International Journal of Mathematical Education in Science and Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tmes20

Building digital cube houses to improve mental rotation skills

Christian Bokhove & Edward S. Redhead

To cite this article: Christian Bokhove & Edward S. Redhead (14 Jan 2025): Building digital cube houses to improve mental rotation skills, International Journal of Mathematical Education in Science and Technology, DOI: 10.1080/0020739X.2024.2441889

To link to this article: https://doi.org/10.1080/0020739X.2024.2441889

Published online: 14 Jan 2025.

INTERNATIONAL JOURNAL OF MATHEMATICAL EDUCATION IN SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/0020739X.2024.2441889

Building digital cube houses to improve mental rotation skills

Christian Bokhove [a] and Edward S. Redhead [b]

[a]Southampton Education School, University of Southampton, Southampton, UK; [b]School of Psychology, University of Southampton, Southampton, UK

ABSTRACT

Prior research indicates that spatial skills, such as mental rotation skills (MRS), are a strong predictor for mathematics achievement, while other studies have also shown that MRS can be improved through training. This paper explores whether a well-known puzzle-oriented tool for building houses with 3D cubes is effective in improving performance in a standardised MRS measure that recorded accuracy and speed. The field experiment took place with 85 year 7 (11–12 year olds) pupils from an independent secondary school in the south of England. We used two conditions in the experiment, with the puzzle-oriented training tool being the intervention condition. The findings show there was a significant effect for accuracy but not for speed. Contrary to prior research our findings did not show any gender effects. The findings and implications are discussed in light of the existing literature around spatial skills, as well as design aspects.

ARTICLE HISTORY
Received 10 June 2024

KEYWORDS
Mathematics education; mental rotation skills; spatial tasks; field experiment

1. Introduction

Since a renewed mathematics curriculum in England from 2014 (Department for Education, 2013) and continuing worries about mathematics achievement in England from an international perspective, for example from the Programme of International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), policy-makers are constantly looking for ways to improve general mathematics achievement. This paper focuses one such suggested ways, through training an influential spatial sub-component of mathematics, namely Mental Rotation Skills (MRS). As MRS are suggested to be strong predictors of achievement in science, technology, engineering and maths (STEM) subjects (Casey et al., 1992), the expectation is that mathematics achievement might improve, although MRS achievement might be subject to gender differences (Voyer et al., 1995). Apart from the research-informed expectations there also are arguments for MRS in terms of concrete classroom applications.

As a former mathematics teacher, the first author regularly used a 3D ‘Building houses with side views’ application in his mathematics lessons. These applications seemed to motivate students and -anecdotally- seemed to improve students’ spatial awareness. In the tool,

CONTACT Christian Bokhove C.Bokhove@soton.ac.uk Southampton Education School, University of Southampton, Highfield Campus, Building 32 Room 2035, Southampton SO17 1BJ, UK

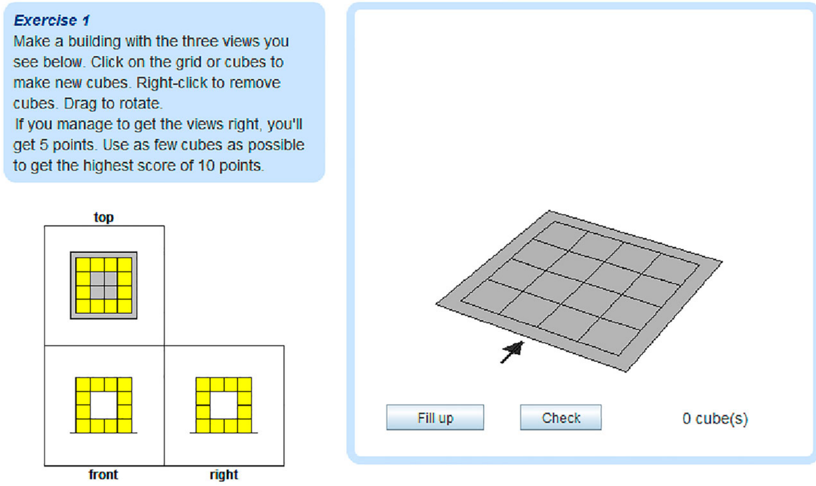

Figure 1. Screenshot of the 'Building houses with side views' tool.

depicted in Figure 1, students need to re-create a 'cube building' with the three views bottom left: one top view, one front view, and one right view. On a tablet, students add blocks by tapping the screen and remove blocks by tapping an existing cube for a slightly longer time. Furthermore, students could rotate the building with their finger.

The application, developed by Boon (2009), became so successful in the Netherlands that it was adopted by the premier publisher of secondary education mathematics books. In England, the application was included in the popular 'Improving Learning in Mathematics' materials under the topic 'Representing 3D shapes' (Young, 2016). For many teachers, it seemed the combination of 3D rotation and a clever interface with challenging tasks in which students had to solve puzzles, with more points given for the most economical solution, was a 'winner'. Students and teachers alike seemed to enjoy the building app as well. In other words, it partly was the 'digital experience' that contributed to the success of the building houses application. Despite its popularity, though, to our knowledge, the software application has not been used a lot in experimental studies. Teachers seemed happy to offer up one or more lessons to this application, and from a motivational perspective this could of course be justified. But would MRS improve after using the app for one or more lessons? The aim of this study was to explore the effects on MRS of using the 'Building houses with side views' application.

This paper explores whether MRS were trainable with a custom tool and reports on an experiment with a validated online MRS measurement tool, created in the same authoring environment as the main application. The structure of the paper is as follows: we start with an overview of the prior literature on the relationship between MRS and mathematics achievement, ways in which these MRS can be improved, and the influence of gender differences. Then we describe the design of our study with specific attention to the 'Building houses' application and the validated online MRS measurement, as well as the tasks in the MRS training tool. Finally, we report on the findings of a classroom experiment in a secondary school, formulating conclusions we can draw from this.

2. Literature review

We explore prior research on MRS by first looking at the relationship between MRS and mathematics achievement, interventions that might improve MRS, and what factors mediate MRS. We then turn to relevant studies on 'spatial skills' in mathematics education.

