Mathematics Education Research Journal (2020) 32:189–211
https://doi.org/10.1007/s13394-020-00331-0

ORIGINAL ARTICLE

The role of spatial, verbal, numerical, and general reasoning abilities in complex word problem solving for young female and male adults

Frank Reinhold[1] · Sarah Hofer[2] · Michal Berkowitz[3] · Anselm Strohmaier[1] · Sarah Scheuerer[1] · Frieder Loch[4] · Birgit Vogel-Heuser[4] · Kristina Reiss[1,2]

Received: 30 October 2019 / Revised: 28 March 2020 / Accepted: 13 April 2020 /
Published online: 5 May 2020

Abstract

This study analyzed the relative importance of different cognitive abilities for solving complex mathematical word problems (CWPs)—a demanding task of high relevance for diverse fields and contexts. We investigated the effects of spatial, verbal, numerical, and general reasoning abilities as well as gender on CWP performance among $N = 1282$ first-year university engineering students. Generalized linear mixed models unveiled significant unique effects of spatial ability, $\beta = 0.284$, verbal ability, $\beta = 0.342$, numerical ability, $\beta = 0.164$, general reasoning, $\beta = 0.248$, and an overall gender effect in favor of male students, $\beta = 0.285$. Analyses revealed negligible to small gender effects in verbal and general reasoning ability. Despite a gender effect in spatial ability, $d = 0.48$, and numerical ability, $d = 0.30$—both in favor of male students—further analyses showed that effects of all measured cognitive abilities on CWP solving were comparable for both women and men. Our results underpin that CWP solving requires a broad facet of cognitive abilities besides mere mathematical competencies. Since gender differences in CWP solving were not fully explained by differences in the four measured cognitive abilities, gender-specific attitudes, beliefs, and emotions could be considered possible affective moderators of CWP performance.

Keywords Spatial ability · Verbal ability · Numerical ability · General reasoning ability · Complex word problems · Gender effects

Introduction

Word problems have been a vital part of mathematics education for centuries. With the shift towards more contextualized and functional modeling tasks, they saw a revival in

✉ Frank Reinhold
frank.reinhold@tum.de

Extended author information available on the last page of the article

the mathematics classroom and in the assessment of mathematical competencies (Strohmaier 2020). Word problems are considered mathematical tasks in which relevant information is presented as text rather than in mathematical notion (Boonen et al. 2016; Daroczy et al. 2015; Verschaffel et al. 2010). They require learners to integrate mathematical, linguistic, and visuo-spatial abilities (Boonen et al. 2013). During the solution, these abilities are typically not applied sequentially, but in parallel (Daroczy et al. 2015). Therefore, it is assumed that they interact during word problem solving, for example in constructing a visuo-spatial representation. However, the nature of this interaction has not yet been investigated to its full extent. In this study, we aim at understanding more about the relative importance of different individual characteristics that may contribute to word problem solving. Here, we focus on *Complex Word Problems* (CWPs) which combine multiple forms of representations (e.g., symbols, graphs, pictures), irrelevant information, notable amounts of text, and functional real-world contexts (Strohmaier 2020).

Both mathematical problem solving and spatial ability have been shown to yield gender differences favoring males. This gender gap is most consistent in the spatial domain, particularly in mental rotation tasks (Levine et al. 2016). Although gender differences in mathematics are considerably smaller and less consistent (Lindberg et al. 2010), word problems are among those areas in mathematics that have shown such differences (Casey et al. 1997; OECD 2016). Furthermore, it has been suggested that males outperform females on mathematical problems on which spatial strategies facilitate correct solutions (Gallagher et al. 2002). Thus, it is of particular interest to find out whether gender differences emerge on CWP and which factors predict such differences. Finally, research on CWP focused primarily on children, while extensive studies among adults seem underrepresented.

This study investigated the effects of spatial ability, verbal and numerical abilities, general reasoning ability, and gender on CWP performance among $N = 1282$ first-year university students in engineering-fields—where CWP solving can be considered of particular importance.

Complex word problem solving

During the last decades, national and international conceptualizations of the goals of mathematics education have shifted from skills and algorithms towards a more functional application of mathematics in the real world (CCSSO 2010; KMK 2015; NCTM 2000; Niss et al. 2016; OECD 2005). As a consequence of this shift, CWPs gained importance in mathematics education. They are considered word problems that (1) present information primarily as text rather than in mathematical notion with a syntax that does not merely mirror the mathematical task, (2) provide information that might be redundant or superficial, (3) contain multiple representations, and (4) revolve around a functional context (Strohmaier 2020). Due to these characteristics, CWPs usually address cognitive processes that go beyond algorithmical calculation and the application of factual knowledge. For example, mathematical modeling is typically addressed in CWP solving (Leiss et al. 2019; Vorhölter et al. 2019; Strohmaier 2020), and linguistic factors are of pivotal importance for their solution (Daroczy et al. 2015). With this contemporary and functional perspective, CWP solving arguably addresses

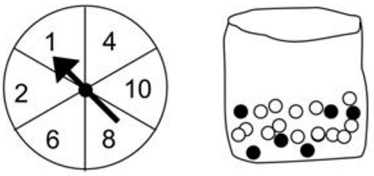
fundamental aspects of the recent goals of mathematics education (OECD 2005; Strohmaier 2020). Examples for CWPs can be found in Fig. 1 and Fig. 2.

Explaining individual differences in mathematics and complex word problem solving performance

Mathematics education typically revolves around the solution of a variety of problems which differ with regard to the cognitive processes involved in their solution. Distinguishing between these various kinds of tasks is important when addressing determinants of mathematics performance. For example, spatial abilities might have a stronger influence on geometric problems than on arithmetic tasks. For word problems, the review by Daroczy et al. (2015) offers an overview of the various factors that influence the solution process. Here, the authors specifically emphasize that linguistic and numerical factors interact during word problem solving. This finding is supported by various studies that report a strong relation between mathematics and reading abilities in CWP solving (Abedi 2006; Boonen et al. 2013; 2016; Leiss et al. 2019). In the *Programme for International Student Assessment* study (PISA)—where both mathematics and reading abilities of 15-year-olds are assessed—the latent correlation between those abilities is considerably high (OECD 2014).

With reading and modeling playing such an important part in CWP solving, cognitive processes that go beyond computational skills and factual mathematical knowledge come into play. Successfully translating a CWP into a mental model suitable for solution includes real-world knowledge, creativity, and the ability to infer missing information (Strohmaier 2020). This multicausal structure of the determinants for successful CWP solving suggests that it addresses a wide range of cognitive and motivational-affective student characteristics (OECD 2019; Schukajlow et al. 2012; Verschaffel et al. 2010). However, when research investigated these characteristics, isolated influences are often analyzed (Daroczy et al. 2015). In contrast, we assume that

A game in a booth at a spring fair involves using a spinner first. Then, **if** the spinner stops on an even number, the player is allowed to pick a marble from a bag. The spinner and the marbles in the bag are represented in the diagram below.



Prizes are given when a black marble is picked. Sue plays the game once.

How likely is it that Sue will win a prize?

A Impossible.
B Not very likely.
C About 50% likely.
D Very likely.
E Certain.

Fig. 1 Example stimulus "Spring Fair" (M471Q01). Adapted from "PISA Released Items Mathematics" (p. 63), by OECD (2006). Copyright 2006 by the OECD. Used under CC BY-NC-SA 3.0 IGO

In Zedland there are two newspapers that try to recruit sellers. The posters below show how they pay their sellers.

ZEDLAND STAR

NEED EXTRA MONEY?

