of dual-task performance between working memory and arithmetic. I used data from the same experiment as Study 2 but with the inclusion of difficulty as a third factor and an addition task as another secondary task. Similar to Study 2, I found consistent main effects of dual-task load suggesting arithmetic performance was primarily driven by difficulty and domain-general resources rather than domain-specific or some combination of general and specific influences. Again, domain-specific effects may still exist as I found a larger effect of the addition secondary task on both multiplication and subtraction compared to both verbal and visuospatial load. Moreover, the effect was larger for subtraction compared to multiplication – opposite of what I predicted. These findings raise concerns regarding the design of dual-task experiments. First, that in order to detect a selective interference effect, the primary and secondary tasks must share more semantically similar resources than what has been previously used in the dual-task arithmetic literature. Second, addition and subtraction being semantically similar but opposite in operations poses a significant switching cost within a dual-task situation. Taken together, this would suggest the use of simple-span tasks involving the storage and rehearsal of letters or spatial positions as secondary tasks are unlikely to produce those elusive selective interference or "crosstalk" effects. Indeed, the nature of dual-task experiments themselves lend more to being able to measure the effect of domain-general resources than domain-specific.

Overall, it is clear that working memory plays a pivotal role in the processing of mental arithmetic as evidenced from Study 1's meta-analysis of dual-task experiments as well as my own dual-task experiment in Studies 2 and 3. However, the results of my experiments in Studies 2 and 3 were not consistent with predictions of arithmetic performance under cognitive load that follow multiple component accounts of working memory (i.e., Baddeley & Hitch, 1994; Logie 2016). Rather, my findings were consistent with more general, attention-based theories that suggest a broader effect from managing the amount of available cognitive resources (e.g., Cowan, 1999, Oberauer, 2009; Barrouillet & Camos, 2020). While I was unable to find domain-specific effects of verbal or visuospatial memory across multiplication and subtraction, I found the addition secondary task to have large differential effects across arithmetic operations. This novel finding presents two possibilities. The first is that working memory operates primarily through an attention-based system where secondary tasks generate varying levels of activation that limit capacity rather than semantically or content related information pathways competing for similar resources (Oberauer, 2009, 2010). The addition task placed too high of a demand on the limited working memory capacity leading to worse performance. However, this account does not entirely explain why addition load was much larger than either verbal or visuospatial load because difficulty was individualized for each participant. This could mean that, controlling for difficulty, there are real selective interference effects but only when the primary and secondary tasks are highly similar to one another. Thus, the second account posits that task features do matter but to a much lesser extent than do general task demands. The other secondary working memory tasks may not have shared enough similarity or overlap (e.g., dots vs numbers, letters vs numbers, numbers vs numbers) with the primary arithmetic tasks to have reliably produced differential effects, but an addition task did. Assuming the second account is true, future work would need to identify what features are most important between tasks to better understand the role working memory plays within dual-task performance.

For example, one key distinction between the arithmetic tasks used in many dual-task experiments is the selected operator (i.e., addition, subtraction, multiplication, division). Neuropsychological studies suggest these operations share more common networks with each other than with other working memory tasks from the literature (Arsalidou & Taylor, 2011; Dehaene, 1992, Dehaene & Cohen, 1995, Dehaene & Cohen, 1997). As such, it may no longer be the case that the working memory tasks typically used in dual-task arithmetic experiments can be as useful for identifying pathways of working memory influence in mental arithmetic, which may explain why differential suppression effects are so difficult to find. Hence for cognitive overlap to create significant interference above and beyond general dual-task costs, it would require the overlap to be highly specific – possibly involving similar operations, problem size, no individual differences in strategy use, or by maintaining similar modalities in presentation formats and responses (e.g., auditory presentation and oral response vs visual presentation and key response). In any case, it is still not clear a priori when and how cognitive overlap could matter in these arithmetic studies. Potentially, the demands of the dual-task design itself must not outweigh the cognitive overlap between the primary and secondary tasks, so that domain-general influences in switching costs have less of an impact on performance. This is possible given that individuals can be trained to improve dual-task performance (Strobach, 2020) though total elimination of dual-task costs may not be possible (Strobach & Schubert, 2017).

