Intelligence 92 (2022) 101649

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

The relation between working memory and mathematics performance among students in math-intensive STEM programs

Michal Berkowitz *, Peter Edelsbrunner, Elsbeth Stern

ETH Zurich, Department of Humanities, Social and Political Sciences, Zurich, Switzerland

A R T I C L E   I N F O

Keywords:
Working memory
Mathematics
STEM
Higher education

A B S T R A C T

This study examined how working memory (WM) and mathematics performance are related among students entering mathematics-intensive undergraduate STEM programs ($N = 317$). Among students of mechanical engineering and math-physics, we addressed two questions: (1) Do verbal and visuospatial WM differ in their relation with three measures of mathematics performance: numerical reasoning ability, prior knowledge in mathematics, and achievements in mathematics-intensive courses? (2) To what extent are the effects of WM on achievements in mathematics-intensive courses mediated by numerical reasoning ability and prior knowledge in mathematics? A latent correlational analysis revealed that verbal WM was at least as strongly associated with the three mathematics measures as visuospatial WM. A latent mediation model revealed that numerical reasoning fully mediated the effects of WM on achievements in math-intensive courses, both directly and in a doubly mediated effect via prior knowledge in mathematics. We conclude that WM across modalities contributes significantly to mathematics performance of mathematically competent students. The effect of verbal WM emerges as being more pronounced than has been assumed in prior literature.

1. Introduction

Complex thinking constantly requires access to relevant information, combining elements of information and ignoring irrelevant information – processes that all take place in working memory (WM) (Baddeley, 2003; Cowan, Morey, Chen, Gilchrist, & Saults, 2008). As such, WM is regarded an essential cognitive resource for learning, reasoning and other complex cognitive activities (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Conway, Getz, Macnamara, & Engel de Abreu, 2011; Hambrick, Kane, & Engle, 2005; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Mathematics performance, particularly across K-12 education levels, has been positively linked with measures of WM in numerous studies (Alloway & Alloway, 2010; Bull, Espy, & Wiebe, 2008; Gathercole, Pickering, Knight, & Stegmann, 2004; Peng, Namkung, Barnes, & Sun, 2016). Among adults, components of WM have been often linked with arithmetic calculations and basic numerical cognition (Ashcraft & Krause, 2007; Trbovich & LeFevre, 2003), while less research has focused on its role for advanced mathematics (Rohde & Thompson, 2007; Tolar, Lederberg, & Fletcher, 2009). In the present research, we examined the relation between WM and mathematics performance in the special group of university students in STEM programs. Mathematics is a core subject in higher STEM education and is necessary for learning other topics, such as physics and engineering. Although students entering STEM may already be mathematically competent, mathematics-intensive courses in these programs are highly challenging even for beginner students who excelled in mathematics in high school. Previous work showed that different cognitive abilities differentiate achievements within high-ability groups, among them STEM students (Coyle, Snyder, Pillow, & Kochunov, 2011; Ferriman-Robertson, Smeets, Lubinski, & Benbow, 2010; Berkowitz & Stern, 2018). Here, we aimed to find out whether this also holds for individual differences in WM. We investigated two aspects of the WM-mathematics relation. First, we compared WM in different modalities, namely verbal and visuospatial, in their association with different measures of mathematics performance among STEM students. There is some evidence for shifts with age and mathematics expertise in the reliance on verbal comparing to visuospatial representations in WM during mathematics performance (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2015). Less, however, is known about this aspect with respect to advanced mathematics. Thus, this part of our study focused on the relevance WM modality to different types of mathematics performance. The second aspect

* Corresponding author at: ETH Zurich, Clausiusstrasse 59, CH-8092 Zurich, Switzerland.
E-mail address: bmichal@ethz.ch (M. Berkowitz).

https://doi.org/10.1016/j.intell.2022.101649
Received 19 April 2021; Received in revised form 3 March 2022; Accepted 3 April 2022
Available online 27 April 2022

we studied was the degree to which WM was predictive of achievements in math-intensive courses beyond other known predictors of mathematics achievements, namely numerical reasoning ability and prior knowledge in mathematics. Comparing to WM, both these factors are on a higher level of cognitive complexity and are more specific to the domain of mathematics. Previous research yielded inconsistent results regarding the relative contribution of WM and higher-order or more complex cognitive factors to achievement prediction in mathematics in different age groups (Alloway & Alloway, 2010; Fuchs et al., 2010; Tolar et al., 2009; Wei, Yuan, Chen, & Zhou, 2012). Thus, we aimed to examine this interplay among STEM students, who are a high-ability group, especially in terms of mathematics competence. In this part of the study, we focused on WM regardless of modality, hence on effects primarily driven by domain-general components of WM. We next introduce the construct of WM, followed by research findings on WM and mathematics.

2. Verbal, visuospatial and domain-general WM

Many WM models assume it to be a multicomponent system (Baddeley & Hitch, 1974; Cowan et al., 2008; Oberauer, 2009). According to the well-known model of Baddeley and Hitch (1974), the *Central Executive* is a core, domain-general mechanism in WM, which controls attention and coordinates incoming information; the *Phonological Loop* is a sub-system for the temporary storage of verbal information; and the *Visuospatial Sketchpad* is a sub-system for the temporary storage of visual and spatial information. Research on the structure of WM supports both a distinction based on modality and a substantial domain-general part, reflected in the overlap between tasks in different modalities (Conway & Kovacs, 2013; Kane et al., 2004; Shah & Miyake, 1996; Süß et al., 2002). What is not entirely agreed upon is the centrality of each part to the construct of WM. Many argue that domain general processes in WM are what drives its link with higher-order cognition (Barrouillet, Portrat, & Camos, 2011; Kane et al., 2004; Wilhelm, Hildebrandt, & Oberauer, 2013). Others think that the dissociation found between spatial and verbal processing in WM indicates that modality-specific mechanisms in WM are not peripheral (Shah & Miyake, 1996; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Kaufman, 2007; Waters & Caplan, 1996). However, most theories do not assume WM to be only domain general or only domain specific, but rather acknowledge that during complex cognitive activity such as reasoning and problem solving, domain-general and domain-specific processes are strongly intertwined. This is also reflected in the way WM is measured. Typically, WM tasks are designed to place attentional demands and require executive control (i.e. domain general processes), while information in some modality (e.g. words) is to be temporarily held or manipulated in WM. Known examples are the complex-span tasks (Conway et al., 2005) or memory updating tasks (Wilhelm et al., 2013). Even the so called simple span tasks, which were designed to tap primarily modality-specific storage, demand domain-general attentional processes as well (Cowan et al., 2008; Kane et al., 2004; Unsworth & Engle, 2007). Thus, many tasks used to assess WM activate both domain-general and modality-specific components. In the following, we refer to WM for either verbal or numerical information as *verbal WM*, and to WM involving non-verbal stimuli (e.g. shapes or locations) as *visuospatial WM*.

3. Working memory and mathematics

WM has been linked with mathematics performance ranging from basic numerical cognition (DeStefano & LeFevre, 2004; Ginsburg, van Dijck, Previtali, Fias, & Gevers, 2014) to school achievements (Alloway & Alloway, 2010; Friso-van den Bos et al., 2013; Fuchs et al., 2010). These links were found across normative development as well as in special populations such as students with learning disabilities and gifted students (Geary, 2011; Hoard, Geary, Byrd-Craven, & Nugent, 2008; Leikin, Paz-Baruch, & Leikin, 2013; Swanson & Beebe-Frankenberger, 2004). The majority of studies on WM and mathematics achievements focus on school children. For example, WM capacity at age five predicted math achievements in school up to six years later, independently of conventional measures of intelligence (Alloway & Alloway, 2010). Children with mathematical learning disabilities typically show poorer WM capacity than children of average math performance (Geary, 2011), whereas mathematically talented children have an advantage in WM task performance comparing to children with average mathematical abilities (Dark & Benbow, 1994; Hoard et al., 2008; Leikin et al., 2013). A recent meta-analysis concluded a mean significant effect of $r = 0.35$ across 110 studies, with stronger effects for mathematics word problems and arithmetic calculations, and for children with learning disabilities (Peng et al., 2016). The reasons for this consistently positive link lie in the nature of mathematics performance. Whether arithmetic calculations or more complex problem solving, one needs to mentally hold, manipulate and update information, to select and switch between solving strategies, and to inhibit inappropriate strategies (LeFevre, DeStefano, Coleman, & Shanahan, 2005). Still, the exact role that WM plays in mathematics performance seems to vary across populations, components of WM and mathematics measures (Friso-van den Bos et al., 2013).