SELL OUR NEWSPAPER

You will be paid:
0.20 zeds per newspaper for the first 240 papers you sell in a week, plus 0.40 zeds for each additional newspaper you sell.

ZEDLAND DAILY

WELL PAID JOB THAT TAKES LITTLE TIME!

Sell the *Zedland Daily* and make 60 zeds a week, plus an additional 0.05 zeds per newspaper you sell.

John decides to apply for a newspaper seller position. He needs to choose the *Zedland Star* or the *Zedland Daily*.

Which one of the following graphs is a correct representation of how the two newspapers pay their sellers? Circle A, B, C or D.

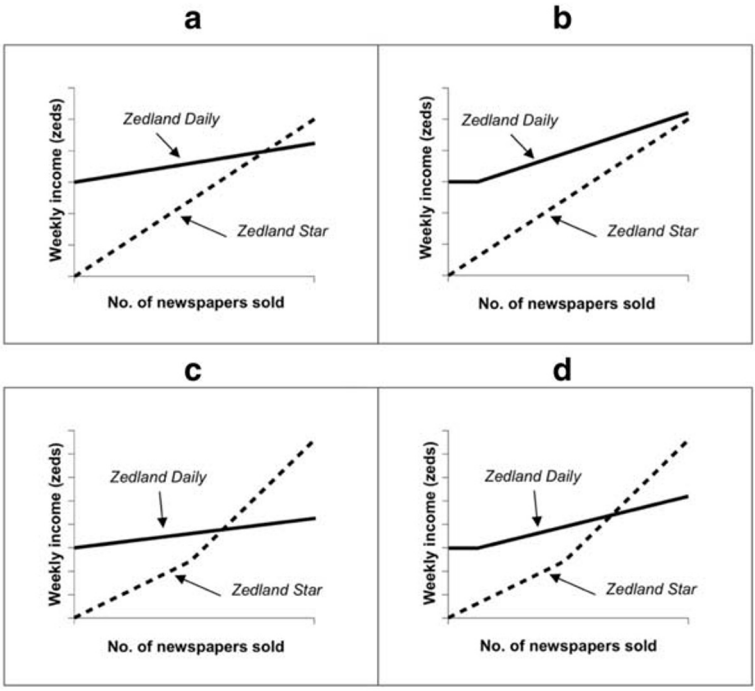


Fig. 2 Example stimulus "Selling Newspapers" (PM994Q03). Adapted from "PISA 2012 Released mathematics Items" (pp. 76–78), by OECD (2013). Copyright 2013 by the OECD. Used under CC BY-NC-SA 3.0 IGO

the variety of these characteristics causes manifold interactions in CWP solving, for example in solution strategies (Leiss et al. 2019; Strohmaier et al. 2020). Thus, this paper aims to address a broad selection of cognitive abilities—i.e., spatial, verbal, numerical, and general reasoning abilities—as well as gender, and investigates how

they can jointly account for individual differences in CWP performance. First, we review previous research that illustrates how these factors may influence CWP solving individually.

Spatial ability

Research findings suggest that there is a positive relation between spatial ability and mathematics performance in general, especially for young children (e.g., Gilligan et al. 2018; Gunderson et al. 2012; Hawes et al. 2019; Mix et al. 2016). In addition to correlational evidence, several studies found positive effects of a spatial training intervention on children's mathematics performance (Cheng and Mix 2014; Hawes et al. 2017; Verdine et al. 2017). For instance, in their classroom intervention study with 6- to 8-year-olds, Cheng and Mix (2014) found a positive effect of mental rotations training on basic arithmetic problems in the form of missing term calculations (e.g., $2 + \_ = 7$), in which reversing the operation and the given order of numbers is needed. Yet, effects of spatial ability interventions on mathematics performance are inconsistent. In another study with 6- to 8-year-olds, Hawes et al. (2015) could not find transfer effects of spatial training on missing-term calculations as well as other forms of calculations. Focusing on elementary school students, Lowrie et al. (2017) could not find a positive effect of a spatial training program on arithmetic problems, but rather on geometry problems, where the role of spatial ability may seem more obvious.

Spatial abilities are commonly assessed by standard tests such as the *Mental Rotations Test* (MRT; Vandenberg and Kuse 1978; see also Peters et al. 1995) among adolescents and adults, or simpler versions for younger age (e.g., *Spatial Transformation Task*; Levine et al. 1999). These tests typically require mental visualization and manipulation of objects, for example by rotation, moving or folding, either in two or three dimensions. The MRT has been widely used among adults. It requires a quick judgment about whether three-dimensional figures are rotated variations or mirror images of a target figure. The test is usually given with a strict time limit, which is intended to minimize reliance on cue-based and analytical solving strategies.

Performance on the MRT and similar tasks has been linked with mathematical performance in quite a number of studies (Casey et al. 1997; Geary et al. 2000), most often in the context of gender differences. Indeed, performance on the MRT—in particular—has been yielding robust and consistent gender differences favoring males (Levine et al. 2016; Linn and Petersen 1985; Peters et al. 1995, 2006). Explanations to the male advantage on spatial tasks encompass a range of biological and social factors (for review, see Miller and Halpern 2014), including gender roles and socialization, evolutionary and hormonal influences, as well as differential strategy use related to differences in accumulated experiences (Boone and Hegarty 2017; Linn and Petersen 1985). Nevertheless, in spite a consistent male advantage on standardized spatial ability tests performance, there is strong empirical evidence that these abilities can be improved by training, which works equally well for both women and men (Baenninger and Newcombe 1989; Lowrie et al. 2018; Uttal et al. 2013). Thus, gaps in spatial ability seem present across different ages, yet are not an insurmountable gender-specific difference.

Verbal, numerical, and general reasoning ability

In addition to spatial ability, there is also empirical evidence for the influence of verbal, numerical, and general reasoning abilities on mathematics performance in general (e.g., Fuchs et al. 2010). In a longitudinal study with English students from age eleven to age 16, general reasoning ability (i.e., *g* factor) correlated highly with mathematics examination performance. Verbal ability, however, did not contribute to mathematics achievement (Deary et al. 2007). In another study across different age groups—from primary to secondary school children—statistically significant direct effects on mathematics achievement were found for fluid reasoning, which comprised of numerical reasoning (e.g., completing numerical sequences), and general reasoning (Taub et al. 2008). A recent study assessing undergraduate mechanical engineering and mathematics-physics students reported that verbal and numerical abilities were associated with students' achievements on most physics and mathematics courses (Berkowitz and Stern 2018). A wide range of research has already shown that verbal abilities are of great importance for mathematical thinking and learning (e.g., Aiken 1972; Fuchs et al. 2006; Jordan et al. 2013; Leung 2017; Morgan et al. 2014; Prediger et al. 2018; Seethaler et al. 2011). They may further gain importance with increasing linguistic complexity of the mathematical content that has to be processed. For (complex) word problem solving in particular, verbal abilities are assumed to contribute to task performance (Abedi 2006; Boonen et al. 2016; Daroczy et al. 2015; Leiss et al. 2019; Strohmaier et al. 2020).

Gender differences

Comparing to gender differences found for spatial tasks, gender differences in mathematics performance—in general—have been both smaller and less consistent across time and cultures. Studies based on large scale assessments such as PISA and TIMSS—both assessing mathematics performance with CWPs—have shown that a male advantage in mathematics on such tests has considerably decreased over time (Ceci and Williams 2010; Lindberg et al. 2010). However, differences have not entirely disappeared, and some debate exists with respect to their practical significance. For instance, whereas meta analytic studies find negligible effect sizes for gender differences in mathematics at the mean level—thus calling attention for similarities between genders (Else-Quest et al. 2010; Hyde et al. 2008), other studies highlight larger differences favoring males at the upper tail of the distribution—thereby keeping open the question of gender gaps in mathematics (Makel et al. 2016; Wai et al. 2010). It is noteworthy that some of the studies linking gender differences in spatial skills with gender differences in mathematics have in fact focused on higher ability populations (Casey et al. 1997; Gallagher et al. 2002).