From the literature and my own work, many different manipulations to examine these domain-specific effects have been tested to varying degrees of success. Careful manipulations of future dual-task experiments can be useful in identifying key factors underlying individual differences in these mental arithmetic tasks, but if my interpretation about the difficulties in finding these effects are true, it would mean that the experimental rigor would come at the cost of any practical significance for the field of math cognition. Research is clear on the relation between working memory and math, though not necessarily the pathways by which working memory influences mathematical processing, especially when it comes to domain-specific influences of visuospatial or verbal memory. Overall, the current studies support the notion that working memory has strong domain-general influences on arithmetic performance but no current model of working memory is able to accurately and reliably predict dual-task arithmetic performance.

Limitations

One of the key limitations from the meta-analysis was a lack of statistical power in detecting interactions between primary and secondary task combinations, which I tried to address in the Studies 2 and 3 with the large within-subject design. Still, despite being sufficiently powered to detect a small effect size in experiment, I was unable to do so. That is not to say there are no domain-specific effects, but the combinations of tasks and particular design used in my experiment may not have led to the correct circumstances to detect them. Individual differences in strategy use also limited our findings in all three studies. Efficiency in arithmetic strategy use is highly correlated with working memory capacity (Bailey, Littlefield, & Geary, 2012; Geary et al., 2004, 2007), but I could test this directly in the meta-analysis nor the experiment. Strategy use has been measured in dual-task arithmetic before (e.g., Tronksky, 2005; Tronsky et al., 2008), but to my knowledge, strategy use has not been directly manipulated. Future work would benefit from trying to control for strategy use to mitigate individual differences in dual-task arithmetic performance. I also broadly discussed different theories of working memory to create predictions for Study 3 rather than explicating exactly how each theory differed from one another, thereby creating more general hypotheses. Some adversarial collaboration work between different labs has already been done to compare how different working memory theories predict general dual-task performance (see Doherty et al., 2019), so there is precedence both here and in the collaboration for work to be extended towards dual-task arithmetic.

Conclusion

Throughout the process of this dissertation, I have attempted to reconcile conflicting findings on how working memory is implicated in mathematical processing. I investigated the effects of working memory on mental arithmetic performance in dual-task experiments both meta-analytically and experimentally and was able to find a robust causal link between them. However, even with a strong experimental design, I was unable to infer clear answers towards the nature of working memory in arithmetic processing. Working memory has wide-reaching, domain-general effects on arithmetic processing, but it also has subtle, domain-specific effects as well that are difficult to predict and test for. No single theory of working memory thus far has been able to explain all of the subtle differences in dual-task arithmetic performance, but determining that is beyond the scope of this project. The goal of this work is to provide a better understanding of how working memory may influence arithmetic as well as new avenues of future research to build from.

References

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review, 94*, 192–210. http://dx.doi .org/10.1037/0033-295X.94.2.192

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221–256. http://dx.doi.org/10.1006/cogp.1996.0007

Arsalidou, M., & Taylor, M. J. (2011). Is 2+ 2= 4? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage, 54*(3), 2382-2393.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1–2), 75–106. http://dx.doi.org/10.1016/0010-0277(92) 90051-I

Ashcraft, M. H., Donley, R. D., Halas, M. A., & Vakali, M. (1992). Working memory, automaticity, and problem difficulty. *Advances in Psychology, 91*, 301–329. http://dx.doi.org/10.1016/S0166-4115(08)60890-0

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*, 201–208. http://dx.doi.org/10 .1016/j.neuroscience.2005.08.023

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A, 49*, 5–28. http://dx.doi.org/ 10.1080/713755608

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. *In Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485.

Baddeley, A., & Hitch, G. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8,* 234–238. http://dx.doi.org/10 .1016/S0959-4388(98)80145-1

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D. H., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. *American Psychologist, 73,* 81–94. http://dx.doi .org/10.1037/amp0000146

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology, 113*(1), 78-92. http://dx.doi.org/10.1016/j.jecp.2012.04.014

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay?. *Journal of Memory and Language*, 45(1), 1-20.