2.1. The relationship between mental rotation skills and mathematics achievement

Mental rotation is a cognitive process that is used in many practical applications, especially ones involving orientation, and as such has been studied extensively by cognitive scientists and neuroscientists. Bruce and Hawes (2015) highlight how mental rotation skills can be applied to a variety of mathematical tasks, including those pertaining to understanding area measurement tasks, the composition and decomposition of 2D and 3D figures, and tasks involving symmetry. Such skills also predict other spatial abilities like wayfinding, orientation and route learning (Nori et al., 2006), especially in children (Fenner et al., 2000; Merrill et al., 2016). Spatial abilities also come into play with verbal and visual-spatial working memory (Kaufman, 2007) and overall problem solving (Geary et al., 2000). Research shows that mental rotation skills may provide children with a 'cognitive tool' that can be used in their school mathematics (Bruce & Hawes, 2015) with the application of 3D mental rotation abilities extending to numerous topics in mathematics curricula, like school geometry (Delgado & Prieto, 2004), algebra (Tolar et al., 2009), numerical estimation (Tam et al., 2019), word problems (Hegarty & Kozhevnikov, 1999), mental mathematics (Kyttälä & Lehto, 2008), and advanced mathematics (Wei et al., 2012). Other researchers also point out the consistently strong relationship between spatial visualisation skills and a breadth of mathematical tasks, indicating far transfer (Mix & Cheng, 2012).

Several studies suggest that spatial skills strongly predict achievement and attainment in science, technology, engineering, and mathematics fields (e.g. Shea et al., 2001; Uttal et al., 2013), with this even extending to very young children (Lauer & Lourenco, 2016). Other evidence also suggests that spatial tasks are related to arithmetical and mathematical performance. For example, Dumontheil and Klingberg (2011) found that brain activity predicted arithmetical performance two years later in healthy participants aged 6–16 years with spatial visualisation skills especially seeming to be related to mathematical thinking (Hawes et al., 2019a; Mix & Cheng, 2012). A quasi-experimental spatial intervention study by Cheung et al. (2019) with 6- to 7-year olds ($N = 62$) showed near transfer gains in mental rotation ability and far transfer to arithmetic performance. Such results of far transfer are not uniformly positive though. A meta-analysis by Xie et al. (2019) with 73 studies and 263 effect sizes found a significant positive association between spatial ability and mathematical ability ($r = 0.27$, 95% confidence interval (CI) [0.24, 0.32]), but that much depends on the mathematical topic, with for example logical reasoning having a stronger association with spatial ability than numerical or arithmetic ability. This prior literature suggests that improvements in MRS could potentially be beneficial for mathematical outcomes.

2.2. Mental rotation skills can be improved through training

King et al. (2019) reported that a meta-analysis of twin studies suggested spatial visualisation abilities are largely heritable, with shared environmental factors having little effect.

Knowing that MRS is an important predictor of mathematical ability is one thing, but if we cannot improve MRS by specific training, this knowledge will not help improve mathematics classrooms. A second underpinning assumption in this study is that MRS can be improved by training. In a meta-analysis of 217 spatial training studies, Uttal et al. (2013) found that mental rotation training can lead to stable gains in MRS and that they transferred to other spatial tasks that were not directly trained. For example, training on mental rotation led to improvements in the same type of task, but also in different spatial tasks like mental paper folding (e.g. Chu & Kita, 2011; Wright et al., 2008).

A study by Meneghetti et al. (2017) showed short-term effects in the mental rotation training groups in terms of accuracy. A study by Cornu et al. (2017) used a tablet-based spatial intervention in kindergarten classrooms ($N = 125$) with an aim to improve children's visuo-spatial and numerical abilities. The 400 min of training over a 10-week period led to domain-specific improvements on spatial skills. Other differential factors played a role in training improvements. Cheng and Mix (2014) found evidence that mental rotation training improved maths performance in 6–8-year olds, but they highlight that certain spatial tasks are more like certain maths tasks. More recently, Mix et al. (2021) assessed whether training to improve spatial skills also improved mathematics performance in elementary-aged children. Spatial training led to better overall mathematics performance for both 7- and 12-year-old children. Spatial skills can be improved in people of all ages and through a wide range of training approaches, including course work, task-based training and video games. In a recent meta-analysis, Hawes et al. (2022) confirmed again that spatial training may benefit mathematics outcomes in middle childhood and early adolescence, although they emphasise that effects were stronger if training and outcomes were aligned. In this study, our starting point is that MRS can be improved by training in secondary mathematics classrooms.

2.3. Factors that explain differences in mental rotation skills training

If we look at the relation between spatial visualisation and numerical reasoning, Hawes and Ansari (2020) note that more variables than cognitive factors are at play, including, age, sex, type of mathematics instruction, past experiences with spatial learning, and various socioemotional and affective factors. In the context of the study, we need to take some of these into account. We have already seen in the section on training that interventions that target audiences ranged considerably from very young children to older students. Focusing on our building app's 3D mental rotation focus specifically, some researchers have asserted that such spatial tasks are too challenging for elementary school children (Hoyek et al., 2012), suggesting additionally that children struggle with mental rotation of 2D representations of 3D cube-figures. According to Jansen et al. (2013), such tasks are 'too difficult for children between the ages of 8–10 years to process' (p. 60). However, Mix et al. (2021) found their training effective for both 7- and 12-year olds. Gender differences in spatial ability are also well documented (e.g. Loring-Meier & Halpern, 1999; Voyer et al., 1995; Voyer et al., 2017), with males showing an advantage. Such differences do, however, depend on the type of spatial test as well. For example, the mental rotations test appears to produce the most robust sex differences among all tests included in their analysis.

Apart from the content, it might also depend on the scoring and testing procedures, including perhaps whether the environment is 'pen and paper' or virtual (Parsons et al., 2004). Parsons et al. (2004) noted that gender differences were replicated with paper-and-pencil measures, while in a virtual environment this was not the case. As this study aims to utilise an online virtual environment, it will be important to look at the gender factor. Uttal et al. (2013) reported that several studies that have used extensive training have indeed found that the gender gap can be reduced and perhaps eliminated (e.g. Feng et al., 2007). There also have been numerous training studies that have shown that individual differences in performance prior to training influence subsequent improvements (Just & Carpenter, 1985; Terlecki et al., 2008). For example, Uttal et al. (2013), Terlecki et al. (2008) who showed that 'female participants who initially scored poorly improved slowly at first but improved more later in training. In contrast, males and females with initially higher scores improved the most early in training' (p. 367). Sorby (2009) highlights strategies from a previous meta-analysis by Linn and Petersen (1985) where males outperform females on mental rotation tasks where speed was included as factor. Males more often utilised a so-called 'holistic strategy', visualising the whole object, while females were more likely to use an 'analytic strategy', a systematic, stepwise approach. A holistic strategy has been found to be less time consuming and therefore potentially favour timed tests. As in this study, we include time and accuracy, it will be important to take gender into account.