Another pattern of results is that a male advantage in mathematics typically emerges on standardized tests, whereas differences in favor of females appear when the outcome criteria are school grades (Halpern et al. 2007; O'Dea et al. 2018; Voyer and Voyer 2014). The reasons for these contradictory findings are not entirely understood, with some hypothesis being that standardized tests tend to cover topics that are less practiced in class—and this being a greater challenge for female students (Halpern et al. 2007), or that girls' approaches to schoolwork are better suited than boys' approaches to the learning demands of schools (Kenney-Benson et al. 2006).

In particular, word problems are one of the mathematical areas that have yielded gender differences (Carr and Alexeev 2011; Ceci and Williams 2010; Gallagher et al. 2002; Geary et al. 2000; Lindberg et al. 2010; OECD 2016), and spatial abilities have been linked with successful solving of word problems (Boonen et al. 2013; Gallagher et al. 2002; Geary et al. 2000; Hegarty and Kozhevnikov 1999). Although other abilities play a role in solving CWPs—most clearly verbal ability (e.g., Delgado and Prieto 2004)—it is assumed that constructing spatial solutions to word problems is a particularly efficient strategy (Hegarty and Kozhevnikov 1999). Consequently, lower spatial skills among females may be one possible explanation for their disadvantage on solving mathematical word problems. Given that CWP may be more challenging than traditional word problems in terms of switching between and integration of different representation types, low spatial skills may explain gender differences also in this type of problems.

The present study

As described throughout the "Introduction" section, we argue that different cognitive abilities—i.e., spatial, verbal, numerical, and general reasoning abilities—contribute to the solution of CWPs in mathematics. We illustrate this with the CWP "Spring Fair" from the PISA mathematics assessment that was also used in the present study (Fig. 1, see also OECD 2006, item number M471Q01). The item concurs with the definition of complex word problems since (1) the stimulus contains a notable amount of text, here describing both the situation at a fair and the rules of the game. (2) The item provides information that can be considered superficial for solving the problem, e.g., that the game takes place in a booth at a spring fair. (3) It contains multiple representations, i.e., the text—providing necessary information about the rules of the game, and the pictures of the spinner and the bag—providing necessary information about the numbers on the spinner and the quantity of black and white marbles in the bag. (4) The item revolves around a functional context, here a spring fair, where the reader has to decide for the player Sue how likely it is to win a price in the described game.

In this item, verbal ability should contribute to the decoding and infering of the necessary mathematical information from the text in order to solve the problem, i.e., the rules of the game. At the same time, the textual information can support the reader in building a meaningful mental model of the situation (Leiss et al. 2019), which in turn can provide a foundation for including real-world knowledge or applying heuristic strategies (possibly ruling out the answer "impossible" or imagining the process of the game). This mental model will arguably also rely on spatial abilities, and the processing of the illustrations is required to estimate (or determine) the two probabilitites of interest—i.e., *spinner stops on an even number* and *black marble is picked from the bag*. Here, an integration of text and pictures seems necessary to solve the item, which will rely on verbal and spatial abilities as well as their interaction. Besides that, numerical ability seems indispensable when aiming at the *precise solution* of how probable winning a prize when playing one time is, i.e., multiplying the two probabilities 5/6 (*spinner stops on an even number*) and 6/20 (*black marble is picked from the bag*). It seems noteworthy that getting the precise solution of 25% for Sue to win a prize is *not* required to score full credit on the item, as five answers are given in the multiple-choice format with "Not very likely" being the correct answer. Thus, starting

backwards from the given solutions is one viable strategy to solve the item—and may involve general reasoning ability. This is particularly true since even more solution processes than the ones described in detail are possible—e.g., counting, drawing a tree diagram, or utilizing the complementary event. Hence, even though the item offers various solutions that may not refer to *all* four cognitive abilities mentioned here, it seems plausible to assume that all of them support students in solving this particular problem. Since the item is exemplary with regard to the definition of CWP adopted here, a similar rationale can be applied to argue that for most CWPs, these cognitive abilities should influence the probability of solving the problem.

We consider understanding the cognitive preconditions that contribute to CWP solving in mathematics and their relative importance in predicting CWP performance an important step towards understanding individual learners' difficulties and designing effective learning environments. In addition to numerical, verbal, and general reasoning abilities, spatial ability may be closely associated with mathematics performance in general and CWP solving in particular. Due to the frequently reported gender differences—to the disadvantage of females—in (complex) word problem solving and especially in spatial abilities, analyses of predictors of CWP solving should allow for possible effects of gender.

Regarding this, we ask, to what extent are cognitive abilities—i.e., spatial, verbal, numerical, and general reasoning ability—correlated to complex word problem solving performance of young female and male adults?

We expect to replicate gender differences in CWP solving in favor of male students. According to recent studies on the role of cognitive abilities and math performance, we expect all four abilities to be correlated to CWP solving. We investigate the hierarchy of spatial, verbal, numerical, and general reasoning ability in predicting CWP performance, and potential gender-specific differences in this hierarchy, on an exploratory basis. Yet, based on the specific structure of CPWs, we assume numerical ability to be of less importance than the other three investigated cognitive abilities.

Method

To answer the research question, we conducted a cross-sectional study with first-year university students and paper-based standardized test instruments for spatial, verbal, numerical, and general reasoning abilities, as well as CWP solving.

Sample

A total of $N = 1282$ first-year university students in engineering domains (i.e., mechanical engineering, electrical engineering, civil engineering, and software engineering) took part in the study, which was conducted at a major German technical university. Among them, 328 were female students. The unbalanced gender distribution may be explained with the focus on students in the engineering domains, where female students are still underrepresented in Germany (Federal Statistical Office Destatis 2019). On average, participants were 19.98 years old ($SD = 2.73$).

Material

We utilized standardized test instruments to assess CWP solving, as well as spatial, verbal, numerical, and general reasoning abilities.

Complex word problem solving

Complex word problem solving ability was measured with items from a pool of published PISA mathematics items. More specifically, we used the following six items: M159Q04 ("Speed of Racing Car", the correct track a racing car has driven has to be assigned to a given distance-velocity diagram), PM942Q02 ("Climbing Mount Fuji," the time a protagonist needs to start his journey has to be calculated given a walking distance and two different paces), M465Q01 ("Water Tank," the correct time-height diagram representing the change in the height of the water surface has to be assigned to given shape and dimensions of a water tank), PM995Q02 ("Revolving Door," the arc length a circular revolving door can have has to be calculated so that no air can flow between entrance and exit), PM994Q03 ("Selling Newspapers," the correct graph representing the payment of two different newspapers has to be selected given information in posters, see Fig. 2), and M471Q01 ("Spring Fair," the likelihood of a protagonist winning in a game involving spinning a wheel and picking a marble has to be estimated given the spinner and the marble bag, see Fig. 1). Taken together, they covered all four content areas used in the PISA assessment (change and relationships, space and shape, quantity, as well as uncertainty and data) and are hence designed to assess different aspects of mathematical literacy (OECD 2019).

Although the items were designed to assess mathematical literacy of 15-year-old students, choosing items of above-average difficulty (i.e., given their item characteristics in the PISA field trials, these items can be considered rather complex in terms of item difficulty) resulted in an overall solution rate of 56.6% for our sample (see the "Gender differences in cognitive abilities" section), indicating that the items reflected a reasonable challenge for undergraduate students. Similar observations had been previously reported (Ehmke et al. 2005; Strohmaier et al. 2019, 2020).