With respect to the relative contribution of verbal and visuospatial WM to mathematics performance among children, there seem to be changes throughout development, but the results are mixed. Based on longitudinal data from second and third graders, Meyer, Salimpoor, Wu, Geary, and Menon (2010) concluded that a shift occurs with age from a stronger reliance on the central executive and the phonological loop to a higher reliance on visuospatial representations when solving mathematics problems. In contrast, a considerable number of studies found that visuospatial WM was more predictive of mathematics performance in early age, while verbal WM became more important with increasing mathematics experience (De Smedt et al., 2009; Raghubar, Barnes, & Hecht, 2010; Van de Weijer-Bergsma et al., 2015). One explanation for a decrease with age in the link between visuospatial WM and mathematics performance is that once basic arithmetic skills have been acquired, verbal-symbolic representations in WM become more dominant during math performance (Van de Weijer-Bergsma et al., 2015). In fact, it has been suggested that visuospatial WM is more strongly activated when information is novel and challenging at any age (van der Ven, van der Maas, Straatemeier, & Jansen, 2013).

High visuospatial processing has been found among mathematically gifted individuals, mostly children and adolescents (Dark & Benbow, 1994; Leikin et al., 2013; Myers & Carey, 2017). In one series of studies among gifted adolescents, mathematically talented students performed particularly well on tasks involving digits and locations, whereas verbally talented students had a special advantage with word stimuli (Dark & Benbow, 1991, 1994). More recently, both general giftedness and mathematical excellence were associated with high WM performance among adolescents, but mathematical excellence was linked with a specific advantage on visuospatial WM (Leikin et al., 2013). Others, however, found no particular advantage among younger gifted children on spatial WM, but rather an overall advantage on all WM components (Hoard et al., 2008). A review on mathematical giftedness concluded that firm conclusions could not be made due to various methodological limitations (Myers & Carey, 2017).

Some evidence in favor of non-verbal processing in mathematics among adults comes from neuro-imaging studies, in which language-independent brain areas were activated in response to complex mathematical statements only among expert mathematicians (Amalric & Dehaene, 2016). Additionally, longitudinal research has linked spatial ability – the mental representation and manipulation of complex visuospatial information - with STEM achievements beyond high school in extremely high-ability samples (Wai, Lubinski, & Benbow, 2009). However, this research focused on psychometrically assessed spatial abilities rather than on underlying mechanisms such as visuospatial WM. In addition, due to the strong focus on spatial skills, other cognitive resources such as verbal skills have received less attention, although some studies have shown significant links with math achievements (Berkowitz & Stern, 2018; Wei et al., 2012). A question thus emerges as to whether among students entering STEM study programs, high visuospatial WM will be particularly linked with mathematics performance when compared to verbal WM.

4. WM and mathematics in higher education

Studies relating WM to mathematics achievements of university students, hence much more advanced mathematics, also found positive links, although less research is available. Among college students, WM was positively correlated with both algebra achievements and performance on standardized mathematics achievements tests (Tolar et al., 2009). While these effects were mediated by computational fluency (i.e. fast arithmetic calculations) and spatial visualization, the indirect effects were moderate, suggesting that individual differences in WM remained important for high-level mathematics achievements. In this study, WM was assessed only with verbal-numerical information, thus the question of WM modality was not addressed. Positive links between WM and standardized mathematics achievement tests were found in other studies with college students (Rohde & Thompson, 2007; Wei et al., 2012). Usually, WM was not the focus of the investigation, but rather one of many correlates of math performance. For example, along with processing speed and spatial ability, WM predicted math achievements on standardized tests of undergraduates enrolled in psychology courses (Rohde & Thompson, 2007). In this study, WM was assessed by a single task with numerical content (Operation Span), and its effects were not significant when general cognitive ability was controlled for. In another study among undergraduates, visuospatial WM was tested along with other measures of visuospatial processing for its effect on advanced mathematics performance (Wei et al., 2012). Here too, a single WM task was included (Corsi Blocks), which was significantly correlated with advanced mathematics also after controlling for other correlates.

Taken together, evidence suggest that WM is positively linked with advanced mathematics achievements beyond school age. However, the available studies considered a rather limited range of WM measures, usually in one modality, and none focused on students in math-intensive programs, who are a rather high ability population. Moreover, achievements on mathematically intensive courses (which would be analogical to school grades in research among children) were not investigated, to our knowledge, for their link with WM.

5. Mediators of the WM-math relation: Prior knowledge in mathematics and numerical reasoning ability

Whereas WM plays a role in mathematics performance, its impact is often mediated by higher-order or more complex cognitive functions. Whether and when WM uniquely contributes to mathematics performance is important for understanding sources of individual differences in mathematics achievements, as well as the interplay between basic and complex cognition in determining mathematics outcomes. As mentioned above, research sometimes found that WM predicted mathematics performance beyond domain specific factors such as numerical cognition and mathematics knowledge, as well as beyond domain general factors such as IQ and non-verbal reasoning (Alloway & Alloway, 2010; Fuchs et al., 2010; Swanson & Beebe-Frankenberger, 2004; Wei et al., 2012). Other studies found that the effects of WM on mathematics outcomes were largely mediated by such factors (Fuchs et al., 2006; Rohde & Thompson, 2007; Tolar et al., 2009). For example, the positive effects of verbal WM on mathematics performance of university students found by Tolar et al. (2009) were fully mediated by basic computational performance and spatial ability, which were in turn partly mediated by prior algebra education. Similarly, WM assessed by the operation-span task (i.e. verbal WM) did not predict students' math achievements beyond general cognitive ability (Rohde & Thompson, 2007). However, visuospatial WM remained significantly correlated with mathematics performance of undergraduate students after controlling for both general and mathematics-specific factors (Wei et al., 2012).

Although both domain-general and domain specific factors influence academic learning and expertise in many domains, their relative roles in determining performance is an issue of some debate (Dochy, De Rijdt, & Dyck, 2002; Grabner, Stern, & Neubauer, 2007; Hambrick & Meinz, 2011; Schneider & Preckel, 2017; Simonsmeier, Flaig, Deiglmayr, Schalk, & Schneider, 2021; Vaci et al., 2019). Based on expertise research, one view has been that in high levels of expertise, domain-specific knowledge is critical, whereas basic abilities play a minor or even no role (Ceci & Liker, 1986; Ericsson & Lehmann, 1996; Schneider, 1993). However, others have shown that some domain-general abilities remain crucial even in high levels of expertise (Grabner et al., 2007; Hambrick & Meinz, 2011; Vaci et al., 2019). Mathematics-intensive courses at the level of higher STEM education must build on previously acquired mathematics knowledge and skills. Therefore, when focusing on advanced mathematics, domain-specific factors such as prior knowledge are expected to have a key role, which could then be contrasted with domain-general factors such as WM. In the case of students entering math-intensive undergraduate studies, the relevant knowledge should reflect mathematics topics learned in high school. Additionally, the ability to reason with numerical information, as defined within the broader construct of intelligence, is also likely to play a role. Unlike measures of school mathematics, such tests focus on the sophisticated application of rules of reasoning (deductive and inductive) on a relatively narrow type of problems (e.g. arithmetic calculations). Nonetheless, the efficiency of numerical reasoning among students who have graduated math-intensive tracks in high school, which is the case for many of those entering undergraduate STEM, is likely to be influenced by schooling as well.

6. The current study: Research questions, hypotheses and methodological approach

The aim of this study was to find out how and to what extent WM and mathematics performance are related among students entering math-intensive STEM programs at a large technological university in Switzerlan (ETH Zurich). Students attending ETH represent a high ability and mathematically competent population: Due to a selective educational system, only people with a diploma of an advanced placement high school (Gymnasium) can enroll for tertiary education. In addition, by self-selection, most of the students entering the technological university already underwent math-intensive courses at school, and the vast majority received good to very good grades in STEM subjects. Apart from these, however, there are no admission exams to the university, which means that students still vary considerably in their prior knowledge and their cognitive abilities (Berkowitz & Stern, 2018). Nonetheless, our research questions and the implications of our findings are to be viewed primarily in the context of a high-ability population. In this context, we addressed two research questions:

Research question 1: Are verbal and visuospatial WM differentially related to individual differences in various measures of mathematics performance among STEM students?

To test whether mathematics performance among our students is characterized by a specific advantage in either WM modality, we administered multiple WM-tasks with either verbal-numerical or visuospatial stimuli. Additionally, we assessed mathematics performance with three different measures: numerical reasoning ability, prior knowledge in mathematics, and academic achievements in mathematics-intensive courses our participants had to take at university. None of these mathematics measures was restricted to a narrow mathematical skill, but they differed in several respects. Numerical reasoning ability was assessed by standardized tests from an intelligence test battery, which required to draw inferences based on numerical relationships and to perform fast calculations. Variance on this test can be traced back to speed differences and to the selection of efficient solving strategies. Performance on a test covering a wide range of mathematics topics learned in high school (e.g., functions) served as a measure of prior knowledge in mathematics. Although all our participants underwent curricula dealing with these topics, considerable achievement variation still exists among these students. Course achievements were grades on the compulsory mathematics classes of calculus and linear algebra, and on math-intensive physics classes (hereafter termed *math-intensive achievements*). These courses are the most demanding and challenging ones during the first undergraduate year, and largely determine future retention of the students. With this broad range of measures, it was possible to examine whether the relations of verbal and visuospatial WM generalized beyond specific measures of mathematics performance.