2.4. Spatial skills in mathematics education research

Although spatial skills have had ample attention in the world of psychological research, there also is a rich history in terms of mathematics education research. In 2020, a special issue edited by Resnick et al. (2020) covered a range of topics, including the theme of transfer from spatial reasoning to mathematics understanding (Lowrie et al., 2020), their role in complex word problem solving (Reinhold et al., 2020) and spatial skills concerning 2D representations of 3D geometrical shapes (Fujita et al., 2020). Other research has also highlighted numerous relevant variables in research on spatial abilities. For example, Harris et al. (2021) included primary and secondary school students in their study, and found that 'different spatial skills can and may be utilised at different time points and in different mathematical areas' (p. 509). Rahe and Quaiser-Pohl (2021) investigated the relationship between gender, mathematical anxiety and (perceived) mental-rotation performance and found gender differences for these outcomes increased with age. Our specific focus on cube buildings also has a rich history in mathematics education research. For example, the seminal work by Battista and Clements (1996, 1998) looked at such cube buildings and front, top and side views, noting that a challenge student face is to integrate the views of a building to form a coherent mental model of it. Fowler et al. (2021) implemented a technologically enhanced STEM programme on the spatial reasoning of a cohort of year 7 students, with results indicating that the learning activities positively impacted attitudes towards STEM and overall spatial skills, but did not improve spatial sub-skills or in relation to specific activities.

Work by Ben-Chaim et al. (1985, 1988) also used cube building interventions with initial differences in spatial visualisation performance by grade (increasing by age), gender (favouring boys) and site (increasing with socio-economic status). However, the performance gain for both genders was similar, despite those initial differences. Safadel et al. (2023) focused on mentally rotating 3D objects and concluded that a significant positive relationship between spatial self-efficacy and spatial ability disappeared when gender and major (STEM or non-STEM) were considered. In other words, variables like context, gender and technology use all seem rather ubiquitous variables in spatial research. In this journal, the ubiquitous nature of spatial skills also is apparent, for example with studies on designing levels for video games (Albarracín et al., 2022), augmented reality teaching materials for 3D geometry thinking skills (İbili et al., 2020), or other intervention units for improving students' spatial skills (Patkin & Dayan, 2013). In other mathematics education research, the actual design of spatial skills experiences has also had extensive attention, for example in geometry content contexts, coding (Dickson et al., 2022; Francis et al., 2016), physical models (Cumino et al., 2021) and fostering spatial ability though computer-aided design (e.g. Dilling & Vogler, 2021). Lowrie et al. (2020) recommended that researchers examine what is happening in classrooms to understand how implementation of spatial intervention impacts student learning, and therefore, we think there is a convincing rationale for this study. They describe a continuum to conceptualise spatial learning, with on one side such learning entirely embedded within a mathematics context, and on the other side an entirely controlled context. In this study, we took research from a setting with undergraduate students (Bokhove & Redhead, 2017) to the current study with secondary school students.

2.5. The current study

In this study, we posit that MRS training in England's grade 7 (11–12-year olds) could be effective but that we would have to take gender and the type of task into account. We therefore designed a classroom experiment with an intervention tool, measuring MRS before and after the intervention. The current study aims to answer the following research questions:

(1) Does the 'Building houses with side views' intervention lead to significant improvements in MRS speed and accuracy of a sample of year 7 students?
(2) Is there a significant gender effect regarding MRS speed and accuracy when a sample of year 7 students uses the 'Building houses with side views' intervention?

3. Methodology

As we set out to evaluate the effectiveness of an MRS intervention, the 'Building houses with side views' tool, we decided on an experimental design for a field experiment in the classroom. However, given the naturalistic setting in classrooms, we want to use a research design with still some form of random assignment in the conditions.

3.1. Design and participants

The field experiment took place in four year 7 classrooms (in June, so most students will have been 12 years old), with 85 students, in an independent secondary school in the south of England. The study uses a design with two groups, with the 85 students randomly allocated to the conditions; they only differed in the order in which the activities are presented, not in the content of the intervention, so we could analyse differential effects. Figure 2 schematically represents the study design. P1 indicates a first 'pre' MRS measurement; P2

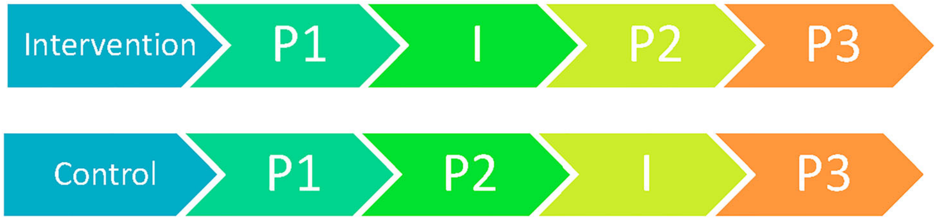

Figure 2. Design for the study. Px are measurements, I is the intervention.

was a 'post' MRS measurement for the intervention group and a second 'pre' measurement for the control group; P3 is a third and final MRS measurement. Students completed all measurements and the intervention during one mathematics lesson of 50 min.

We assert that this design allows us to use one of the groups as control, but still offer the intervention to all the students; normally this is a limitation for schools partaking with a control group. We opted for this design because administering a possibly effective intervention to only a selection of students was deemed ethically problematic. Sometimes this is solved by offering the tool after a delay period, but this not deemed a suitable solution here, as the school would continue with the regular curriculum content. The chosen design allowed for randomisation at the student level without disadvantaging or priming students. An *active control* with a replacement task was not deemed useful in the context of schooling, as for example a replacement task for a control involving crosswords is not educationally useful. Instead, we wanted to make sure all students eventually had the opportunity to do the tasks. By comparing measurements at P1 and P2, we should be able to see the effect of the intervention. Although conditions were randomly assigned to the conditions, after assignment it turned out there was a disbalance with one group's size being 36 and the other group's size 49, together making up the total sample of 85. In the data analysis, this was reduced further because of missing data (see below). These students were part of four classes with around 20 students each. Of the 85, 51 students were boys and 34 students were girls. Classes in year 7 at this school had individual tablet computers available. Figure 3 gives an impression of the typical setup in each classroom. Ethical approval was sought and gained from the university ethics board, adhering to the University Ethics Policy. Participants consented to the collection of data.