Spatial ability

Spatial ability was measured with the MRT, as used by Peters et al. (1995), which is based on the original paper-pencil test by Vandenberg and Kuse (1978). The test consists of 24 items (possible scores range from 0 to 24; $\alpha = 0.91$, 95% CI [0.90, 0.91]). In each item, a three-dimensional geometrical structure built out of cubes is given. Two correctly rotated structures corresponding to the initial structure must be picked out of a selection of four (Fig. 3). Full credit is given only when both correct structures are marked (Peters et al. 1995, 2006).

Verbal ability

Verbal ability was measured with a verbal analogies scale (20 items; possible scores range from 0 to 20; $\alpha = 0.65$, 95% CI [0.67, 0.70]). In each item, a pair of words

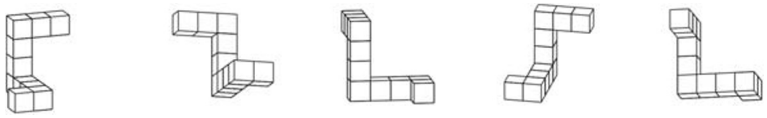


Fig. 3 Example of an item in the Mental Rotation Test (MRT). "The target is given on the left; the second and third of the four items on the right are correct matches for the target. Both have to be checked in order for subject to be given a single point. No other response (single correct, one correct, one incorrect) is credited" (Peters et al. 2006, p. 1007). Reprinted from "Mental Rotation Test Performance in Four Cross-Cultural Samples (N = 3367): Overall Sex Differences and the Role of Academic Program in Performance," by M. Peters, W. Lehmann, S. Takahira, Y. Takeuchi, and K. Jordan, 2006, *Cortex*, *42*(7), 1005–1014. Copyright 2006 by Elsevier. Reprinted with permission

representing a specific relation is given. The task is to reproduce this relation by completing another pair of words where the second word is missing. Participants can pick the correct solution out of five possible answer alternatives (part of the Intelligence Structure Test IST 2000 R, see Liepmann et al. 2007).

Numerical ability

Numerical ability was measured with a calculations scale (20 items; possible scores range from 0 to 20; $\alpha = 0.82$, 95% CI [0.81, 0.83]). Basic arithmetic problems (e.g., $1/3 \cdot 30 + 2/3 \cdot 45$) are given that must be solved without a calculator. The solutions are natural numbers (part of the Intelligence Structure Test IST 2000 R, see Liepmann et al. 2007).

General reasoning ability

General reasoning ability was measured with a short form of Raven's Advanced Progressive Matrices (12 items; possible scores range from 0 to 12; $\alpha = 0.67$, 95% CI [0.65, 0.70]). It contains a "series of homogeneous, progressively more difficult items that require the examinee to choose which piece (from eight options) best completes a pattern series presented across three rows of designs" (Arthur and Day 1994, p. 394).

Procedure

Assessment was led by a group of the authors as test instructors. The test was administered in the major lectures during the first semester of the students' university studies. We used a paper-based instrument and conducted the assessment under controlled and standardized conditions. The students received all five scales in a 40-page booklet (DIN A4). Before each scale was administered, students were told to read the specific scale's introduction, covering information about the nature of the items within that scale as well as one solved example item. If no questions were raised after reading the introduction to each scale, the test instructor told students to turn the page and start solving the items within a specific period of time: Students had 8 min for complex word problems, 3 + 3 min for the spatial ability scale (12 items in each cluster), 7 min for the verbal ability scale, 10 min for the numerical ability scale, and 12 min for the general reasoning ability scale. These time limits are recommended for the specific scales.

All students took part in the study on a voluntary basis and without reimbursement. The test was administered as a low-stakes-test and students were told that effects of study program on specific cognitive abilities were of interest.

All students were asked for their informed consent, which could be withdrawn within a time period of 2 months. Data were analyzed only after these 2 months. A proportion of 15.4% of students attending the lectures did not give consent and were excluded from the analyses.

Data and statistical analyses

In order to get an impression of the verbal, numerical, and general reasoning abilities as well as the spatial abilities of our sample, we compared them with published reference values based on similar samples via single sample *t* tests. An overview of the four different reference samples is given in Table 1.

Furthermore, we compared these abilities between the female and male students within our sample via Welch two-sample *t* tests. Analyzing a rather large sample of 1282 students, significant differences in group comparisons are very likely to occur. *P* values are hence accompanied by interpreting the corresponding effect sizes (Bakker et al. 2019). According to Cohen (1969), we consider effects lower than $d = 0.20$—and despite being significant—negligible for the purpose of this study.

Generalized linear mixed models (GLMMs) were used to estimate effects of cognitive abilities and gender on complex word problem solving (for a discussion about advantages of GLMMs over other statistical methods, see Anderson et al. 2010; Brauer and Curtin 2018). The Gender-only model contained only *Gender* (− 0.5 = female, 0 = not given, 0.5 = male) as a fixed effect. The Full model contained fixed effects for *Spatial ability*, *Verbal ability*, *Numerical ability*, and *General reasoning ability* (i.e.,

Table 1 Overview of the reference samples A to D used to characterize the sample in this study

Population characteristics (*N*)	% female/age	Original citation	Verbal	Numerical	General reasoning	Spatial (f)	Spatial (m)
This study: German first year university engineering students (1282)	26/19.98	–	12.16	15.39	8.35	9.85	12.69
A: German higher education entrance qualification (132)	–/19–20	Liepmann et al. (2007)	11.74	13.15	–	–	–
B: US American university students (202)	40/21.4	Arthur and Day (1994)	–	–	8.1	–	–
C: German university science students (73)	100/–	Peters et al. (2006)	–	–	–	12.8	–
D: German university science students (219)	100/–	Peters et al. (2006)	–	–	–	–	16.6

number of items answered correctly on the corresponding scales, all standardized at the total sample), as well as *Gender* and the four corresponding *interactions* of gender and cognitive abilities. All models allowed for random intercepts for *Students* and *Items*—considering both variance between the abilities of the students as well as the difficulty and different kinds of items. Particularly, no mean score for the six CWPs is calculated but they are considered separate observations—taking into account that they are designed to assess different aspects of mathematical literacy (OECD 2019).

Estimates are given as log-odds. They can be transformed into probabilities of obtaining a correct answer. In consequence of standardization and centering of the variables, the *Intercept* in the Full model describes the estimated probability of getting a correct answer on a CWP of average difficulty from a—neither female nor male—student with average spatial, verbal, numerical, and general reasoning abilities. We report both marginal and conditional R^2_{GLMM}—as proposed by Nakagawa and Schielzeth (2013)—as estimates for variance explained by the fixed effects only (i.e., $R^2_{GLMM(m)}$) and the entire model including random effects (i.e., $R^2_{GLMM(c)}$). In addition, we report the *Proportion Change in Variance* on the student random intercept (PCV; Merlo et al. 2005a, b; see also Nakagawa and Schielzeth 2013) as estimates for variance explained by unique fixed effects—i.e., how including specific predictors reduces variance components on the student level. All analyses were conducted in *R* (R Core Team 2008) using the 'lme4' package (Bates et al. 2015). The "pirate plots" were created using the 'yarrr' package (Phillips 2017).

Results

We first characterize our sample according to reference values on the standardized scales for spatial, verbal, numerical, and general reasoning ability before answering the research questions. We address gender differences in cognitive abilities present in our sample and different effects of cognitive abilities and gender on CWP solving.