As discussed in the introduction, some literature suggests that visuospatial WM is more strongly related to mathematics performance than verbal WM in young age, when basic mathematical concepts are novel (Tronsky, 2005; Van de Weijer-Bergsma et al., 2015; van der Ven et al., 2013). Applied to our group of young adults, who had already acquired substantial mathematics knowledge in high school, we would only expect special effects of visuospatial WM when acquiring new and particularly challenging mathematics knowledge. Because this happens almost continuously during undergraduate studies, we hypothesized that visuospatial WM would be more strongly related to math-intensive achievements than verbal-numerical WM. Following the same rationale, such an advantage was not expected with respect to prior knowledge in mathematics, which largely involves familiar content. As for numerical reasoning ability, although such tests are designed to introduce problems not typically learned in schools, performance does partly build on acquired and automatized knowledge. For students in our sample, we did not expect these tests to pose particularly novel problems, therefor a stronger link with visuospatial than with verbal WM was not expected for numerical reasoning ability.

In order to test these hypotheses, we conducted a latent correlational analysis. In the model shown in Fig. 1, we estimated the strength of the WM-math relation across modalities of WM and measures of mathematics performance. For each measure of mathematics performance, we tested whether its links with verbal and visuospatial WM differed by means of model comparison tests. When modeling WM by modality, as shown in Fig. 1, we did not aim to isolate domain-specific from domain-general elements in WM, but rather only to separate verbal from visuospatial elements. We assumed that each WM task involved both domain-general WM components as well as those that are specific to either modality (Kane et al., 2004; Unsworth and Engle, 2007). Thus, each of the verbal and visuo-spatial WM factors still includes domain-general elements. We did not statistically isolate modality-specific from domain-general elements because we hold the view that domain general processes (e.g., executive attention) are essential and integral to the construct as a whole and are intertwined with domain-specific components during cognitive performance. We also did not aim to capture only the *phonological loop* or only the *visuo-spatial sketchpad*, but rather the combination of each of these with the core domain general mechanism.

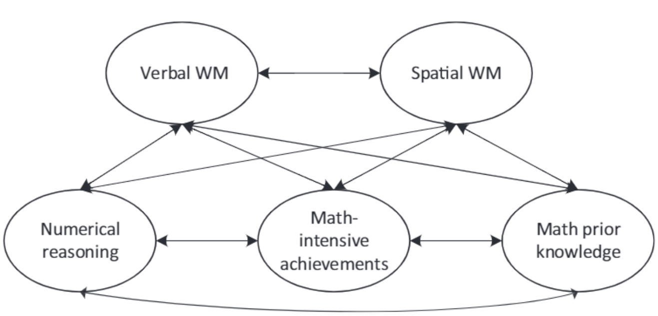

Fig. 1. Model for estimating the relation between verbal and visuospatial WM and three measures of math performance. The paths in focus of research question 1 are the overall six correlations of verbal and spatial WM with the three mathematics performance-constructs.

Research question 2: Do prior knowledge in mathematics and numerical reasoning ability mediate the relation between WM and mathematics-intensive achievements of STEM students?

As described earlier, various cognitive factors often mediate the effects of WM on mathematics performance. Our question focused on the domain-specific factors numerical reasoning ability and prior knowledge in mathematics, which are known to be strong predictors of future learning also among university students (e.g. Tolar et al., 2009). Thus, we were interested in the extent to which these domain-specific factors mediated the effects of WM on math-intensive achievements. The effects of these three factors would indicate their predictive validity: WM, prior knowledge in mathematics and numerical reasoning ability were all assessed at entry to STEM programs, whereas math-intensive achievements were gained at the end of the first academic year.

In this analysis, we were interested in comparing the effects of WM with content-specific reasoning abilities and knowledge on one mathematics outcome (i.e. achievements in first year courses), rather than comparing the effects of WM in different modalities. Because we assumed that domain-general processes in WM are essential in driving such effects, it was important that our model captured domain-general WM. This was particularly important when WM was to "compete" with other cognitive predictors. For this purpose, and in order to keep the model parsimonious, we thought that a single WM factor is most suitable. We thus selected a WM model that has been validated previously by Lewandowsky, Oberauer, Yang, and Ecker (2010). The advantages of this model are its simplicity (only four WM tasks are included), and that it was designed to minimize method variance. The tasks contain either verbal, numerical or visuospatial stimuli, and capture different functional aspects (storage and processing, relational integration). Although the spatial domain is underrepresented in this model, the verbal-spatial distinction was less crucial in this analysis, and we preferred this model with its coverage of different functional aspects. In addition, as Lewandowsky et al. (2010) noted, these tasks have been shown to load highly on domain-general WM factors; and to correlate substantially both with general reasoning tests and with other newly developed WM tasks, further supporting their selection as a good representation of general WM. Thus, in this model, prior knowledge and numerical reasoning were the domain-specific predictors, while WM was a domain-general construct with which their effects could be compared.

In accordance with the view that WM serves as a general cognitive resource for more complex cognitive activities such as learning and reasoning, we expected both prior knowledge in mathematics and numerical reasoning ability to be also related with WM, and to mediate to some extent its effects on mathematics achievements. However, while high domain-specific competencies may be necessary when facing the ambitious learning goals at a technical university, they might not be sufficient for excelling in math-intensive courses. The flexibility in restructuring existing knowledge and adapting it to new strategies places high demands on working memory (Sweller, van Merriënboer, & Paas, 2019). It is therefore entirely plausible that the relationship between WM and achievement in math intensive courses will not be fully mediated by prior knowledge and domain-specific reasoning. To find out whether there is a unique contribution of WM to math-intensive achievements, the mediation model depicted in Fig. 2 was tested.

We hypothesized that both numerical reasoning and prior knowledge in mathematics will mediate at least partly any effects of WM on math-intensive achievements. Except for the direct effect of WM on math-intensive achievements, there were three possible mediation effects in this model: one through numerical reasoning, one through prior knowledge independently of numerical reasoning, and one double mediation effect through both numerical reasoning and prior

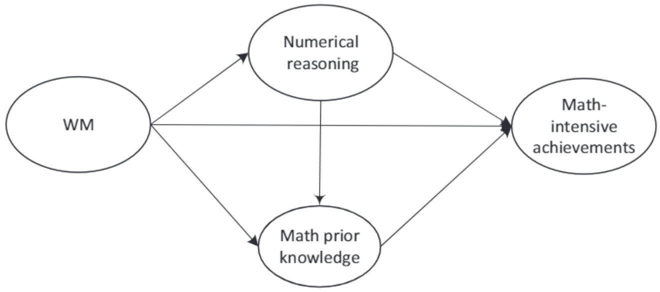

Fig. 2. A mediation model for predicting math-intensive academic achievements.

knowledge.

7. Method

7.1. Sample

Participants were 317 students (44 females) at ETH Zurich, a large technological university in Switzerland. Students were in their first undergraduate year either in the mechanical engineering program ($N =$ 150, 9% females), or the mathematics and physics programs ($N = 167$, 18% females; the math and physics programs share an identical curriculum during the first year). The major compulsory mathematics courses in both study programs are Analysis (calculus) and Linear Algebra, although the courses are taught differently in each group (e.g. more applied versus theoretical). Similarly, both groups must take physics courses of mechanics, which are extended and deepened in the mechanical engineering program, while additional physics topics are taught in the math-physics program. In both programs, physics courses rely heavily on mathematics.

Students were recruited across three consecutive cohorts (n1 = 65, n2 = 158 and n3 = 94 respectively). Mean age was 19.5 (SD = 1.4) and ranged from 17 to 25. There were no significant differences between cohorts on age ($F_{2,314} = 0.19$, $p = .83$), proportions in study programs ($\chi^2_2 = 4.29$, $p = .12$) or proportion of women ($\chi^2_2 = 2.11$, $p = .35$). The proportion of women within study programs matched those usually found at ETH. Except for a small advantage for students in the first cohort on one subtest of numerical reasoning (the number series test: $F_{2,311} = 4.39$, $p = .01$), there were no differences between cohorts on the study measures. We therefore aggregated cohorts for further analyses.