3.2. The 'Building houses with side views' application

The technical implementation of the intervention is based on tried and tested applets for building houses (Boon, 2009), as already presented in Figure 1. Throughout the years, as designer of the app, Boon has explained the principles behind the design in several places (2009), including how it is important it is to find 'a balance between giving the user freedom in his constructions and explorations in the virtual environment and imposing constraints to guide the user to intended experiences' (Boon, 2006, p. 39). In line with the principles of Realistic Mathematics Education (e.g. Van den Heuvel-Panhuizen & Drijvers, 2014), the elementary objects that are the 'building blocks' are 'real' for the students: they have a meaningful foundation (Boon, 2006, p. 43). Furthermore, the app provides direct feedback

Figure 3. Study in action in the classroom.

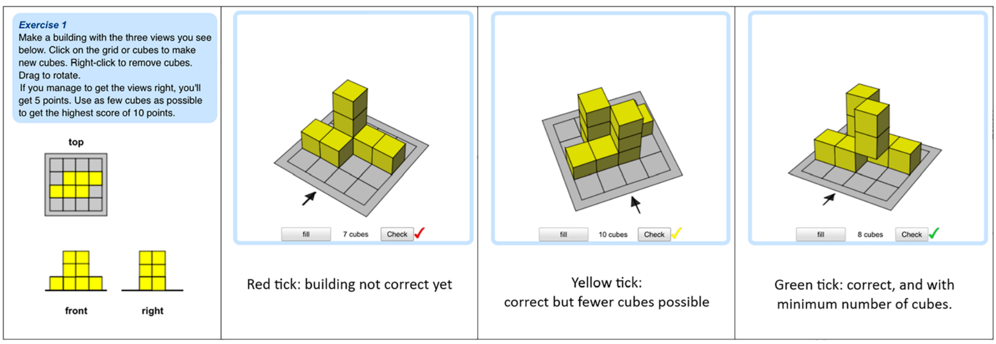

Figure 4. Three types of feedback ('ticks') in the 'Building houses with side views' tool.

with every mouse-click showing 'what the consequences are, for example dragging the mouse or rotating' (Mackrell & Bokhove, 2017, p. 69).

Students were first given one example task as explanation and then were given 2 times 10 tasks, 20 in total, without necessarily the intention the students would complete all of the tasks. In building the 'houses', cubes are allowed to 'float'. There is a scoring and feedback system that motivates pupils to find optimal solutions for the given tasks. Students can check their building by tapping the 'check' button. If a solution is correct, a yellow tick will appear (correct views). However, only if the minimal number of cubes is used a green tick will appear and the maximum score gained. Figure 4 illustrates the three possible outcomes.

The strategies students could use were left open, and retrospectively it became clear that some students focussed on getting a 'yellow' tick for as many tasks as possible, while others wanted to get the ideal minimum (green) solution first, before continuing to the next task. The intervention was run within a dedicated learning environment for mathematics, storing student answers and providing log files. Figure 5 shows a screenshot of the backend. As researchers we could check the scores by all participants, on the left-hand side of the figure, and a real-time representation of the students' answers on the right-hand side of

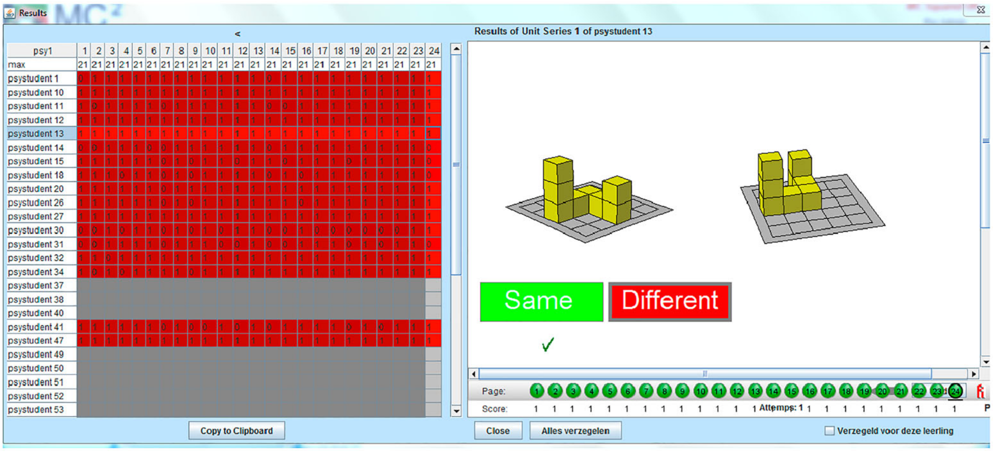

Figure 5. The researcher 'back-end' where we could see scores, student work and can download raw data.

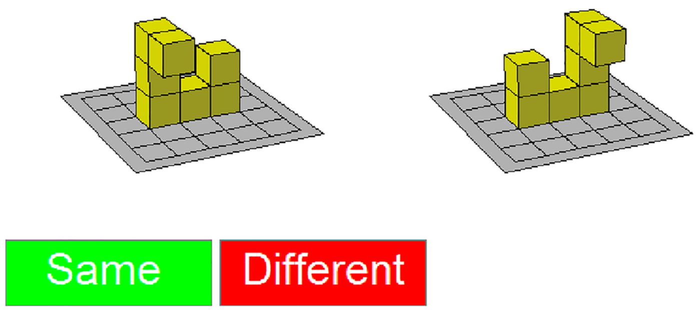

Figure 6. Example of Mental Rotation Skill assessment item.

the figure. Tasks completed correctly as well as timings are stored by the environment. In addition to the MRS measurement and the intervention tool, we added some questions regarding demographics, mental effort and an open question asking what they liked or did not like about the intervention. As both conditions had used the tool ('I' in Figure 1), we could use the results from both groups.

3.3. Mental rotation skills measurement

The MRS measurement task was based on Ganis and Kievit's (2015) redesign of an existing standardised MRS measure (Shepard & Metzler, 1971). Stimuli are composed of a pair of three-dimensional objects, where the baseline object has to be compared to a target object rotated over 0, 50, 100 or 150 degrees. Participants are asked to mentally rotate the target object to determine whether it can be brought into alignment with the baseline object. The measurement includes two types of stimuli: 'same' stimuli, where the two objects can be brought into alignment via rotation, and 'different' stimuli, where this is not possible. Figure 6 shows what one of the test items looks like. Students had to indicate whether two block buildings were the 'same' or 'different' by clicking on the corresponding choice.