Descriptive analyses

To characterize our sample, we compared the students' cognitive abilities on all four scales to published reference values (see Table 1 for a detailed overview of the reference samples). For that, we conducted single sample *t* tests. Our first-year university students sample showed significantly—yet negligibly—higher verbal abilities ($M = 12.16$, $SD = 3.11$) than Reference Sample A ($M = 11.74$, $SD = 3.35$), $t(1285) = 4.79$, $p < .001$, $d = 0.13$. In addition, our sample showed significantly higher numerical abilities ($M = 15.39$, $SD = 3.65$) than Reference Sample A ($M = 13.15$, $SD = 3.67$), $t(1286) = 22.04$, $p < .001$, $d = 0.61$. Regarding general reasoning, our sample also showed significantly—yet again negligibly—higher abilities ($M = 8.35$, $SD = 2.50$) than Reference Sample B ($M = 8.1$, $SD = 2.5$), $t(1284) = 3.73$, $p < .001$, $d = 0.10$. Since spatial abilities are commonly reported for gender-specific samples, we compared our female and male subsample with the corresponding reference groups. Our female students showed significantly lower spatial abilities ($M = 9.85$, $SD = 5.23$) than the

respective Reference Sample C ($M = 12.8$, $SD = 4.8$), $t(328) = -10.21$, $p < .001$, $d = 0.56$. Our male subsample also showed significantly lower spatial abilities ($M = 12.69$, $SD = 6.15$) than the corresponding Reference Sample D ($M = 16.6$, $SD = 4.6$), $t(925) = -19.36$, $p < .001$, $d = 0.64$.

To summarize, our sample of first-year university students showed average verbal and general reasoning abilities, higher numerical abilities, but lower spatial abilities regarding students of similar age and with comparable formal higher education (see Table 1). This comparison only serves the purpose to characterize the sample and is not used to answer specific research questions. Especially, it should be noted that those reference samples differ for each cognitive ability assessed.

Gender differences in cognitive abilities

We investigated whether gender differences in the four cognitive abilities were present in the sample. For that, we conducted Welch two-sample t tests. Regarding spatial ability, male students ($M = 12.69$, $SD = 6.15$) outperformed female students ($M = 9.85$, $SD = 5.23$) significantly, $t(671.22) = 8.05$, $p < .001$, $d = 0.48$. There was a significant—yet negligible—difference in verbal abilities between male students ($M = 12.28$, $SD = 3.19$) and female students ($M = 11.87$, $SD = 2.82$), $t(642.67) = 2.21$, $p < .001$, $d = 0.13$. Moreover, numerical abilities differed significantly between male students ($M = 15.71$, $SD = 3.45$) and female students ($M = 14.62$, $SD = 4.00$) in favor of the male students, $t(511.22) = 4.39$, $p < .001$, $d = 0.30$. In general reasoning ability no significant difference was found between male students ($M = 8.43$, $SD = 2.36$) and female students ($M = 8.23$, $SD = 2.51$), $t(546.93) = 1.25$, $p = .21$. To summarize, our male subsample outperformed our female subsample in both spatial and numerical abilities and showed comparable verbal and general reasoning abilities.

Descriptive values for each scale are depicted as "pirate plots" with means and standard deviations in Fig. 4, showing the distribution for the female and male subsample, as well as the total sample.

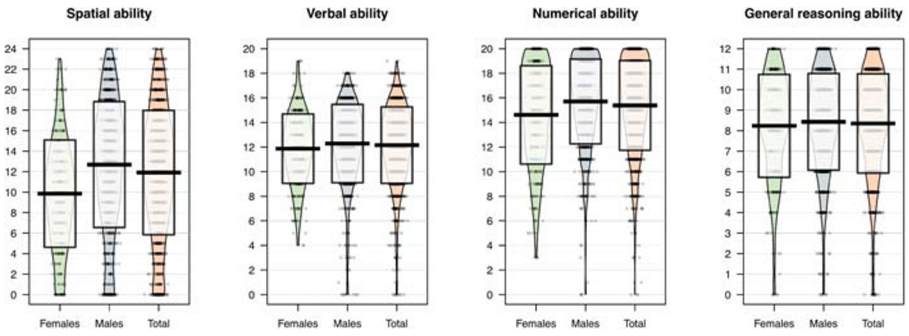


Fig. 4 Pirate plot showing the distribution of the four cognitive abilities for female and male students, as well as the total sample. Violin and scatter plots show the distribution, thick lines represent mean values, and error boxes represent ± 1 standard deviation

Effects of cognitive abilities and gender on complex word problem solving

We asked to what extent different cognitive abilities are correlated to CWP solving performance in young female and male adults. In particular, parameter estimates in a GLMM and estimates for the proportion change in variance (PCV) in the student random intercept can yield a hierarchical order of the unique effects of spatial, verbal, numerical, and general reasoning ability, as well as gender on CWP solving performance. GLMM unveiled an estimated probability of 56.6%, 95% CI [37.8, 73.7], for an average student with average cognitive abilities to get a correct answer on a CWP of average difficulty in the given assessment. Regarding this, we assumed the difficulty of the selected items to be suitable for first-year university students. As summarized in Table 2 (Full model), all further reported unique effects were significant with $p < .001$.

Spatial ability had a unique effect on CWP solving, $\beta = 0.284$, with a proportion change in variance on the student random intercept of 13.0%. Verbal ability had a unique effect on CWP solving, $\beta = 0.342$, with a proportion change in variance on the

Table 2 Parameter estimates based on the generalized linear mixed models for complex word problem solving

	Null model		Gender-only model			Full model		
Fixed effects	Est.	*SE*	Est.	*SE*	PCV	Est.	*SE*	PCV
Intercept	0.329	0.387	0.200	0.388	–	0.266	0.389	–
Spatial ability	–	–	–	–	–	0.284 ***	0.043	13.0%
Verbal ability	–	–	–	–	–	0.342 ***	0.042	19.3%
Numerical ability	–	–	–	–	–	0.164 ***	0.036	5.2%
General reasoning ability	–	–	–	–	–	0.248 ***	0.039	11.3%
Gender	–	–	0.554 ***	0.083	6.4%	0.285 ***	0.077	4.5%
× Spatial ability	–	–	–	–	–	−0.071	0.086	–
× Verbal ability	–	–	–	–	–	−0.040	0.085	–
× Numerical ability	–	–	–	–	–	−0.025	0.074	–
× General reasoning ability	–	–	–	–	–	−0.000	0.080	–
Random effects	Var.	*SD*	Var.	*SD*		Var.	*SD*	
Student	0.848	0.921	0.794	0.891		0.361	0.601	
Item	0.895	0.946	0.895	0.946		0.899	0.948	
Model fit indices	Index		Index			Index		
$R^2_{GLMM(m)}$	–		9.4%			10.2%		
$R^2_{GLMM(c)}$	–		40.1%			35.1%		
AIC	9210		9169			8785		
BIC	9231		9197			8868		

Note. 7690 observations, 1282 students, 6 items. Predictors are *z*-standardized at the total sample. Estimates are given as log-odds. *SE* = standard error, *SD* = standard deviation. PCV = proportion change in variance on the student random intercept, and R^2_{GLMM} = marginal and conditional estimates for variance explained (see Nakagawa and Schielzeth 2013). Levels of significance: ***$p < .001$

student random intercept of 19.3%. Numerical ability had a unique effect on CWP solving, $\beta = 0.164$, with a proportion change in variance on the student random intercept of 5.2%. General reasoning ability had a unique effect on CWP solving, $\beta = 0.248$, with a proportion change in variance on the student random intercept of 11.3%. Thus, verbal ability was the best predictor for individual differences in CWP solving, hierarchically followed by spatial ability, general reasoning ability, and lastly numerical ability—being the weakest cognitive predictor. In addition, there was a unique gender effect in favor for male students, $\beta = 0.285$, with a proportion change in variance on the student random intercept of 4.5%. Here, male students were on average 13.2% more likely to solve a CWP correct than female students—after controlling for cognitive abilities. Furthermore, none of the interaction effects between the four cognitive abilities and gender was significant, $.40 < ps < .99$. Thus, the effects of all measured cognitive abilities on CWP solving were comparable for both women and men, despite persistent gender differences in both spatial and numerical abilities.