7.2. Procedure

The study was conducted in the framework of a larger research project that investigated cognitive correlates of academic achievements among STEM students. At the beginning of the first bachelor year, students performed a battery of WM tasks, cognitive ability tests and a mathematics knowledge test. Students' achievements on math and physics courses were collected at the end of the academic year, shortly after performing the first-year exams. These first-year exams are taken very seriously by both the students and the lecturers, as students are excluded from the study program if they fail twice. Grades in these written exams were determined only by performance on the exams and did not include other aspects of evaluation. The students were recruited at the very beginning of the fall semester and were offered a monetary compensation equivalent to a standard student's hourly wage for participation. Students were invited to two group-testing sessions. Numerical reasoning tests and other paper-and-pencil tests not included in this study were administered in the first session. A battery of working memory tasks was given in the second session. The sequencing of tests was fixed for all participants, sessions and samples. Scores on the School-Mathematics Knowledge Test were reported to us directly by the students or by the mathematics department, given student's consent. In fall of the following academic year, we collected students' grades on the mathematics and physics exams, which take place in summer. All participants gave their informed consent for participating and disclosing their grades before the beginning of the study. The study was approved by the ethics committee of ETH Zurich.

7.3. Measures

7.3.1. Working memory tasks

Ten working memory tasks were included in the study, five verbal-numerical and five visuospatial. All tasks were adapted and modified from previous studies, as listed below. All tasks were programmed in Matlab. All tasks began with an instructions screen followed by practice trials. A trial began with a fixation point displayed at the center of the screen for 1.5 s. Set size and number of trials are mentioned next to each task below. The order of set sizes was randomized across trials but not across participants. Each task took approximately 8–10 min to complete. Scoring of all span tasks was mean proportion of correct responses across trials (as described in Conway et al., 2005). Scoring on SSTM was based on the degree of deviation from the original pattern, as described by Lewandowsky et al. (2010).

Word span. Participants were presented with sequences of words. All words were familiar German nouns with one to three syllables. Each word was displayed for 1 s and the interval between words was 500 msec. Set size ranged from 4 to 8 and there were 15 trials in total. At the end of each trial, participants were asked to type in the words in the correct serial order one at a time using a keyboard (task adapted from Oberauer, Süß, Wilhelm, & Wittman, 2003).

Digit span. Participants were presented with sequences of digits ranging from 1 to 9. Set size ranged from 5 to 9 and there were 15 trials in total. All other features were similar to those of the Word Span task.

Corsi blocks. This task has been frequently used as a measure of visuospatial WM (e.g., Miyake et al., 2001). Participants were presented a series of 'block' locations. 'Blocks' were 2 × 2 cm white squares, which turned black one at a time. Following a fixation point, an array of 9 unevenly scattered squares appeared on white background within an area of 20 × 20 cm at the center of the screen. One of the squares turned black for 650 msec, followed by a 500 msc interval, until the next square turned black. Set size ranged from 3 to 8 and there were 18 trials in total. A self-paced break was given every 6 trials. In order to prevent learning of the squares array, six different configurations of squares were used, each appearing three times along the task. At the end of each trial, the array of white squares appeared again. Participants were asked to click on the squares which previously turned black in the correct serial order.

Arrow span. This task was a modification of a task used by Shah and Miyake (1996) and Kane et al. (2004). Participants were presented a series of arrow orientations. On each trial, an arrow appeared at the center of the screen pointing at one of 8 angles: 20°, 65°, 110°, 155°, 200°, 245°, 290°, 335° (we avoided 45° gaps because these could be easily coded as clock numbers). Arrows were displayed one at a time for 1 s each, with intervals of 500 msc between arrows. Set size ranged from 3 to 8 and there were 18 trials in total. A self-paced break was given every 6 trials. At the end of each trial, a grid appeared displaying the eight orientations. Participants were asked to click on the grid to indicate the arrow orientations in the correct serial order.

Operations span. This was a complex span task, adapted from Lewandowsky et al. (2010). The storage part included sequences of letters, all consonants excluding Y and Q. The processing part consisted of simple arithmetic equations (e.g., $3 + 2 = 5$) which were to be judged as correct or incorrect. Half of the equations were correct. On each trial, participants saw an alternating sequence of letters and equations. Each letter appeared for 1 s. Each equation appeared until a response was made or until 3 s elapsed. The interval between letters and equations was 100 ms. At the end of a trial, participants were asked to recall the letters in the correct serial order. Set size ranged from 4 to 9 and there were 18 trials in total. A self-paced break was given every 6 trials.

Reading span. This complex span task was adapted from Lewandowsky et al. (2010). The task was similar to the operations span task, except for the processing part. Instead of equations, participants were asked to judge the correctness of sentences. For example, the sentence ‘A horse is bigger than a cat’ is correct, whereas ‘In every cloud there is a spider’ is incorrect. Set size ranged from 4 to 8 and there were 15 trials in total.

Rotations span. This complex span task was modified from Kane et al. (2004) and Shah and Miyake (1996). The storage part of the task included a sequence of arrow orientations identical to those in the Arrow Span task. In the processing part, a letter (G, F or R) appeared either in its normal or mirrored form, and was rotated at one of seven orientations: 0°, 45°, 135°,180°, 225°, 270°, or 315°. Participants had to decide whether the letter was in its normal form or mirrored. The letter was normal half of the time. On each trial, participants saw an alternating sequence of arrows and rotated letters. Each arrow appeared for 1 s. Each latter appeared until a response was made or until 3 s elapsed. The interval between arrows and letters was 100 ms. Recall was done similarly to Arrow Span task. Set size ranged from 3 to 8 and there were 18 trials in total, including a self-paced break every 3 trials.

Symmetry span. This complex span task was modified from Kane et al. (2004). The storage part included a sequence of block locations identical to those in the Corsi Blocks task. In the processing part, participants saw a grid of 8 × 8 squares (1x1cm each) at the center of the screen. Some of the squares in the grid appeared black. Participants were asked to decide whether the black-square design was symmetrical along its vertical axis. The design was symmetrical half the time. On each trial, participants saw an alternating sequence of block locations and symmetry patterns. Each block-location appeared for 650 msec and each symmetry pattern appeared until a response was made or until 3 s elapsed. The interval between locations and symmetry was 100 ms. Recall was done similarly to the Corsi Blocks task. Set size ranged from 3 to 8 and there were 18 trials in total, including a self-paced break every 3 trials.

Memory updating. Task adapted from Lewandowsky et al. (2010). Participants were asked to mentally update digits through arithmetic operations. On each trial, participants saw a set of three to five digits, displayed in individual frames one by one for 1 s each. Next, arithmetic operations such as “+2” or “-4” were displayed in some or all of the frames in which digits previously appeared. Operations were displayed for 1.3 s each, followed by a 250-msec blank interval. Participants had to apply the operation to the digit that last appeared in the frame and to remember the new result. The number of operations on each trial ranged from two to six. After the last operation, final recall was signaled by question marks appearing one by one in each frame. Participants were asked to type in the remembered digit for each frame. Arithmetic operations ranged from −7 to +7, excluding 0. Interim and final results ranged from 1 to 9. There were 15 trials in total, generated by crossing set sizes of three, four, and five with the number of updating operations (two, three, four, five, and six).

Spatial short-term memory. Task adapted from Lewandowsky et al. (2010). Participants had to remember the relative location of dots in a grid. On each trial, a 10 × 10 cells grid appeared and a number of black dots were displayed one by one in individual cells. Each dot appeared for 900 msec followed by a 100 msec interval. Participants were instructed to remember the spatial relations between the dots. Absolute dot positions were irrelevant, only the overall pattern of dots was to be remembered. At the end of each trial, participants were cued to reproduce the pattern of dots by clicking the right cells. Set size ranged from two to six and there were 18 trials in total.

7.3.2. Numerical reasoning

Numerical reasoning ability was assessed by three subscales of the German intelligence structure test 2000 revised (Amthauer, Brocke, Liepmann, & Beauducel, 2001; Liepmann, 2007). These include Arithmetic Calculations (solving simple equations); Number Series (completing series by rule inference, such as 9 7 10 8 11 9 12?); and Numerical Signs (deciding which numerical operations are missing in an equation, e.g.,: 6? 2? 3 = 5). Each subscale consisted of of 20 multiple choice questions of increasing difficulty and had a time limit of 8 to 10 min. One point was given for each correct answer, so that scores range on each scale was 0–20.

7.3.3. Prior knowledge in mathematics

School-Mathematics Knowledge Test. This test was developed at the mathematics department at ETH Zurich for assessing the level of mathematical knowledge among new undergraduate students. It is offered to all new students during the first week of the fall semester and is voluntary. To participate in this study, students were asked to take this test and provide us access to their final score. The test consists of 29 multiple choice questions covering ten content areas: Algebra, trigonometry, functions, series, differential equations, integrals, vector geometry, probability, plane geometry and combinatorics. The test is not time limited and takes about one hour to complete (see https://osf.io/a8t3g for the test and further information).