The measurement instrument was validated in a separate study, in which we found that the online tool behaved similar to the original MRS measurement (Bokhove & Redhead, 2017). Our study made a few changes to the original measurement by Ganis and Kievit (2015). The measurement tool in this study uses two different blocks of 24 trials each (P1 and P2), with the P1 block also being used for P3. The four orientations (0, 50, 100 and 150 degrees) occur equally often. On half of the trials, the objects in a pair were the same. The order of the trials was kept the same as the original validation study that randomised them, with no more than three same or different trials occurring consecutively. Another difference with the original study by Ganis and Kievit (2015) was that we relaxed the 7500 ms time limitation per test item. We felt this would not be appropriate for a classroom setting, as the emphasis was on spatial skills not on speed. We will return to this aspect in the discussion section of this article.

3.4. Variables and data analysis

We statistically analyse the data with two dependent variables, namely the accuracy (percentage correct in the standardised mental rotation skills measure) and the speed (total time spent to answer the test items of standardised mental rotation skills measure). Although measures of speed and accuracy are sometimes combined, we argue with Vandierendonck (2017) that separate inclusion can also be usefully inspected. In addition, we recorded gender. In total, we have collected the following variables:

gender denoting the gender of the student (0 = girl/female, 1 = boy/male).

accuracy1 and **accuracy2**: percentage correct for P1 and P2 (accuracy, 0–100).

time1 and **time2**: completion times for P1 and P2 in seconds (speed).

We also initially collected and included a measure of mental effort, a 9-point Likert scale indicating 'very very low mental effort' to 'very very high mental effort' (Paas, 1992). This scale has been widely used in cognitive load theory. However, we note that there were quite a few missing values for this measure. A final question provided a textbox and asked students' experiences with the intervention, as we wanted to know whether they enjoyed using the tool. Descriptive statistics were calculated for all the variables. We then conducted a multivariate linear regression. Analyses were performed with JASP 0.11.1.0. Before we present the results, it is useful to highlight that some students in the study only completed one measurement out of P2 and P3; we do not know why this was the case. As could be seen in Figure 1, this would not really make a difference for the intervention group, however if the students in the control group had only made P3, they would not be 'control' students any more, as they would have taken the intervention before the post-measurement. For this reason, we removed those students, which reduced the total sample size. We also removed students with the outlier of answering the questions in less than 10 s.[1] This left 62 students for analysis, 25 females and 37 males, with unfortunately a disbalance between control and intervention, 19 in the control and 43 in the intervention. This disbalance also affected the gender balance, something we discuss towards the end of the paper.

4. Results

We first report the descriptive statistics. Table 1 presents the descriptive statistics for the continuous variables by group (c = control, i = intervention). We can see that on average at both measurement points a similar percentage of MRS assessment items was answered correctly, seemingly indicating no improvement on accuracy. As could be expected the completion times the second time round were much lower, which seems an improvement in speed. However, on both accuracy and speed the standard deviation at the second measurement point was much greater.

Given our specific gender focus, Table 2 shows the four key variables for accuracy and speed for two genders. Only for **accuracy1** was there a significant difference between males and females, $t(60) = -4.65, p < .001$. Although this could have been caused by the gender disbalance between control and intervention groups, it is notable this did not result in differences in the other measurements.

To further check the suspicion of no improvement in accuracy but improvement in speed, a multivariate regression analysis provided further insight. We first used the post-test accuracy2 as dependent variable, and as covariates the pre-test accuracy1 (as 'prior knowledge'), intervention and gender, as well as the interaction between intervention and gender. Accuracy1 and the intervention were significant, as can be seen in Table 3. There were no gender effects. The directions of the estimates seem intuitively plausible as positive: knowing more beforehand predicts better accuracy afterwards and the intervention predicts accuracy2. In sum, our results do seem to indicate a modest improvement of accuracy after our intervention.

In Table 4 ,we tabulate a similar analysis with the post-test time2 as dependent variable, and as covariates the pre-test speed time1, intervention and gender, as well as the interaction between intervention and gender. Only time1 was significant. There again were no gender effects. Those students that needed more time beforehand were more likely to need more time in the post-measurement time2 as well. Note that, although not significant, the direction of the intervention estimate is negative, which makes sense, as more practice should reduce the time to answer similar questions accurately.

Table 1. Descriptive statistics for continuous variables, c = control, i = intervention.

	accuracy1		time1		accuracy2		time2		gender		clt	
	c	I	c	i	c	i	c	i	c	i	c	i
Valid	19	43	19	43	19	43	19	43	19	43	15	28
Missing	0	0	0	0	0	0	0	0	0	0	4	15
Mean	91.31	85.74	299.79	328.93	86.11	87.30	270.21	240.93	0.74	0.54	6.93	6.93
Std. Deviation	10.72	13.34	91.00	99.36	23.31	14.66	142.50	121.01	0.45	0.51	1.75	1.59

Table 2. Differences between males and females on key variables.

		M	SD			*M*	SD
accuracy1***	Female	79.5	15.1	**accuracy2**	Female	82.1	17.6
	Male	92.8	7.2		Male	90.2	17
time1	Female	329	104	**time2**	Female	266	144
	Male	314	93		Male	239	116

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 3. Coefficients for regression, dependent variable post-test accuracy2.

Coefficients	Unstandardised	Standard error	Standardised	*t*	*p*
(Intercept)	−3.356	14.440		−0.232	.817
accuracy1***	0.907	0.163	0.660	5.572	<.001
intervention*	16.621	6.898	0.440	2.410	.019
gender	8.986	7.594	0.253	1.183	.242
intervention * gender	−15.994	8.347	−0.443	−1.916	.060

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4. Coefficients for regression, dependent variable post-test speed time2.

Coefficients	Unstandardised	Standard Error	Standardised	*t*	*p*
(Intercept)	148.817	82.169		1.811	.075
time1*	0.388	0.169	0.296	2.293	.026
intervention	−13.298	62.512	−0.048	−0.213	.832
gender	6.723	66.151	0.026	0.102	.919
intervention * gender	−48.503	76.744	−0.185	−0.632	.530

* $p < .05$, ** $p < .01$, *** $p < .001$.

The qualitative response to the question about enjoyment showed that both genders very much liked the puzzle-oriented intervention tool. Almost exclusively, unrelated to gender, comments were very favourable. Some examples were:

- ‘It was highly enjoyable but challenging’
- ‘Great fun’
- ‘I love this I love this I love this’

Sometimes there were more critical comments along the lines of ‘challenging’ and ‘This is annoying’ because the student could not find the required solution. This aspect of ‘challenge’ also could be seen in the way students approached the ‘green ticks’, the ideal solution. The best students in the trial could at a maximum only find these ‘perfect’ solutions for 10 of the 20 available tasks. Some students seemed a bit frustrated when they could not get the green ticks.