Barrouillet, P., & Camos, V. (2020). The time-based resource-sharing model of working memory. *Working Memory: State of the Science*, 85-115.

Barrouillet, P., Corbin, L., Dagry, I., & Camos, V. (2015). An empirical test of the independence between declarative and procedural working memory in Oberauer's (2009) theory. *Psychonomic bulletin & review, 22*(4), 1035-1040.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. Child *Development Perspectives, 8,* 36–41. http://dx.doi .org/10.1111/cdep.12059

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language, 61*(3), 457-469.

Camos, V., Lagner, P., & Loaiza, V. M. (2017). Maintenance of item and order information in verbal working memory. *Memory, 25*(8), 953-968.

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*, *39*(2), 231-244.

Cavdaroglu, S., & Knops, A. (2016). Mental subtraction and multiplication recruit both phonological and visuospatial resources: Evidence from a symmetric dual-task design. *Psychological Research, 80*, 608–624. http://dx.doi.org/10.1007/s00426-015-0667-8

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*, 141– 160. http://dx.doi.org/10.1016/j.jecp.2012.02.005

Chen, E. H., & Bailey, D. H. (2021). Dual-task studies of working memory and arithmetic performance: A meta-analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(2), 220.

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences, 21*, 111–124. http://dx.doi.org/10.1016/j.tics.2016.12.007

Clearman, J., Klinger, V., & Szu˝cs, D. (2017). Visuospatial and verbal memory in mental arithmetic. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 70*, 1837–1855. http://dx.doi .org/10.1080/17470218.2016.1209534

Cohen, L., & Dehaene, S. (1995). Number processing in pure alexia: The effect of hemispheric asymmetries and task demands. *Neurocase, 1*(2), 121-137.

Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., . . . Karama, S. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence*, *41*, 712–727. http://dx.doi.org/10.1016/j.intell.2013.09.002

Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cragg, L., Richardson, S., Hubber, P. J., Keeble, S., & Gilmore, C. (2017). When is working memory important for arithmetic? The impact of strategy and age. *PloS One*, *12*(12), e0188693.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychological Research, 62*, 72–77. http://dx .doi.org/10.1007/s004260050041

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory & Cognition, 29*, 267–273. http:// dx.doi.org/10.3758/BF03194920

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1-2), 1-42.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1(1), 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, 33(2), 219-250.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3-6), 487-506.

Deschuyteneer, M., & Vandierendonck, A. (2005). The role of response selection and input monitoring in solving simple arithmetical products. *Memory & Cognition, 33,* 1472–1483. http://dx.doi.org/10.3758/ BF03193379

DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353– 386. http://dx.doi.org/10.1080/09541440244000328

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135–168. http://dx.doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333,* 959–964. http://dx.doi.org/10.1126/science.1204529

Doherty, J. M., & Logie, R. H. (2016). Resource-sharing in multiple-component working memory. *Memory & Cognition, 44*(8), 1157-1167.

Doherty, J. M., Belletier, C., Rhodes, S., Jaroslawska, A., Barrouillet, P., Camos, V., . . . Logie, R. H. (2019). Dual-task costs in working memory: An adversarial collaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(9), 1529-1551. http://dx.doi.org/10.1037/xlm0000668

Duverne, S., Lemaire, P., & Vandierendonck, A. (2008). Do workingmemory executive components mediate the effects of age on strategy selection or on strategy execution? Insights from arithmetic problem solving. *Psychological Research, 72,* 27–38. http://dx.doi.org/10.1007/ s00426-006-0071-5

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal (Clinical Research Ed.), 315*, 629–634. http://dx.doi.org/10 .1136/bmj.315.7109.629

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23. http://dx.doi.org/10.1111/1467-8721.00160

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation, 44*, 145–200. http://dx.doi.org/10.1016/ S0079-7421(03)44005-X

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

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 Method* , 39, 175-191.

Fischer, R., & Plessow, F. (2015). Efficient multitasking: parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, 6, 1366.