7.3.4. Math-intensive achievement

Grades were collected on the courses Analysis and Linear algebra in each study group. In both groups, these courses are given yearly (i.e. in two parts across two semesters) and have one final exam at the end of the year. In addition, grades on the course Mechanics were collected in the mechanical engineering group. This course is also taught in two parts across two semesters, with a single exam at the end of the year. In the math-physics group, physics in the first semester focuses on mechanics and in the second semester on waves, electricity and magnetism. Students perform two exams at the end of the year, corresponding to each part. We therefore created one physics score for this group by averaging the grades on both exams. Exam scores are transformed into grades ranging from 1 (worst) to 6 (best) with steps of 0.25 points. We used this grades scale in our analyses.

8. Results

We begin this section with information about missing data and descriptive statistics of all study measures. Next, we report the analyses regarding the first research question (the relation of verbal and visuospatial WM with three mathematics measures). Before examining the central parameters regarding this question, we conducted a preliminary analysis of the measurement model of WM. This is followed by the main correlational analyses, in which we compare each of the three correlations of WM with mathematics achievements between verbal and visuospatial WM. Then, the analyses corresponding to the second research question regarding mediation of WM effects are described. We begin again with the measurement model of WM, followed by a mediation model.

8.1. Data screening and missing data

Missing data included the following: (1) Values over 3.5 SDs from the mean (overall 10 data points) were considered outliers and treated as missing values. (2) Grades: For 44 students (17 from engineering and 27 from math-physics), no course grades were available. The reasons for missing grades were varied and included drop out, switching to another department, postponing examination, or an unknown reason. (3) Working memory tasks: Participants in the first cohort ($n = 65$) performed only four WM tasks (digit span, operations span, memory updating and spatial STM), thus had missing values on six tasks. Data on all WM tasks were missing for two participants, who did not attend the second study session. Due to a technical error, data was not recorded for four participants on the Word Span task and for one participant on Memory Updating. (4) Scores on Prior Knowledge in Mathematics were missing for 42 participants. In order to handle these missing data analytically, we used full information maximum likelihood estimation, which can handle partially missing data of participants and avoids information loss induced by listwise deletion.

8.2. Descriptive statistics

Descriptive statistics and reliability estimates of the observed test scores are shown in Table 1. All reliability estimates were calculated within our sample, except for the test Prior Knowledge in Mathematics, as noted below. The means on the three numerical reasoning tests confirmed the high ability level of our sample: The raw values in our sample varied between 16.68 and 17.62 (Table 1), while the raw values reported in the manual for German Gymnasium students were 14.05 to 14.17, and for the corresponding general age group between 12.15 and 12.39 (Liepmann, 2007).

The observed correlations between all measures are shown in Table 2. A few scales tended to have slight negative skew (i.e. values < -1.0) (Symmetry Span, Reading Span, Number Series), but none of the scales showed visible ceiling effects. Descriptive statistics and correlations split by study program (Math/Physics and Mechanical Engineering-students) appear in the supplementary materials S1. Score distributions were highly similar across the groups.The means in both groups appeared very similar for most of the scales. The only noticeable difference was on math prior knowledge, where the math-physics group had a mean value 3.83 higher than the engineering group, which equaled a difference in Cohen's *d* of 0.74.

8.3. Statistical analyses and model fit

All latent variable analyses were conducted in M*Plus* version 8.3 (Muthen & Muthen, 2017). In part 1 (correlational analyses for the first research question), robust maximum likelihood estimation (MLR) was used, which corrects standard errors and chi-square based fit statistics for multivariate kurtosis. In part 2 (mediation models for the second research question), bias-corrected bootstrapping of standard errors was used with 10,000 bootstrap draws. This is important in models in which indirect effects such as mediation are tested, because indirect effects represent multiplied parameters, the standard errors of which do not follow known distributions (Shrout & Bolger, 2002).

We followed recommendations described by Kline (2015) for estimating model fit: Exact fit: (1) Model Chi-square with its degrees of freedom and *p* value. Approximate fit indices: (2) Steiger-Lind Root Mean Square Error of Approximation (RMSEA) (3) Bentler Comparative Fit Index (CFI) (4) Standardized Root Mean Square Residual (SRMR). We consider a model as poorly fitting when the 90% CI of the RMSEA does not exclude values >0.10, when the CFI is <0.90, or when the SRMR is >0.10. As recommended by Kline (2015), we examined the residual correlations matrices of all models regardless of fit and judged the significance of misfit based on theoretical explanations for salient residual correlations.

8.4. Addressing research question 1

As noted above, research question 1 was to find out whether verbal and visuospatial WM are differentially related to individual differences in measures of mathematics performance. In a first step, we modelled WM as consisting of two correlated but distinct latent variables, representing verbal and visuospatial WM. A confirmatory factor analysis (CFA) was conducted in which two correlated factors were specified, each explaining variance in five verbal and five visuospatial tasks respectively. Modification indices to the initial model indicated a residual association between the arrow and rotation span tasks. Such an association is reasonable, given that the two tasks assess memory for arrow orientation, while the other three spatial tasks assess memory for block location. Therefore, this residual correlation was estimated in the final model, yielding adequate model fit ($\chi^2_{33} = 73.88$, $p < .001$; *RMSEA* $= 0.063$ [CI90 $= 0.044$; 0.082]); *CFI* $= 0.956$; *SRMR* $= 0.047$). The latent correlation between verbal and visuospatial WM was 0.71 ($p < .001$), indicating about 50% overlap in true score variance. This model fit the data better than a single common factor model ($\chi^2_{34} = 154.81$, $p < .001$; *RMSEA* $= 0.106$ [CI90 $= 0.090$; 0.123]); *CFI* $= 0.871$; *SRMR* $= 0.074$).

Next, we added the three math-performance constructs to the model, which were correlated with the two latent WM-variables. Numerical reasoning and math-intensive achievements were modelled as latent variables that were identified by their three respective indicator variables. The loadings of all indicator variables were freely estimated, and the latent variables' variances fixed to unity for identification. In addition, we added a dummy-latent variable for math prior knowledge, as there was only one manifest indicator variable available for this variable. The loading of students' scores on this latent variable was freely estimated, whereas the latent variable's variance was fixed to unity, and the manifest variable's residual variance fixed to 0. In this way, this variable represented the whole variance in students' scores on the math prior knowledge-test. A modification index in this model indicated that

Table 1
Means, standard deviations and reliability estimates of all study measures.

		Mean	SD	Reliability (Cronbach's alpha)
Visuospatial WM	Arrow span	0.67	0.14	0.83
	Corsi blocks	0.80	0.08	0.65
	Rotation span	0.58	0.17	0.88
	Symmetry span	0.82	0.10	0.81
	Spatial STM	0.89	0.04	0.58
Verbal WM	Digit span	0.78	0.11	0.75
	Word span	0.70	0.12	0.83
	Operations span	0.74	0.11	0.79
	Reading span	0.74	0.18	0.89
	Memory updating	0.68	0.17	0.86
Numerical reasoning	Calculations	17.04	2.48	0.70
	Number series	17.22	2.48	0.77
	Numerical signs	17.27	2.52	0.71
PKM		19.89	5.52	0.88[a]
Course grades				
Analysis		3.98	1.17	
Linear Algebra		4.07	1.18	
Physics		4.41	0.96	

[a] Estimated in a sample of $N = 1912$ (estimates in our sample were not available). PKM = Prior knowledge in mathematics.

Table 2
Observed correlations between study variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. AS																	
2. CB	0.43																
3. RoS	0.70	0.37															
4. Sym	0.51	0.50	0.57														
5. SSTM	0.40	0.28	0.36	0.52													
6. DS	0.39	0.28	0.44	0.36	0.21												
7. WS	0.25	0.19	0.30	0.40	0.20	0.48											
8. OS	0.32	0.22	0.45	0.46	0.21	0.56	0.50										
9. RS	0.40	0.19	0.50	0.53	0.32	0.53	0.55	0.64									
10. MU	0.43	0.25	0.50	0.46	0.33	0.54	0.39	0.53	0.57								
11. Calc	0.25	0.21	0.21	0.22	0.14	0.28	0.11	0.25	0.19	0.30							
12. NR	0.33	0.33	0.34	0.27	0.23	0.26	0.22	0.26	0.24	0.39	0.40						
13. NS	0.35	0.27	0.35	0.30	0.18	0.23	0.10	0.29	0.24	0.40	0.50	0.47					
14. PKM	0.18	0.12	0.13	0.09	0.13	0.18	0.14	0.14	0.17	0.24	0.35	0.28	0.16				
15. ANL	0.14	0.14	0.09	0.09	0.11	0.26	0.12	0.11	0.13	0.17	0.31	0.18	0.17	0.36			
16. LAG	0.11	0.13	0.07	0.09	0.14	0.26	0.16	0.13	0.20	0.21	0.31	0.20	0.20	0.34	0.77		
17. PHY	0.14	0.11	0.13	0.07	0.10	0.27	0.22	0.15	0.21	0.20	0.31	0.17	0.19	0.31	0.77	0.79	

AS = Arrow span; CB=Corsi blocks; RoS = Rotation span; Sym = Symmetry span; SSTM = spatial short-term memory; DS=Digit span; WS=Word span; OS=Operations span; RS = Reading span; MU = Memory updating; Calc = Calculations; NR = Number series; NS=Numerical signs; PKM = Prior knowledge in mathematics; ANL = Analysis; LAG = Linear algebra; PHY=Physics.