5. Discussion

This classroom experiment set out to answer two research questions: (1) Does the ‘Building houses with side views’ intervention lead to significant improvements in MRS speed and accuracy of a sample of year 7 students? and (2) Is there a significant gender effect regarding MRS speed and accuracy when a sample of year 7 students uses the ‘Building houses with side views’ intervention? As expected, prior accuracy and speed predicted post-test achievement. For accuracy as dependent variable being part of the intervention group was a significant predictor of post-test achievement. However, this was not the case for speed. In other words, answering the first research question, the intervention seems to help with MRS accuracy but not with speed. Answering the second research question, no gender effects were found for both MRS accuracy and speed, nor any interactions between the intervention and gender. We will now discuss these findings more widely.

5.1. Absence of gender differences

Our findings did not show any gender effects. We hypothesise this might be due to several reasons. As our literature review also showed that age was an important factor in spatial skills, our choice for year 7 (12-year olds), was expected to result in gender differences, as around 10 years of age, boys tend to outperform girls on measures of mental rotation (Halpern et al., 2007; Titze et al., 2010). One possible explanation for an absence of gender effects could have been the absence of a time limit in the MRS measurement. Linn and Petersen (1985), for example, already noted that males outperform females on mental rotation tasks where speed of performance is a factor, because males more often utilised a so-called 'holistic strategy', visualising the whole object, while females were more likely to use an 'analytic strategy', a systematic, stepwise approach. A holistic strategy has been found to be less time consuming and therefore potentially favour timed tests. In that context the absence of the 7500-ms time limit for each MRS measurement task, as described in the methodology, perhaps actually helped females in doing just as well as males. Interestingly, further scrutiny of the average times used for the 24 test items showed that students on average took around 10–14 s. Of course, this also begs the question whether to include such a time limit or not in future classroom trials. What if the time limit just heightens anxieties? The absence of a time limit in our study can be seen as a limitation.

Another reason could be the relatively short duration of the intervention. Rodán et al. (2016) noted improvements in MRS in 14–15-year olds as well, but also no gender differences. The programme was about 3 h in length and had no time limits. Nevertheless, the short duration of the intervention can be seen as a limitation in our study. Finally, it must be noted that the lack of a gender effect could have been caused by the disbalance in the conditions. Although we aimed to have balanced conditions with initially 85 students using the intervention, the messy nature of classroom practices meant that the groups were not balanced, something which is a limitation of the study. We also note again that the study took place at one independent school in the south of England. Future studies should take place in other school contexts.

Even though we could not confirm gender differences in our study, we could ask ourselves whether we even should want to close the gender gap in MRS. Uttal et al. (2013) argued that efforts that focus on closing the gender gap of specific spatial skills, such as mental rotation, may be misplaced. From a theoretical perspective differences in performance on isolated spatial skills might be interesting; however, on a practical level such differences might only be relevant if they have undesirable consequences. And, according to Uttal et al. (2013), if we focus on such undesirable outcomes, for example decreasing the gender gap in measures of STEM success (i.e. grades and achievement in STEM disciplines), then maybe there are better ways to address these than focussing on a 'niche' spatial skill like MRS. Nazareth et al.'s (2013) results suggest that the relation between gender and MRS is partially mediated by the number of masculine spatial activities participants had engaged in as youth. Closing the gap between males and females in spatial ability may then be accomplished in part by 'encouraging female youth to engage in more particular kinds of spatial activities' (p. 201).

5.2. Why should mental rotation skills even matter?

If we focus on the duration of the intervention, an even more relevant theme for discussion again would be the 'added value' of MRS in the mathematics curriculum. Given the limited specific time devoted in the current curricula to 3D geometry, it might not be realistic to expect more time being devoted to custom interventions for mental rotation. If we classify these tasks as non-standard or non-normative, the greater cognitive complexity of tasks could require more study-time (Widder, 2018). For two reasons we think it still would be useful to include limited training in the mathematics curriculum. Firstly, because students and teachers are generally very positive and motivated about the specific building app; it simply is a lot of fun and has a geometry focus, a mathematical topic that often buckles under the pressures of algebra, statistics and probability. Secondly, because although our study did not provide overwhelming evidence of improving MRS, with a limited time investment accuracy improved. Having said that, a priori we expected the effect to be more sizeable. Where Hawes et al. (2015) found that spatial training led to near transfer on other spatial tasks, our study did not show such gains. Possibly the lack of effect on mathematics had to do with the short duration of the intervention. The tool could be placed in a more extensive programme of spatial training, including other spatial skills than mental rotation. It must be said, however, that recent curriculum changes have resulted in less rather than more time for spatial skills. Perhaps more comprehensive interventions are better in taking advantage of the historically tight relationship between spatial thinking and mathematics. An absence of training effects also could lie in the fact that according to a meta-analysis of twin studies spatial visualisation abilities are largely heritable, with shared environmental factors having little effect (King et al., 2019).

5.3. Mechanisms between mental rotation skills and mathematical reasoning

Maybe our results can also be explained by looking in more detail at the underlying cognitive mechanism at play in the relationship between spatial abilities and mathematical reasoning. Hawes and Ansari (2020) came to four specific mechanisms or accounts for the relationship. For example, the shared neural processing account points to similar areas of the brain being used for higher-level spatial skills like mental rotation (e.g. see Hawes et al., 2019b; Figure 4). Zacks (2008) meta-analysis on the neural correlates of mental rotation also points to this, with particular parts of the brain robustly and consistently activated during mental rotation tasks. Another mechanism formulates a 'spatial modelling account', which involves a 'mental blackboard' on which numerical relations and operations can be modelled and visualised. Previously we suggested that perhaps the strategies might differ between males and females, and the strategies might be related to that 'mental blackboard'. However, we must remember that the tool in question also provides a very concrete 'manipulation' students can use to visualise their answers.

Anecdotally, the first author can report that several students would regularly turn the 3D building to the three perspectives to check if they were correct. Hawes and Ansari (2020) mention another mechanism that might be involved in spatial performance, namely working memory. They mention the example of mentally rotating cube figures and describe how 'individuals with low-spatial abilities often lose 'sight' of the mental image and require multiple attempts at rotation' (p. 474). It is asserted, in this explanation, that individual differences in spatial visualisation are related to differences in working memory (e.g. Hegarty & Waller, 2005). Although we did not include our measure of mental effort in our analyses because of missing data, the average approached 7 out 9, with females slightly higher than males. We realise this is inconclusive but such a high level of self-reported mental effort might indicate an important role for working memory capacity. All these findings are not easy to interpret, but regardless of the effectiveness, we also draw optimism from the fact that teachers and students very much enjoyed the intervention tool, and in that sense even some moderate promise, especially for accuracy, of our limited intervention could be enough for teachers to use it in their classrooms. This also is testament to the high quality of the tool's design (Boon, 2009).