In the Gender-only model (Table 2), gender had a significant and larger effect ($\beta = 0.554$) on CWP solving than in the full model ($\beta = 0.285$), with a proportion change in variance on the student random intercept of 6.4%. Thus, cognitive abilities—included in the full model—could reduce the gender effect, yet not resolve it completely. Furthermore, effects of all four cognitive abilities have shown to yield larger PCVs on the student random intercept than gender.

Discussion

We discuss the results of our study in general before going into detail on gender differences in CWP solving. Furthermore, limitations and directions for future research are given.

Complex word problem solving as a highly demanding mathematical task

Our results underpin that CWP solving is a highly demanding task, requiring a broad facet of cognitive abilities besides mere mathematical competencies. Given the estimated solution probability of our whole sample, we conclude that CWPs that showed above-average difficulty in PISA—even though developed for a younger population—were still demanding enough for our sample of first-year university students in engineering fields to investigate necessary prerequisites. This noteworthy result is in line with former findings regarding adults solving items from the PISA mathematics survey. For instance, Ehmke et al. (2005) could show that a sample of German adults solving 14 PISA mathematics items demonstrated mathematical competence comparable to the 15-year-olds sample in PISA 2000. In addition to the broad facet of cognitive abilities required to solve CWPs, another explanation for these tasks still being demanding for first-year university engineering students might be that the mathematical content knowledge acquired in regular schools after the age of 15 does not essentially contribute to solving the items, so that additional schooling in higher mathematics might not have a large impact on performance in PISA mathematics items—i.e., CWPs.

We found a significant effect of spatial ability on CWP performance. This finding is consistent with evidence from previous research for links between spatial abilities and mathematics performance more generally (Casey et al. 1997; Geary et al. 2000; Mix et al. 2016), and with arguments that spatial abilities are important in solving mathematical word problems specifically (Boonen et al. 2013; Hegarty and Kozhevnikov 1999). Differently from many studies, we examined the predictive power of spatial ability while also accounting for the effects of verbal, numerical and general reasoning abilities, thereby confirming a unique contribution of spatial performance to mathematics word problem solving. At the same time, several studies that included spatial and non-spatial ability measures did not find unique effects for spatial ability (Berkowitz and Stern 2018; Rutherford et al. 2018; Taub et al. 2008; Tolar et al. 2009). One possible explanation for this discrepancy may be the current focus on CWP solving rather than on broader measures of mathematical performance. CWP solving may be more spatially demanding than other types of mathematical competencies, as CWPs require a correct understanding of both spatial and non-spatial representations, and the efficient switching between them. Switching between representation types—especially between symbolic and non-symbolic representations—is known for its importance in mathematics more generally (Bruner 1960; Duval 2006; Lesh 1981). Additionally, it is possible that elements in the process of solving mental rotation items that are not specific to visualization have played a role as well. Recent studies on solution strategies on tests such as the MRT revealed that cue-based strategies are efficient alternatives to visualizations of rotation, and that a key for success on these tests may be the identification and selection of an optimal strategy (Boone and Hegarty 2017; Stieff et al. 2012). Yet, since we used a strictly timed task design for conducting the MRT, this could have minimized the use of cue based strategies in the present study.

We also found significant positive effects of verbal, numerical, and general reasoning abilities on CWP performance. Although the impact of these factors on mathematics performance or word problem solving, respectively, had been investigated by previous studies (Berkowitz and Stern 2018; Daroczy et al. 2015; Deary et al. 2007), their interplay was illustrated by our findings in a novel way—with our results showing verbal abilities to have the largest impact, spatial abilities and general reasoning being of equal importance, and numerical abilities showing the least positive effect. Regarding this, it seems noteworthy that—even in the investigated CWPs—verbal ability was still the strongest predictor. This finding is in line with studies supporting a strong relation between mathematical and reading abilities in CWP solving (Abedi 2006; Leiss et al. 2019; Strohmaier et al. 2019, 2020). Again, the study presented here provides a novel contribution to research on the role of linguistic factors in mathematics, since the effect remains superior when other cognitive abilities are controlled for. This supports the notion that reading is uniquely important for successful word problem solving, and arguably for learning mathematics in general. Judging from these results and the emerging importance of CWPs in mathematics education, a higher sensitivity for language-related obstacles and resources in mathematics classrooms seems plausible and timely (Boonen et al. 2016; Strohmaier 2020). At the same time, acknowledging a possible spillover in students struggling with lower verbal abilities can provide a fairer and more adequate assessment of their mathematical abilities. Finally, it supports efforts to stronger implement interdisciplinary education, integrating reading and mathematics training.

Gender differences

In line with results of existing research, male adults showed significantly higher spatial abilities than female adults in our study. Nevertheless, the results of our study imply that spatial ability does not contribute differently to CWP solving for male and female students with comparable spatial ability: spatial ability had a comparable positive effect on CWP solving for both women and men implying that the gender difference in spatial ability may not be the main factor leading to lower performance of women on CWP.

Moreover, our results are not in line with previous hypotheses that gender differences in mathematics problem solving could be explained by poorer spatial skills among females: in the present study, gender differences remained after controlling for spatial ability. This is further supported by the finding that gender differences in CWP solving could not be fully explained by differences in the four measured cognitive abilities—opening up questions about attitudes, beliefs, and emotions (Hannula et al. 2016) as possible affective moderators of mathematics achievement.

Limitations and future directions

In our study, spatial ability was operationalized only by a test of mental rotations (the MRT, see Peters et al. 1995). This test has been extensively studied before both for links with mathematical performance and with gender differences in spatial ability. However, a variety of different and well-established instruments exist that cover other aspects of spatial ability than those operationalized by the MRT. Given the present study, we cannot answer how aspects of spatial ability different from mental rotation may be related to CWP solving—or mathematical performance in general. Here, future studies might investigate the role of other aspects of spatial ability in solving CWPs to gain deeper insights in cognitive prerequisites that may explain mathematical performance.

When referring to the comparison of the sample in the present study to reference samples, one should bear in mind that these comparisons serve the purpose to characterize our sample and not to answer specific research questions. One result noteworthy in that context is that both female and male students in our sample showed lower spatial ability than the reference group: the reference sample comprised German university students in scientific fields (Peters et al. 2006)—yet, we do not know whether they were tested at the beginning or the end of their studies, while our sample was tested in the first week at university, where dropouts are still quite common.

Another noteworthy result of our study is that numerical ability had the least effect on CWP performance. When interpreting this result, it should be noted that—given the nature of the items—not all CWPs necessarily require calculations (i.e., the numerical ability that is primarily tapped in the numerical ability items used in this study) to be solved, as they are designed within the PISA mathematics framework that aims at covering different aspects of mathematical literacy with arithmetic being only one of them (OECD 2019). Regarding the items used in our study, a more qualitative approach investigating necessary prerequisites to solve each item separately might be fruitful to gain insights into what item characteristics require which cognitive ability. Given the item "Spring Fair" (M471Q01, see Fig. 1), for instance, the likelihood of Sue winning a prize only needs to be estimated—and especially only be given in prosaic form and not

as a numerical value. One could expect that if the student had to calculate the probability of Sue winning a prize, the outcome in this CWP would more strongly depend on numerical ability and perhaps less on general reasoning ability. This could be investigated via altering the questions in the CWPs to focus on exact numerical results that need to be calculated. However, given that "Spring Fair"—as an item from the PISA mathematics survey—aims at the application of mathematical knowledge in real-world contexts, it should also be discussed whether calculating the exact numeric solution would really reflect a real-world situation more authentically than the estimation of the likelihood in prosaic form.