Table 3
Tests of equalities of the covariances of verbal and visuospatial WM with the three outcome measures.

Model		χ^2	*df*	*p*	*SCF*	Δdf	$\Delta\chi^2$	*p*
0	No equality constraints	183.28	108	< 0.001	1.022			
1	VWM/VSWM equality with numerical reasoning	182.75	109	< 0.001	1.026	1	0.13	0.72
2	VWM/VSWM equality with prior math knowledge	184.63	109	< 0.001	1.024	1	1.41	0.23
3	VWM/VSWM equality with math-intense achievements	185.93	109	< 0.001	1.023	1	2.59	0.11

VWM = Verbal working memory; VSWM = Visuospatial WM.

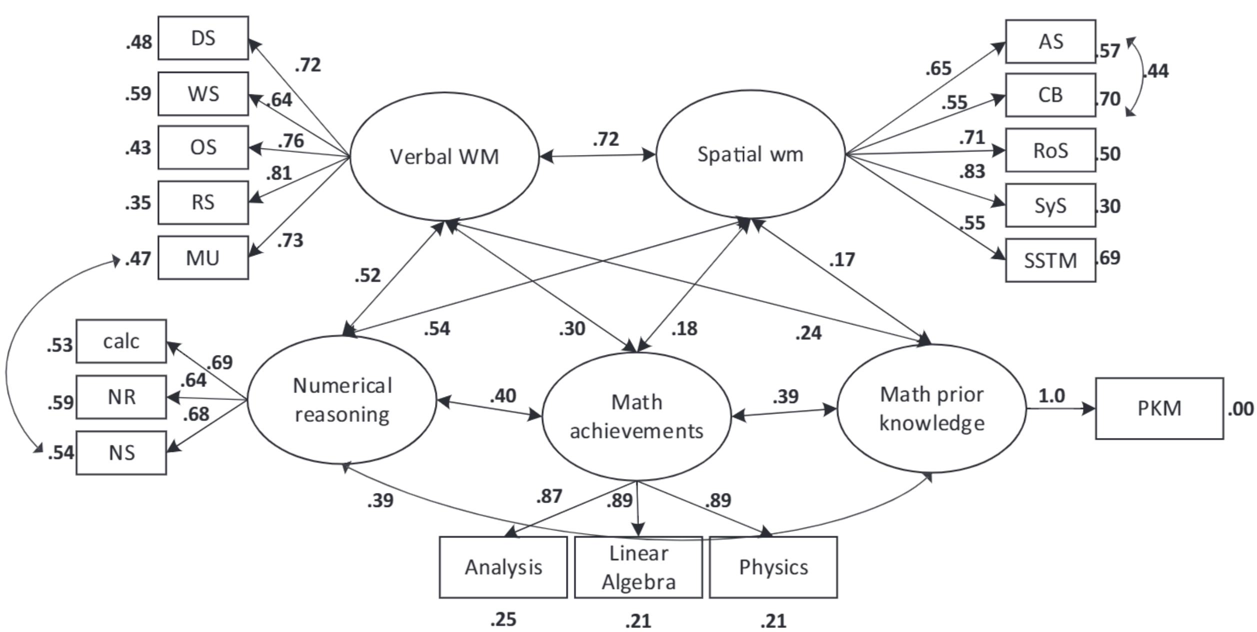

Fig. 3. Measurement model with WM and mathematics measures. AS = Arrow span; CB=Corsi blocks; RoS = Rotation span; Sym = Symmetry span; SSTM = spatial short-term memory; DS=Digit span; WS=Word span; OS=Operations span; RS = Reading span; MU = Memory updating; Calc = Calculations; NR = Number series; NS=Numerical signs; PKM = Prior knowledge in mathematics.

the residual variance of the numerical reasoning subscale Numerical Signs correlated with that of the Memory Updating WM task. We included this residual covariance between the two variables, after which the model showed adequate fit (Model 0 in Table 3; $\chi^2_{108} = 183.28$, $p < .001$, *RMSEA* = 0.047 [CI90 = 0.035; 0.058], *CFI* = 0.959, *SRMR* = 0.050) and no further outstanding residual correlations were present between the indicator variables. The resulting model is depicted in Fig. 3.

Finally, we conducted model comparison tests to examine whether there was evidence for differences in the relations of verbal and visuospatial working with each of the outcome measures. As indicated in Table 3, there were no significant differences in the strength of the relations of the two WM-facets with any of the three outcome variables. It should nonetheless be noted that descriptively, the correlation of verbal WM with math-intensive achievements appeared stronger ($r = 0.30$) than that of visuospatial WM ($r = 0.18$), with a similar trend for prior math knowledge ($r = 0.24$ and $r = 0.17$ for verbal- and visuospatial WM respectively). Thus, the central result regarding our first research question is that verbal WM was related at least as equally strongly as visuospatial WM to mathematics performance across three different measures. The hypothesis that visuospatial WM plays a unique role when it comes to the acquisition of new mathematical knowledge could not be confirmed for these STEM students. Consistent with this and with our rationale for research question 2, verbal and spatial WM were collapsed into a single latent variable in our next analyses.

In the supplementary materials S2, we provide an analysis in which we considered differences in the measurement model and correlation parameters between the two groups of students (i.e. mechanical engineering and math/physics). This additional analysis provided evidence that in both groups, the link between verbal WM and math-intensive achievements might be as strong or even stronger than that of visuospatial WM.

8.5. Addressing research question 2

The second research question was whether and to which extent the effect of WM on math-intensive achievements was mediated by prior knowledge in mathematics and numerical reasoning ability, and whether there remained a unique effect. We first examined a revised measurement model with a single latent variable of general WM, as previously established by Lewandowsky et al., 2010. In this model, scores on four tasks served as indicators of general WM: Reading Span, Operations Span, Memory Updating and Spatial Short-term memory. As in the original model, the residual correlation between Reading- and Operation Span was freely estimated. This model of WM yielded appropriate fit to the data in the entire sample: $\chi^2_1 = 2.15$, $p = .143$; $RMSEA = 0.060$ [CI90 = 0.000; 0.175]; $CFI = 0.994$; $SRMR = 0.014$.

As with the first research question, we then added the three mathematics measures to the model, which were modelled in the same manner as above. The resulting measurement model of the four constructs, which is depicted in Fig. 4, yielded appropriate fit, $\chi^2_{38} = 46.72$, $p = .157$; $RMSEA = 0.027$ [CI90 = 0.000; 0.050]; $CFI = 0.992$; $SRMR = 0.036$.

Next, we tested the central mediation model for the second research question. The fit of this mediation model, which was now estimated with bootstrapping instead of robust maximum likelihood estimation, was $\chi^2_{38} = 46.30$, $p = .167$; $RMSEA = 0.026$ [CI90 = 0.000; 0.050); $CFI = 0.993$; $SRMR = 0.036$. The resulting regression paths from the mediation model are depicted in Fig. 5, and all indirect effects of WM on math-intensive achievements are listed in Table 4.

As shown in Table 4, there was a significant total effect of WM on math-intensive achievements, which was primarily attributable to indirect effects. Specifically, most of this effect was mediated by numerical reasoning ability, which in turn had both a direct effect on math-intensive achievements, and an indirect effect via prior math knowledge. The indirect effect of WM via prior math-knowledge, and the direct effect of WM, were both small and non-significant. Taken together, the two mathematics-specific factors (numerical reasoning and prior math-knowledge) completely mediated the effects of WM on math-intensive achievements in this group of students, with numerical reasoning taking a key-role through both a direct effect and a double-

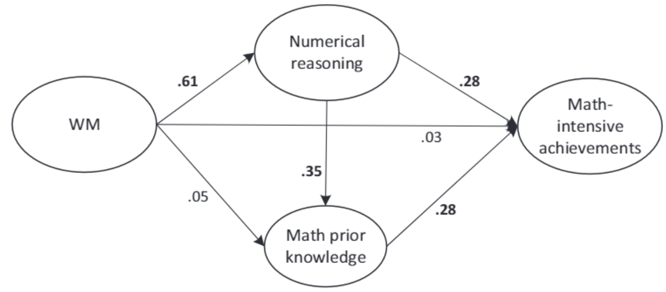

Fig. 5. Estimated regression paths in the mediation model. Values are standardized estimates (significance determined by bootstrapped CIs). Values in bold: $p < .05$.