5.4. Improving some design choices

We can't rule out that our design choices in the tool influenced the outcomes, which we deem a limitation of the study. In a future trial it might be good to have two stages, rather than let students decide: first create *a* correct building for all 20 tasks, *after that* aim for the minimum number for each of the buildings by removing cubes. Sometimes, limitations and constraints in a tool can actually be conducive for learning. For example, Montag et al. (2021) found that limited rotation perhaps would be more effective than free rotation. In the context of our study, the training tool allows for freedom in the way the houses are manipulated. This again shows the design tensions in mathematical tools and tasks between constraints and open-ness (e.g. Bokhove & Jones, 2018; Mackrell & Bokhove, 2017). Another interesting question for future research would be to look *within* the activities how students went about (strategies) solving the task. For instance, some students added cubes, while others started with a rough building with a lot of cubes and then started removing cubes.

6. Conclusion

Although the 'messy' practices of research in classrooms made this study challenging, we fervently agree with Stylianides and Stylianides (2013) that mathematics education research should be conducted in classrooms through collaboration of teachers and researchers to ensure relevance and practicality. The enthusiasm of students and teachers meant the study with its intervention tool was relevant and practical. Our study, despite some limitations, provides empirical evidence that with a low-intensity intervention *accuracy* could be improved. Furthermore, mental rotation skills are difficult to learn and difficult to teach directly. Having a tool that can be used anyplace at any time, with a playful puzzle element, could be a useful way to both build rotation skills, as well as enjoyment.

Note

1. This could be caused by two things: (1) students not doing the work seriously and (2) incorrect recording of the times through the browser.

Acknowledgments

We thank the teachers and students from the participating school.

Disclosure statement

No potential conflict of interest was reported by the authors.

Author contribution statement

C.B. and E.R. collaboratively conceptualised the research, analysed the data and wrote the manuscript. C.B. collected the data in the field experiment.

ORCID

Christian Bokhove http://orcid.org/0000-0002-4860-8723
Edward S. Redhead http://orcid.org/0000-0002-7771-1228

References

Albarracín, L., Hernàndez-Sabaté, A., & Gorgorió, N. (2022). Designing levels of a video game to promote spatial thinking. *International Journal of Mathematical Education in Science and Technology*, *53*(11), 3138–3150. https://doi.org/10.1080/0020739X.2021.1953626

Battista, M., & Clements, D. H. (1998). Research into practice: Finding the number of cubes in rectangular cube buildings. *Teaching Children Mathematics*, *4*(5), 258–264. https://doi.org/10.5951/TCM.4.5.0258

Battista, M. T., & Clements, D. H. (1996). Students' understanding of three-dimensional rectangular arrays of cubes. *Journal for Research in Mathematics Education*, *27*(3), 258–292. https://doi.org/10.2307/749365

Ben-Chaim, D., Lappan, G., & Houang, R. T. (1988). The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal*, *25*(1), 51–71. https://doi.org/10.3102/00028312025001051

Ben-Haim, D., Lappan, G., & Houang, R. T. (1985). Visualizing rectangular solids made of small cubes: Analyzing and effecting students' performance. *Educational Studies in Mathematics*, *16*(4), 389–409. https://doi.org/10.1007/BF00417194

Bokhove, C., & Jones, K. (2018). Stimulating mathematical creativity through constraints in problem-solving. In N. Amado, S. Carreira, & K. Jones (Eds.), *Broadening the scope of research on mathematical problem solving* (pp. 301–319). Springer. https://doi.org/10.1007/978-3-319-99861-9_13

Bokhove, C., & Redhead, E. (2017). Training mental rotation skills to improve spatial ability. *Online Proceedings of the BSRLM*, *36*(3). https://eprints.soton.ac.uk/407203/1/BSRLM_CP_36_3_02.pdf

Boon, P. (2006). Designing didactical tools and microworlds for mathematics educations. In C. Hoyles, J.-B. Lagrange, L. H. Son, & N. Sinclair (Eds.), *Proceedings of the 17th ICMI Study Conference* (pp. 38–45). https://www.mathunion.org/fileadmin/ICMI/files/Digital_Library/icmi-study-17/ICMI17proceedingsPart2.pdf

Boon, P. (2009). A designer speaks: Designing educational software for 3D geometry. *Educational Designer*, *1*(2), 1–11. https://www.educationaldesigner.org/ed/volume1/issue2/article7/

Bruce, C. D., & Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: What is it? Why does it matter? And what can we do about it? *ZDM*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Casey, M. B., Pezaris, E., & Nuttall, R. L. (1992). Spatial ability as a predictor of math achievement - The importance of sex and handedness patterns. *Neuropsychologia*, *30*(1), 35–45. https://doi.org/10.1016/0028-3932(92)90012-B

Cheng, Y.-L., & Mix, K. M. (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

Cheung, C. N., Sung, J. Y., & Lourenco, S. F. (2019). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, *84*(7), 2000–2017. https://doi.org/10.1007/s00426-019-01202-5

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Cornu, V., Schiltz, C., Pazouki, T., & Martin, R. (2017). Training early visuo-spatial abilities: A controlled classroom-based intervention study. *Applied Developmental Science*, *23*(1), 1–21. https://doi.org/10.1080/10888691.2016.1276835

Cumino, C., Pavignano, M., Spreafico, M. L., & Zich, U. (2021). Geometry to build models, models to visualize geometry. *Digital Experiences in Mathematics Education*, *7*(1), 149–166. https://doi.org/10.1007/s40751-020-00080-6

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

Department for Education. (2013). *National curriculum in England: Mathematics programmes of study*. https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study

Dickson, B. A., Kotsopoulos, D., & Harris, L. (2022). The use of coding clubs to develop middle-school students' spatial reasoning abilities. *Digital Experiences in Mathematics Education*, *8*(1), 50–69. https://doi.org/10.1007/s40751-022-00099-x

Dilling, F., & Vogler, A. (2021). Fostering spatial ability through computer-aided design: A case study. *Digital Experiences in Mathematics Education*, *7*(2), 323–336. https://doi.org/10.1007/s40751-021-00084-w