Our results indicate the importance of verbal abilities. In the analysis presented here, the importance of verbal ability was quantitatively shown, yet a more qualitative approach could help specifying the processes that account for this relation. This remains a key mission for future research (Daroczy et al. 2015; Strohmaier 2020). For example, the analysis of process data and think-aloud protocols could reveal how verbal and mathematical thinking interact (Strohmaier 2020; Strohmaier et al. in press). Fostering an interplay between abilities, rather than preferring one form of thinking over another, may thus be an important pathway in preparing students for successful performance on CWP solving.

Another open question that may arise from the results of this study concerns possible intervention effects. For the development of CWP solving in particular—or mathematical competence in general—it might be of interest whether interventions on specific cognitive capacities, such as spatial or verbal ability, will result in an increase in CWP performance.

Finally, the item random intercept in the present study is rather high, suggesting difference in the solutions between the different items. Yet, while in all CWPs used in our study, the amount of text is substantial and needs to be captured adequately to solve the item, this difference may likely be due to other item characteristics, such as the presence of a complex visual stimulus (see Fig. 2), which may result in a larger need for spatial ability. Regarding that, future studies could seek to derive such qualitatively different characteristics of CWPs (e.g., complex visual stimuli) that may account for specific cognitive abilities (e.g., a stronger connection to spatial reasoning).

Conclusion

As shown in this study, prerequisites for CWP solving are manifold. Thus, we suggest to combine their instruction in mathematics education in schools—understanding both CWP solving in particular and mathematics in general not only as mere calculation, but as an interplay of various cognitive abilities. Instruction in school following this understanding of mathematics should build on the integration of spatial, verbal, general reasoning, and mathematical training.

Funding Information Open Access funding provided by Projekt DEAL.

References

Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Lawrence Erlbaum.

Aiken, L. R. (1972). Language factors in learning mathematics. *Review of Educational Research, 42*, 359–385. https://doi.org/10.3102/00346543042003359.

Anderson, C. J., Verkuilen, J., & Johnson, T. R. (2010). *Applied generalized linear mixed models: Continuous and discrete data for the social and behavioral sciences*. New York: Springer.

Arthur, W., & Day, D. V. (1994). Development of a short form for the raven advanced progressive matrices test. *Educational and Psychological Measurement, 54*(2), 394–403. https://doi.org/10.1177/0013164494054002013.

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*(5–6), 327–344. https://doi.org/10.1007/BF00287729.

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics, 102*(1), 1–8. https://doi.org/10.1007/s10649-019-09908-4.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

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.

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1005–1019. https://doi.org/10.1037/xlm0000370.

Boonen, A. J. H., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology, 38*(3), 271–279. https://doi.org/10.1016/j.cedpsych.2013.05.001.

Boonen, A. J. H., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology, 7*, 191. https://doi.org/10.3389/fpsyg.2016.00191.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389–411. https://doi.org/10.1037/met0000159.

Bruner, J. S. (1960). *The process of education*. Cambridge: Harvard University Press.

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology, 103*(3), 617–631. https://doi.org/10.1037/a0023864.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*(4), 669–680. https://doi.org/10.1037/0012-1649.33.4.669.

CCSSO. (2010). *Common Core state standards for mathematics*. Washington DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19*(5), 275–279. https://doi.org/10.1177/0963721410383241.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences* (1st ed.). New York, NY: Academic Press.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*, 348. https://doi.org/10.3389/fpsyg.2015.00348.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence, 32*(1), 25–32. https://doi.org/10.1016/S0160-2896(03)00061-8.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics, 61*(1–2), 103–131. https://doi.org/10.1007/s10649-006-0400-z.

Ehmke, T., Wild, E., & Müller-Kalhoff, T. (2005). Comparing adult mathematical literacy with PISA students: Results of a pilot study. *Zentralblatt für Didaktik der Mathematik, 37*(3), 159–167. https://doi.org/10.1007/s11858-005-0005-5.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. https://doi.org/10.1037/a0018053.

Federal Statistical Office Destatis. (2019). *Studierende an Hochschulen. Vorbericht. Fachserie 11 Reihe 4.1. Wintersemester 2018/2019* [Students at Universities. Winter Semester 2018/2019]. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/studierende-hochschulen-vorb-2110410198004.pdf. Accessed 27 February 2019.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & 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), 29–43. https://doi.org/10.1037/0022-0663.98.1.29.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., Bryant, J. D., & Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*(6), 1731–1746. https://doi.org/10.1037/a0020662.

Gallagher, A., Levin, J., & Cahalan, C. (2002). Cognitive Patterns of Gender Differences on Mathematics Admissions Tests. *ETS Research Report Series, 2002(2), i–30*. https://doi.org/10.1002/j.2333-8504.2002.tb01886.x

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*(4), 337–353. https://doi.org/10.1006/jecp.2000.2594.

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2018). The developmental relations between spatial cognition and mathematics in primary school children. *Developmental Science, 22*(4), e12786. https://doi.org/10.1111/desc.12786.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48*(5), 1229–1241. https://doi.org/10.1037/a0027433.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x.

Hannula, M. S., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanim, E., et al. (2016). *Attitudes, beliefs, motivation and identity in mathematics education. An overview of the field and future directions*. Hamburg, Germany: Springer Open. https://doi.org/10.1007/978-3-319-32811-9.

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4*(3), 60–68. https://doi.org/10.1016/j.tine.2015.05.001.

Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing children's spatial and numerical skills through a dynamic spatial approach to early geometry instruction: Effects of a 32-week intervention. *Cognition and Instruction, 35*(3), 236–264. https://doi.org/10.1080/07370008.2017.1323902.

Hawes, Z., Moss, J., Caswell, B., Seo, J., & Ansari, D. (2019). Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach. *Cognitive Psychology, 109*, 68–90. https://doi.org/10.1016/j.cogpsych.2018.12.002.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*(5888), 494–495. https://doi.org/10.1126/science.1160364.

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116*(1), 45–58. https://doi.org/10.1016/j.jecp.2013.02.001.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology, 42*(1), 11–26. https://doi.org/10.1037/0012-1649.42.1.11.

KMK. (2015). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife [Educational standards for mathematics as part of the general higher education entrance qualification]*. Cologne, Germany: Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany.

Leiss, D., Plath, J., & Schwippert, K. (2019). Language and mathematics—Key factors influencing the comprehension process in reality-based tasks. *Mathematical Thinking and Learning, 21*(2), 131–153. https://doi.org/10.1080/10986065.2019.1570835.

Lesh, R. (1981). Applied mathematical problem solving. *Educational Studies in Mathematics, 12*(2), 235–264. https://doi.org/10.1007/BF00305624.

Leung, F. K. S. (2017). Making sense of mathematics achievement in East Asia: Does culture really matter? In G. Kaiser (Ed.), *Proceedings of the 13th International Congress on Mathematical Education* (pp. 201–218). Springer, Cham. https://doi.org/10.1007/978-3-319-62597-3_13.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35*(4), 940–949. https://doi.org/10.1037/0012-1649.35.4.940.