Table 4
Direct and indirect effects of WM on math-intensive achievements.

Effect of WM on math-intensive achievements	Unstd. effect	Bootstrapped CI (95%)
Total effect	0.38	0.18; 0.61
Total indirect	0.34	0.14; 0.57
Indirect via NR	0.24	0.04; 0.47
Indirect via PKM	0.02	−0.06; 0.12
Indirect via NR and PKM	0.09	0.02; 0.17
Direct effect	0.04	−0.26; 0.35

NR = Numerical reasoning; PKM = Prior knowledge in mathematics.

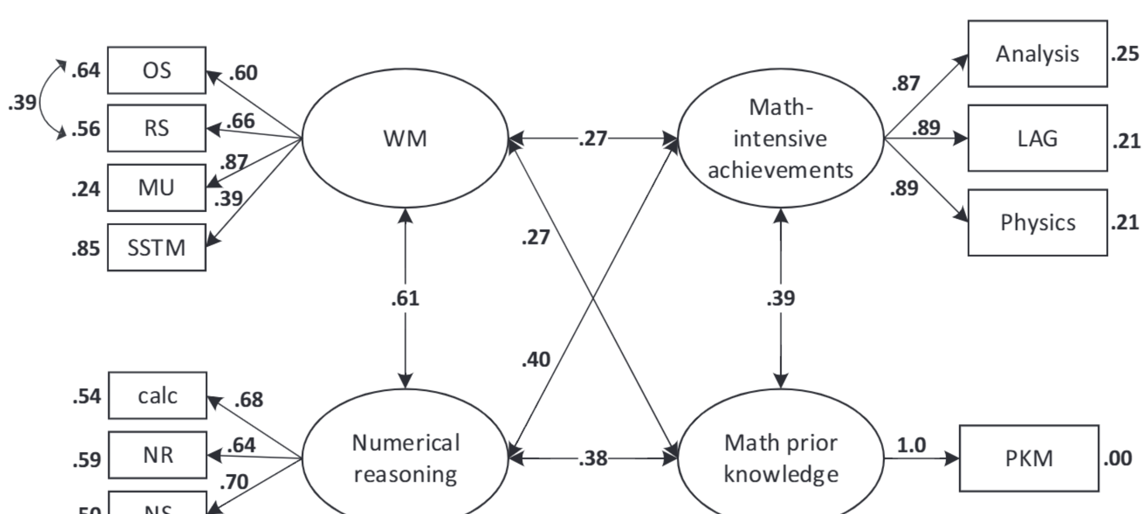

Fig. 4. Measurement model corresponding to the second research question. SSTM = spatial short-term memory; OS=Operations span; RS = Reading span; MU = Memory updating; Calc = Calculations; NR = Number series; NS=Numerical signs; PKM = Prior knowledge in mathematics; LAG = Linear algebra.

mediated indirect effect via math prior knowledge.

As for research question 1, we provide a separate version of this mediation analysis for the two groups of students in the supplementary materials (S3), in order to examine the robustness of the results. This analysis indicated that numerical reasoning might be a more substantial mediator among mechanical engineering than among math-physics students.

9. Discussion

WM plays a role in mathematics learning and performance across ages and levels of education. The magnitude of this role seems to vary across populations, mathematics measures, and WM components (Friso-van den Bos et al., 2013). In this study, we investigated the relation between WM and mathematics performance among beginner STEM students, who are a high ability and mathematically competent group. We found that WM was significantly associated with three different measures of mathematics performance among these students, and significantly contributed to the prediction of achievements on highly demanding math-intensive courses. Thus, similar to findings from K-12 education (Alloway & Alloway, 2010; Bull et al., 2008; Fuchs et al., 2010), our results point out the importance of WM also for high-level mathematics performance. Moreover, while previous research showed that higher order cognitive abilities differentiate even within the highest achievement level (Coyle et al., 2011; Ferriman-Robertson et al., 2010), the present results confirmed that this is also the case for WM.

The first goal of our study was to compare the strength of the WM-math relation between WM in the verbal and visuospatial modalities. This was to indicate whether WM in either modality is advantageous when it comes to high-level mathematics. Across three different mathematics measures, we found that visuospatial WM had no special advantage over verbal WM. We rather found that both verbal and visuospatial WM were significantly related to mathematics performance to similar extent. At the descriptive level, verbal WM even tended to have an advantage in some cases. Specifically, the strength of the relation between WM and numerical reasoning ability was highly similar between WM modalities. For both prior knowledge in mathematics and math-intensive achievements, but especially the later, verbal WM tended to show stronger correlations than visuospatial WM, though the difference was not statistically significant. An impact of verbal WM is in line with the view that verbal WM becomes more important to mathematics performance with increasing age and mathematics knowledge (Van de Weijer-Bergsma et al., 2015). These results contradict findings from younger populations, in which a specific advantage for visuospatial WM for mathematics performance emerged with age (Meyer et al., 2010). One explanation for an early advantage of visuospatial WM to mathematics performance has been its higher demand when mathematical knowledge is new. It has been additionally suggested that the link between novelty and visuospatial WM may emerge independently of age (Tronsky, 2005; van der Ven et al., 2013). Because beginner STEM students are often overwhelmed with new mathematics content, we found it plausible that visuospatial WM will emerge as particularly important in this population as well, at least with respect to mathematics learned during the first undergraduate year. However, the data did not support an advantage to visuospatial WM, and even implied a slightly stronger link between verbal WM and mathematics-intensive achievements. Therefore, the hypothesis that visuospatial WM plays a unique role when it comes to the acquisition of new mathematical knowledge could not be confirmed among STEM students. It may nonetheless be argued that because math-intensive achievements in our study were taken after a learning period of two academic semesters, it was no longer novel. We therefore cannot rule out the possibility that visuospatial WM would show stronger links with learning outcomes taken more closely to the beginning of the learning process. In addition, given that the difference in correlations strength of verbal and visuospatial WM with math-intensive achievements was not statistically significant, we suggest replicating the present findings in future studies with samples of high ability STEM-students.

The present results may seem to contradict findings regarding high visuospatial WM among high ability students that are particularly competent in mathematics (Dark & Benbow, 1991; Leikin et al., 2013). However, different from these previous studies, our study focused on individual differences within a group of students with high-mathematical competence, rather than comparing between groups with high and average mathematical competence. Thus, among students who are overall highly competent mathematically, visuospatial WM was not more advantageous to mathematics performance than verbal WM.

Overall, our results align with those showing positive links between all components of WM and mathematics performance (e.g. Bull et al., 2008), and with the idea that WM is essential in acquiring new knowledge regardless of modality (Oberauer, Süβ, Wilhelm, & Wittmann, 2008; Wilhelm et al., 2013). The positive correlations between WM in both modalities and all mathematics outcomes suggest that WM components that are not restricted to modality are central in driving these links. This brings us to the second goal of our study, which was to examine whether domain-general WM was predictive of math-intensive achievements beyond mathematics-specific factors that are more complex cognitively, namely numerical reasoning ability and prior knowledge in mathematics. The main finding from this analysis was that WM significantly predicted achievements, with mediation paths through numerical reasoning, which in turn predicted math-intensive achievements both directly and indirectly via math prior knowledge. The remaining direct effect of WM on achievements was negligible. Thus, the two constructs of numerical reasoning and math prior knowledge seem to provide one perspective on capturing those elements in WM that are relevant to math-intensive achievements. In other words, the components of WM that are essential for mathematics achievements were largely captured by indices of more complex cognitive activity that are specific to mathematical thinking. While previous studies found unique, direct effects of WM on mathematics outcomes beyond complex cognition, most of these studies were conducted among children, and involved domain general cognitive factors (e.g., IQ) as additional predictors (e.g. Alloway & Alloway, 2010). Our findings align with those showing complete mediation by such factors, as emerged with university students and when indices of complex cognition were specific to mathematics (Rohde & Thompson, 2007; Tolar et al., 2009). Nonetheless, a unique contribution of WM to mathematics performance has been previously reported also with university students (Wei et al., 2012). This inconsistency may be explained with the type of the mathematical outcome measures. For example, Wei et al. focused on novel mathematics concepts shortly after being introduced to the students, while our study and the aforementioned ones examined broader academic achievements and tests, which likely draw more strongly on prior mathematics knowledge. Finally, both numerical reasoning and prior math-knowledge had significant direct effects on math-intensive achievements in our study, indicating their complementary roles for achievements in mathematics-intensive courses.