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. arXiv preprint arXiv:1503.02220.

Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & Van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, 10, 29–44. http://dx.doi.org/10.1016/j.edurev.2013.05.003

Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, 28, 774 –782. http://dx.doi.org/10.3758/ BF03198412

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology,* 88, 121–151. http://dx.doi.org/10.1016/j.jecp.2004.03.002

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,* 1343–1359. http://dx.doi.org/10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Widaman, K. F., & Little, T. D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition, 14*, 478–487. http://dx.doi.org/10.3758/BF03202519

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*, 236–242. http://dx.doi.org/10.1016/j.tics.2007.04.001

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition, 30*, 447–455. http://dx.doi.org/10.3758/BF03194945

Higgins, J. P., White, I. R., & Anzures-Cabrera, J. (2008). Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. *Statistics in Medicine, 27*, 6072–6092. http://dx.doi.org/10.1002/ sim.3427

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. Cognitive Psychology, 10, 302–323. http://dx.doi.org/10 .1016/0010-0285(78)90002-6

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience, 6,* 435–448. http://dx.doi.org/10.1038/nrn1684

Hubber, P. J., Gilmore, C., & Cragg, L. (2014). The roles of the central executive and visuospatial storage in mental arithmetic: A comparison across strategies. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 67,* 936–954. http://dx.doi.org/10.1080/ 17470218.2013.838590

Hurst, M., & Cordes, S. (2017). Working memory strategies during rational number magnitude processing. *Journal of Educational Psychology, 109*, 694–708. http://dx.doi.org/10.1037/edu0000169

Imbo, I., & LeFevre, J. A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese- and Canadian-educated adults. *Memory & Cognition, 38*, 176–185. http://dx.doi.org/ 10.3758/MC.38.2.176

Imbo, I., & Vandierendonck, A. (2007a). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96,* 284–309. http:// dx.doi.org/10.1016/j.jecp.2006.09.001

Imbo, I., & Vandierendonck, A. (2007b). The role of phonological and executive working memory resources in simple arithmetic strategies. *European Journal of Cognitive Psychology, 19*, 910–933. http://dx.doi .org/10.1080/09541440601051571

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *Quarterly*

Journal of Experimental Psychology: Human Experimental Psychology, 60, 708–731. http://dx.doi.org/10 .1080/17470210600762447

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2012). Cogmed and working memory training—Current challenges and the search for underlying mechanisms. *Journal of Applied Research in Memory & Cognition, 1*, 211–213. http://dx.doi.org/10.1016/j.jarmac.2012.07.002

JASP Team (2022). JASP (Version 0.16.2)[Computer software].

Kalaman, D. A., & LeFevre, J. A. (2007). Working memory demands of exact and approximate addition. *European Journal of Cognitive Psychology, 19*, 187–212.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of experimental psychology: General, 133*(2), 189.

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., ... & Fukuda, H. (2004). A functional MRI study of simple arithmetic—a comparison between children and adults. *Cognitive Brain Research*, *18*(3), 227-233.

Ketelsen, K., & Welsh, M. (2010). Working memory and mental arithmetic: A case for dual central executive resources. *Brain and Cognition, 74*, 203–209. http://dx.doi.org/10.1016/j.bandc.2010.07.011

Koch, I. (2009). The role of crosstalk in dual-task performance: Evidence from manipulating response-code overlap. *Psychological Research, 73*, 417–424. http://dx.doi.org/10.1007/s00426-008-0152-8

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: a functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *48*(4), 657-661.

Lee, K. M., & Kang, S. Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition, 83,* B63– B68. http://dx.doi.org/10.1016/S0010-0277(02)00010-0

LeFevre, J. A., & Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition, 3*(1), 31-62.

LeFevre, J. A., Lei, Q., Smith-Chant, B. L., & Mullins, D. B. (2001). Multiplication by eye and by ear for Chinese-speaking and English-speaking adults. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 55*(4), 277.