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *WIREs Cognitive Science, 7*(2), 127–155. https://doi.org/10.1002/wcs.1380.

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R [Intelligence Structure Test 2000 R]* (2nd ed.). Göttingen, Germany: Hogrefe.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123–1135. https://doi.org/10.1037/a0021276.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479–1498. https://doi.org/10.2307/1130467.

Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *British Journal of Educational Psychology, 87*(2), 170–186. https://doi.org/10.1111/bjep.12142.

Lowrie, T., Logan, T., Harris, D., & Hegarty, M. (2018). The impact of an intervention program on students' spatial reasoning: Student engagement through mathematics-enhanced learning activities. *Cognitive Research: Principles and Implications, 3*, 50. https://doi.org/10.1186/s41235-018-0147-y.

Makel, M. C., Wai, J., Peairs, K., & Putallaz, M. (2016). Sex differences in the right tail of cognitive abilities: An update and cross cultural extension. *Intelligence, 59*, 8–15. https://doi.org/10.1016/j.intell.2016.09.003.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Råstam, L. (2005a). A brief conceptual tutorial on multilevel analysis in social epidemiology: Interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *Journal of Epidemiology & Community Health, 59*(12), 1022–1029. https://doi.org/10.1136/jech.2004.028035.

Merlo, J., Yang, M., Chaix, B., Lynch, J., & Råstam, L. (2005b). A brief conceptual tutorial on multilevel analysis in social epidemiology: Investigating contextual phenomena in different groups of people. *Journal of Epidemiology & Community Health, 59*(9), 729–736. https://doi.org/10.1136/jech.2004.023929.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*(1), 37–45. https://doi.org/10.1016/j.tics.2013.10.011.

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145*(9), 1206–1227. https://doi.org/10.1037/xge0000182.

Morgan, C., Craig, T., Schuette, M., & Wagner, D. (2014). Language and communication in mathematics education: An overview of research in the field. *ZDM Mathematics Education, 46*(6), 843–853. https://doi.org/10.1007/s11858-014-0624-9.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Niss, M., Bruder, R., Planas, N., Turner, R., & Villa-Ochoa, J. A. (2016). Survey team on: Conceptualisation of the role of competencies, knowing and knowledge in mathematics education research. *ZDM Mathematics Education, 48*(5), 611–632. https://doi.org/10.1007/s11858-016-0799-3.

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications, 9*(1), 3777. https://doi.org/10.1038/s41467-018-06292-0.

OECD. (2005). *Definition and selection of key competencies: Executive summary*. Paris: OECD Publishing. https://www.oecd.org/pisa/35070367.pdf.

OECD. (2006). *PISA released items mathematics*. Paris: OECD Publishing. https://www.oecd.org/pisa/38709418.pdf.

OECD. (2013). *PISA 2012 released mathematics items*. Paris: OECD Publishing. http://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf.

OECD. (2014). *PISA 2012 technical report*. Paris: OECD Publishing. https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

OECD. (2016). *PISA 2015 results (volume I): Excellence and equity in education*. Paris: OECD Publishing. https://doi.org/10.1787/9789264266490-en.

OECD. (2019). *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing. https://doi.org/10.1787/b25efab8-en.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test – Different versions and factors that affect performance. *Brain and Cognition, 28*(1), 39–58. https://doi.org/10.1006/brcg.1995.1032.

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (N = 3367): Overall sex differences and the role of academic program in performance. *Cortex, 42*(7), 1005–1014. https://doi.org/10.1016/S0010-9452(08)70206-5.

Phillips, N. (2017). Yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R". https://CRAN.R-project.org/package=yarrr. Accessed 19 April 2017.

Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E., & Benholz, C. (2018). Language proficiency and mathematics achievement: Empirical study of language-induced obstacles in a high stakes test, the central exam ZP10. *Journal für Mathematik-Didaktik, 39*(S1), 1–26. https://doi.org/10.1007/s13138-018-0126-3.

R Core Team. (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rutherford, T., Karamarkovich, S. M., & Lee, D. S. (2018). Is the spatial/math connection unique? Associations between mental rotation and elementary mathematics and English achievement. *Learning and Individual Differences, 62*, 180–199. https://doi.org/10.1016/j.lindif.2018.01.014.

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics, 79*(2), 215–237. https://doi.org/10.1007/s10649-011-9341-2.

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21*(5), 536–542. https://doi.org/10.1016/j.lindif.2011.05.002.

Stieff, M., Ryu, M., Dixon, B., & Hegarty, M. (2012). The role of spatial ability and strategy preference for spatial problem solving in organic chemistry. *Journal of Chemical Education, 89*(7), 854–859. https://doi.org/10.1021/ed200071d.

Strohmaier, A. R., (2020). *When reading meets mathematics. Using eye movements to analyze complex word problem solving*. [Doctoral dissertation, Technical University of Munich]. https://mediatum.ub.tum.de/?id=1521471.

Strohmaier, A. R., Lehner, M. C., Beitlich, J. T., & Reiss, K. M. (2019). Eye movements during mathematical word problem solving—Global measures and individual differences. *Journal für Mathematik-Didaktik, 40*(2), 255–287. https://doi.org/10.1007/s13138-019-00144-0.

Strohmaier, A. R., Schiepe-Tiska, A., Chang, Y.-P., Müller, F., Lin, F.-L., & Reiss, K. M. (2020). Comparing eye movements during mathematical word problem solving in Chinese and German. *ZDM Mathematics Education*. https://doi.org/10.1007/s11858-019-01080-6.

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (in press). Eye-tracking methodology in mathematics education research: A systematic literature review. Educational Studies in Mathematics. https://doi.org/10.1007/s10649-020-09948-1.

Taub, G. E., Keith, T. Z., Floyd, R. G., & Mcgrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*(2), 187–198. https://doi.org/10.1037/1045-3830.23.2.187.

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.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402. https://doi.org/10.1037/a0028446.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). *Link between spatial and mathematical skills across the preschool years*. Hoboken: Wiley-Blackwell.

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

Vorhölter, K., Greefrath, G., Borromeo Ferri, R., Leiß, D., & Schukajlow, S. (2019). Mathematical modelling. In H. N. Jahnke & L. Hefendehl-Hebeker (Eds.), *Traditions in German-speaking mathematics education research* (pp. 91–114). Cham: Springer. https://doi.org/10.1007/978-3-030-11069-7_4.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.1037/a0036620.

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence, 38*(4), 412–423. https://doi.org/10.1016/j.intell.2010.04.006.

Affiliations

Frank Reinhold[1] · Sarah Hofer[2] · Michal Berkowitz[3] · Anselm Strohmaier[1] · Sarah Scheuerer[1] · Frieder Loch[4] · Birgit Vogel-Heuser[4] · Kristina Reiss[1,2]

Sarah Hofer
sarah.hofer@tum.de

Michal Berkowitz
michal.berkowitz@ifv.gess.ethz.ch

Anselm Strohmaier
anselm.strohmaier@tum.de

Sarah Scheuerer
sarah.scheuerer@tum.de

Frieder Loch
frieder.loch@tum.de

Birgit Vogel-Heuser
vogel-heuser@tum.de

Kristina Reiss
kristina.reiss@tum.de

[1] Heinz Nixdorf-Chair of Mathematics Education; TUM School of Education, Technical University of Munich, Munich, Germany

[2] Centre for International Student Assessment (ZIB); TUM School of Education, Technical University of Munich, Munich, Germany

[3] Chair for Research on Learning and Instruction, ETH Zurich, Zürich, Switzerland

[4] Institute of Automation and Information Systems; Department of Mechanical Engineering, Technical University of Munich, Munich, Germany