The present study adds to existing literature in several respects. First, the results align with those from previous research in showing that individual differences in cognitive abilities even at the high range can still play a role in academic learning and achievements (e.g, Ferriman-Robertson et al., 2010). Here, we showed this to be the case not only for complex cognition, but also at the level of basic information processing. Second, although the findings do not necessarily generalize to populations that are less mathematically competent or more heterogeneous in cognitive ability, we think they are relevant to other university students who take mathematics classes. Our results indicate that differences in WM functions contribute to explaining differences in high-level mathematics achievement. There is no clear reason to assume that this will not be the case also for students taking less advanced mathematics courses, as long as these courses are comparatively demanding to the students. Finally, the role of verbal WM resources emerged to be at least as important as visuospatial resources for mathematics performance.

10. Future directions

WM is an important cognitive resource in highly demanding learning contexts, even for individuals with generally high cognitive ability and substantial prior knowledge. One hypothesis that can be tested in future research is whether WM is more likely to show unique contributions to mathematics outcomes, the more challenging the mathematics is to previously learned skills and knowledge. We have analyzed math-intensive achievements beyond specific courses and study programs, but tracking these relations while considering the particular demands of mathematics courses (e.g. the different levels of prior knowledge needed), might reveal more about the role of WM to advanced mathematics performance. Testing this hypothesis would in fact be relevant not only to high ability groups. It is additionally possible that some of the effects found here were attenuated comparing to what one might expect in a more heterogeneous sample. Therefore, WM may show even stronger links with mathematics outcomes in less selected samples. Regarding types of mathematics outcomes, although our study covered different measures, these were all relatively broad indicators of mathematics performance. Our indicators for numerical reasoning ability, for example, likely varied in the degree to which they tapped into fluid reasoning ability (e.g. number series) and quantitative knowledge (arithmetic calculations), as these are distinctively defined in known models of intelligence (McGrew, 2009). Prior mathematics knowledge and math-intensive achievements were based on a broad range of mathematics topics. Different and more nuanced results may emerge if more specific mathematics topics or problems are to be examined. For example, verbal and visuospatial WM may show different links with mathematics measures if those are selected from narrower contents, so that some clearly build on visual-spatial representations and others do not. Finally, our study focused on mathematics-specific mediators of WM, but it did not control for general reasoning ability when testing its effects on achievements. Thus, the interplay between WM, general intelligence and mathematics performance in mathematically competent students, such as those entering STEM, is still to be addressed in future research.

Acknowledgments

We are very grateful to the departments of mechanical engineering, mathematics and physics at ETH Zurich, especially to Alexander Caspar for sharing the School-Mathematics Knowledge Test; to Norbert Hungerbühler and to Maddalena Velonà. We thank Martin Frick for invaluable work in programming all WM tasks; Jessica Büetiger, Samuel Winiger, Nicole Bruegger and Anita Wildi for their assitance. We are grateful to the anonymous reviewers for their helpful feedback on this manuscript. This research was funded by the Donald C. Cooper-Fonds at ETH Zurich (2-90065-91).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2022.101649.

References

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20–29.

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences, 113*(18), 4909–4917.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Der Intelligenz-Struktur-Test 2000R: Manual [The intelligence-structure-test 2000R: Manual]*. Göttingen, Germany: Hogrefe.

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243–248. https://doi.org/10.3758/BF03194059

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Vol. 8. Psychology of learning and motivation* (pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1.

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review, 118*(2), 175.

Berkowitz, M., & Stern, E. (2018). Which cognitive abilities make the difference? Predicting academic achievements in advanced STEM studies. *Journal of Intelligence,* 6(4), 48. https://doi.org/10.3390/jintelligence6040048

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228.

Ceci, S. J., & Liker, J. K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General, 115*(3), 255.

Colom, R., Abad, F. J., Quiroga, M.Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence, 36*(6), 584–606.

Conway, A. R., Getz, S. J., Macnamara, B., & Engel de Abreu, P. M. J. (2011). Working memory and intelligence. In *, 394. The Cambridge handbook of intelligence* (p. 418).

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786.

Conway, A. R., & Kovacs, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. *Psychology of Learning and Motivation, 58,* 233–270.

Cowan, N., Morey, C. C., Chen, Z., Gilchrist, A. L., & Saults, J. S. (2008). Theory and measurement of working memory capacity limits. *Psychology of Learning and Motivation, 49,* 49–104.

Coyle, T., Snyder, A., Pillow, D., & Kochunov, P. (2011). SAT predicts GPA better for high ability subjects: Implications for Spearman's law of diminishing returns. *Personality and Individual Differences, 50*(4), 470–474.

Dark, V. J., & Benbow, C. P. (1991). Differential enhancement of working memory with mathematical versus verbal precocity. *Journal of Educational Psychology, 83*(1), 48.

Dark, V. J., & Benbow, C. P. (1994). Type of stimulus mediates the relationship between working-memory performance and type of precocity. *Intelligence, 19*(3), 337–357.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*(3), 353–386. https://doi.org/10.1080/09541440244000328

Dochy, F., De Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning: How far have we progressed since bloom? Implications for educational practice and teaching. *Active Learning in Higher Education, 3*(3), 265–284.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*(1), 273–305.

Ferriman-Robertson, K. F., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science, 19*(6), 346–351.

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

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., … Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., … Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*(6), 1731.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 18*(1), 1–16.

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics: JDBP, 32*(3), 250–263. https://doi.org/10.1097/DBP.0b013e318209edef

Ginsburg, V., van Dijck, J.-P., Previtali, P., Fias, W., & Gevers, W. (2014). The impact of verbal working memory on number–space associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(4), 976–986. https://doi.org/10.1037/a0036378

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica, 124*(3), 398–420.

Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2005). The role of working memory in higher-level cognition. In *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 104–121).

Hambrick, D. Z., & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science, 20*(5), 275–279.
Hoard, M. K., Geary, D. C., Byrd-Craven, J., & Nugent, L. (2008). Mathematical cognition in intellectually precocious first graders. *Developmental Neuropsychology, 33*(3), 251–276. https://doi.org/10.1080/87565640801982338
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.
Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35*(3), 211–223.
Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
LeFevre, J.-A., DeStefano, D., Coleman, B., & Shanahan, T. (2005). Mathematical cognition and working memory. In *Handbook of mathematical cognition* (pp. 361–377). Psychology Press.
Leikin, M., Paz-Baruch, N., & Leikin, R. (2013). Memory abilities in generally gifted and excelling-in-mathematics adolescents. *Intelligence, 41*(5), 566–578. https://doi.org/10.1016/j.intell.2013.07.018
Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571–585.
Liepmann, D. (2007). *Intelligenz-Struktur-Test 2000 R: IST 2000 R. Hogrefe*.
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10 (Elsevier).
Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences, 20*(2), 101–109.
Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621–640. https://doi.org/10.1037/0096-3445.130.4.621
Muthen, L. K., & Muthen, B. (2017). *Mplus version 8 user's guide*. Muthen & Muthen. https://books.google.ch/books?id=dgDlAQAACAAJ.
Myers, T., Carey, E., & SzHucs, D. (2017). Cognitive and neural correlates of mathematical giftedness in adults and children: A review. *Frontiers in Psychology, 8*, 1646.
Oberauer, K. (2009). *Chapter 2 Design for a working memory. In psychology of learning and motivation* (Vol. 51, pp. 45–100). Academic Press. https://doi.org/10.1016/S0079-7421(09)51002-X
Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167–193.
Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence, 36*(6), 641–652.
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*(4), 455.
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*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005
Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83–92.
Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*(6), 565.
Schneider, W. (1993). *Acquiring expertise: Determinants of exceptional performance*.
Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4.
Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422.
Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2021). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, 1–24.
Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—And a little bit more. *Intelligence, 30*(3), 261–288. https://doi.org/10.1016/S0160-2896(01)00100-3
Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*(3), 471.
Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292.
Tolar, T. D., Lederberg, A. R., & Fletcher, J. M. (2009). A structural model of algebra achievement: Computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educational Psychology, 29*(2), 239–266. https://doi.org/10.1080/01443410802708903
Trbovich, P. L., & LeFevre, J.-A. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition, 31*(5), 738–745. https://doi.org/10.3758/BF03196112
Tronsky, L. N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory & Cognition, 33*(5), 927–940.
Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038.
Vaci, N., Edelsbrunner, P., Stern, E., Neubauer, A., Bilalić, M., & Grabner, R. H. (2019). The joint influence of intelligence and practice on skill development throughout the life span. *Proceedings of the National Academy of Sciences, 116*(37), 18363–18369.
Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Memory & Cognition, 43*(3), 367–378. https://doi.org/10.3758/s13421-014-0480-4
van der Ven, S. H., van der Maas, H. L., Straatemeier, M., & Jansen, B. R. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learning and Individual Differences, 27*, 182–192.
Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817–835. https://doi.org/10.1037/a0016127
Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology Section A, 49*(1), 51–79. https://doi.org/10.1080/713755607
Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics: Cognitive correlates of advanced mathematics. *British Journal of Educational Psychology, 82*(1), 157–181. https://doi.org/10.1111/j.2044-8279.2011.02049.x
Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*, 433.