Lemaire, P. (1996). The role of working memory resources in simple cognitive arithmetic. *European Journal of Cognitive Psychology, 8,* 73–104. http://dx.doi.org/10.1080/095414496383211

Lien, M. C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory period effects: Implications for response selection. *Psychonomic Bulletin & Review, 9,* 212–238. http://dx.doi.org/10 .3758/BF03196277

Logan, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107*, 32–63. http://dx.doi.org/10.1037/0096-3445.107.1.32

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. Journal of Experimental Psychology: *Human Perception and Performance, 5*, 189–207. http://dx.doi.org/10.1037/ 0096-1523.5.2.189

Logie, R. (2018). Human cognition: Common principles and individual variation. *Journal of Applied Research in Memory and Cognition, 7*(4), 471-486.

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science, 20*(4), 240-245.

Logie, R. H. (2016). Retiring the central executive. *Quarterly Journal of Experimental Psychology, 69*(10), 2093-2109.

Logie, R. H., & Baddeley, A. D. (1987). Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 310–326. http://dx.doi.org/10.1037/0278-7393.13.2.310

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*, 395– 410. http://dx.doi.org/10.3758/BF03200866

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology, 18*, 62–78. http://dx.doi.org/ 10.1080/09297049.2011.575772

McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educational and Child Psychology, 20*, 93–108.

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

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science, 11,* 512–534. http://dx.doi.org/ 10.1177/1745691616635612

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,* 101–109. http://dx.doi.org/10.1016/j.lindif .2009.08.004

Miller, J. (2006). Backward crosstalk effects in psychological refractory period paradigms: Effects of second-task response types on first-task response latencies. *Psychological Research, 70,* 484–493. http://dx.doi .org/10.1007/s00426-005-0011-9

Miyake, A., & Shah, P. (Eds.). (1999). Models of working memory: Mechanisms of active maintenance and executive control. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO 9781139174909

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100. http://dx.doi.org/10 .1006/cogp.1999.0734

Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review, 20*(1), 163-170.

Morey, C. C., Morey, R. D., Van Der Reijden, M., & Holweg, M. (2013). Asymmetric cross-domain interference between two working memory tasks: Implications for models of working memory. *Journal of Memory and Language, 69*(3), 324-348.

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 435–448. http://dx.doi.org/10.1037/0096-1523.13 .3.435

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*(3), 193-251.

Noël, M. P., Désert, M., Aubrun, A., & Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory & Cognition, 29*, 34–42. http://dx.doi.org/10.3758/BF03195738

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, 51, 45-100.

Oberauer, K. (2010). Declarative and procedural working memory: Common principles, common capacity limits?. *Psychologica Belgica*, *50*(3-4).

Park, S., & Beretvas, S. N. (2019). Synthesizing effects for multiple outcomes per study using robust variance estimation versus the threelevel model. *Behavior Research Methods, 51*, 152–171.

Park, S., Kim, M. S., & Chun, M. M. (2007). Concurrent working memory load can facilitate selective attention: evidence for specialized load. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1062.

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 358.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220–244. http://dx.doi.org/10.1037/0033- 2909.116.2.220

Pashler, H. (1994). Graded capacity-sharing in dual-task interference?. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 330.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*(1), 629-651.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. 10.3758/s13428-018-01193-y

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*, 455–473. http://dx.doi.org/10 .1037/edu0000079

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., & Booth, J. R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*(11), 1932-1947.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110– 122. http://dx.doi.org/10.1016/j.lindif.2009.10.005

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *Journal of Numerical Cognition, 3,* 468–495. http://dx.doi.org/10.5964/jnc.v3i2.31

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology*, *91*, 137–157.

Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., Gold, L., . . . Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory: A randomized clinical trial. *Journal of the American Medical Association Pediatrics, 170,* e154568 – e154568. http://dx.doi.org/10.1001/ jamapediatrics.2015.4568

Rouder, J. N. (2005). Are unshifted distributional models appropriate for response time? *Psychometrika, 70,* 377–381. http://dx.doi.org/10.1007/ s11336-005-1297-7

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Ruthruff, E., Pashler, H. E., & Hazeltine, E. (2003). Dual-task interference with equal task emphasis: Graded capacity sharing or central postponement? *Perception & Psychophysics, 65,* 801–816. http://dx.doi.org/10 .3758/BF03194816

Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin & Review, 8,* 73–80. http://dx.doi.org/10 .3758/BF03196141

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology, 53*, 671– 685. http://dx.doi.org/10.1037/ dev0000265

Salthouse, T. A. (1988). The role of processing resources in cognitive aging. In M. L. Howe & C. J. Brainerd (Eds.), Cognitive development in adulthood (pp. 185–239). New York, NY: Springer. http://dx.doi.org/10 .1007/978-1-4612-3852-2_7

Schmidt, F. L. (2017). Statistical and measurement pitfalls in the use of meta-regression in meta-analysis. *The Career Development International, 22*, 469–476. http://dx.doi.org/10.1108/CDI-08-2017-0136

Seitz, K., & Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *European Journal of Cognitive Psychology, 12*, 552– 570. http://dx.doi.org/10.1080/095414400750050231

Seitz, K., & Schumann-Hengsteler, R. (2002). Phonological loop and central executive processes in mental addition and multiplication. *Psychological Test and Assessment Modeling, 44*(2), 275.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138,* 628–654. http://dx.doi .org/10.1037/a0027473

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405-410.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children’s addition. *Journal of Experimental Psychology: General, 116*, 250–264. http://dx.doi.org/10.1037/0096-3445.116.3.250

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*, 258–275. http://dx.doi.org/10.1037/0096-3445.117.3.258

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest, 17,* 103– 186. http://dx.doi.org/10.1177/1529100616661983

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*, 60–78. http://dx.doi.org/10.1002/jrsm.1095

Strobach, T. (2020). The dual-task practice advantage: Empirical evidence and cognitive mechanisms. *Psychonomic Bulletin & Review*, *27*(1), 3-14.

Strobach, T., & Schubert, T. (2017). No evidence for task automatization after dual-task training in younger and older adults. *Psychology and Aging*, *32*(1), 28.

Strobach, T., Schütz, A., & Schubert, T. (2015). On the importance of Task 1 and error performance measures in PRP dual-task studies. *Frontiers in Psychology*, *6*, 403.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. Journal of Experimental Psychology: *Human Perception and Performance*, 29(1), 3.

Trbovich, P. L., & LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31*, 738–745. http:// dx.doi.org/10.3758/BF03196112

Treisman, A. M., & Davies, A. (1973). Dividing attention to ear and eye. In S. Kornblum (Ed.), Attention and Performance IV (pp. 101–117). San Diego, CA: Academic Press.

Tronsky, L. N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory & Cognition, 33,* 927–940. http://dx.doi.org/10.3758/BF03193086

Tronsky, L. N., McManus, M., & Anderson, E. C. (2008). Strategy use in mental subtraction determines central executive load. *The American Journal of Psychology, 121*, 189 – 207. http://dx.doi.org/10.2307/ 20445456

Uittenhove, K., Chaabi, L., Camos, V., & Barrouillet, P. (2019). Is working memory storage intrinsically domain-specific? *Journal of Experimental Psychology: General*, *148*(11), 2027–2057. https://doi.org/10.1037/xge0000566

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1072.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48. http://dx.doi.org/10 .18637/jss.v036.i03

Vreugdenburg, L., Bryan, J., & Kemps, E. (2003). The effect of selfinitiated weight-loss dieting on working memory: The role of preoccupying cognitions. *Appetite, 41*, 291–300. http://dx.doi.org/10.1016/ S0195-6663(03)00107-7

Welford, A. T. (1952). The psychological refractory period and the timing of high-speed performance-a review and a theory. *British Journal of Psychology, 43*(1), 2.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455.

Wickham et al., (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686, https://doi.org/10.21105/joss.01686

Xenidou-Dervou, I., van Lieshout, E. C., & van der Schoot, M. (2014). Working memory in nonsymbolic approximate arithmetic processing: A dual-task study with preschoolers. *Cognitive Science, 38*, 101–127.