Dumontheil, I., & Klingberg, T. (2011). Brain activity during visuospatial working memory task predicts arithmetical performance 2 years later. *Celebral Cortex*, *22*(5), 1078–1085. https://doi.org/10.1093/cercor/bhr175

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850–855. https://doi.org/10.1111/j.1467-9280.2007.01990.x

Fenner, J., Heathcote, D., & Jerrams-Smith, J. (2000). The development of wayfinding competency: Asymmetrical effects of visuo-spatial and verbal ability. *Journal of Environmental Psychology*, *20*(2), 165–175. https://doi.org/10.1006/jevp.1999.0162

Fowler, S., Cutting, C., Kennedy, J., Leonard, S. N., Gabriel, F., & Jaeschke, W. (2021). Technology enhanced learning environments and the potential for enhancing spatial reasoning: A mixed methods study. *Mathematics Education Research Journal*, *34*(4), 887–910. https://doi.org/10.1007/s13394-021-00368-9

Francis, K., Khan, S., & Davis, B. (2016). Enactivism, spatial reasoning and coding. *Digital Experiences in Mathematics Education*, *2*(1), 1–20. https://doi.org/10.1007/s40751-015-0010-4

Fujita, T., Kondo, Y., Kumakura, H., Kunimune, S., & Jones, K. (2020). Spatial reasoning skills about 2D representations of 3D geometrical shapes in grades 4 to 9. *Mathematics Education Research Journal*, *32*(2), 235–255. https://doi.org/10.1007/s13394-020-00335-w

Ganis, G., & Kievit, R. A. (2015). A new set of three-dimensional shapes for investigating mental rotation processes: Validation data and stimulus set. *Journal of Open Psychology Data*, *3*(1), Article e3. https://doi.org/10.5334/jopd.ai

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

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

Harris, D., Logan, T., & Lowrie, T. (2021). Unpacking mathematical-spatial relations: Problem-solving in static and interactive tasks. *Mathematics Education Research Journal*, *33*(3), 495–511. https://doi.org/10.1007/s13394-020-00316-z

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior. *Psychonomic Bulletin & Review*, *27*(3), 465–482. https://doi.org/10.3758/s13423-019-01694-7

Hawes, Z. C. K., Gilligan-Lee, K. A., & Mix, K. S. (2022). Effects of spatial training on mathematics performance: A meta-analysis. *Developmental Psychology*, *58*(1), 112–137. https://doi.org/10.1037/dev0001281

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., Seo, J., & Ansari, D. (2019a). 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

Hawes, Z., Sokolowski, H. M., Ononye, C. B., & Ansari, D. (2019b). Neural underpinnings of numerical and spatial cognition: An fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation. *Neuroscience & Biobehavioral Reviews*, *103*, 316–336. https://doi.org/10.1016/j.neubiorev.2019.05.007

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

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121–169). Cambridge University Press. https://doi.org/10.1017/CBO9780511610448.005

Hoyek, N., Collet, C., Fargier, P., & Guillot, A. (2012). The use of the Vandenberg and Kuse mental rotation test in children. *Journal of Individual Differences*, *33*(1), 62–67. https://doi.org/10.1027/1614-0001/a000063

İbili, E., Çat, M., Resnyansky, D., Şahin, S., & Billinghurst, M. (2020). An assessment of geometry teaching supported with augmented reality teaching materials to enhance students' 3D geometry thinking skills. *International Journal of Mathematical Education in Science and Technology*, *51*(2), 224–246. https://doi.org/10.1080/0020739X.2019.1583382

Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., & Heil, M. (2013). Mental rotation performance in primary school age children: Are there gender differences in chronometric tests? *Cognitive Development*, *28*(1), 51–62. https://doi.org/10.1016/j.cogdev.2012.08.005

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137–172. https://doi.org/10.1037/0033-295X.92.2.137

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. https://doi.org/10.1016/j.intell.2006.07.009

King, M. J., Katz, D. P., Thompson, L. A., & Macnamara, B. N. (2019). Genetic and environmental influences on spatial reasoning: A meta-analysis of twin studies. *Intelligence*, *73*, 65–77. https://doi.org/10.1016/j.intell.2019.01.001

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, *23*(1), 77–94. https://doi.org/10.1007/BF03173141

Lauer, J. E., & Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychological Science*, *27*(10), 1291–1298. https://doi.org/10.1177/0956797616655977

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

Loring-Meier, S., & Halpern, D. F. (1999). Sex differences in visuospatial working memory: Components of cognitive processing. *Psychonomic Bulletin & Review*, *6*(3), 464–471. https://doi.org/10.3758/BF03210836

Lowrie, T., Resnick, I., Harris, D., & Logan, T. (2020). In search of the mechanisms that enable transfer from spatial reasoning to mathematics understanding. *Mathematics Education Research Journal*, *32*(2), 175–188. https://doi.org/10.1007/s13394-020-00336-9

Mackrell, K., & Bokhove, C. (2017). Designing technology that enables task design. In A. Leung & A. Baccaglini-Frank (Eds.), *Digital technologies in designing mathematics education tasks* (pp. 55–73). Springer. https://doi.org/10.1007/978-3-319-43423-0_4

Meneghetti, C., Cardillo, R., Mammarella, I. C., Caviola, S., & Borella, E. (2017). The role of practice and strategy in mental rotation training: Transfer and maintenance effects. *Psychological Research*, *81*(2), 415–431. https://doi.org/10.1007/s00426-016-0749-2

Merrill, E. C., Yang, Y., Roskos, B., & Steele, S. (2016). Sex differences in using spatial and verbal abilities influence route learning performance in a virtual environment: A comparison of 6-to 12-year old boys and girls. *Frontiers in Psychology: Developmental Psychology*, *7*, Article 258. https://doi.org/10.3389/fpsyg.2016.00258

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior*, *42*, 197–243. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D., & Bower, C. (2021). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology*, *113*(2), 304–314. https://doi.org/10.1037/edu0000494

Montag, M., Bertel, S., de Koning, B. B., & Zander, S. (2021). Exploration vs. limitation–An investigation of instructional design techniques for spatial ability training on mobile devices. *Computers in Human Behavior*, *118*, Article 106678. https://doi.org/10.1016/j.chb.2020.106678

Nazareth, A., Herrera, A., & Pruden, S. M. (2013). Explaining sex differences in mental rotation: Role of spatial activity experience. *Cognitive Processing*, *14*(2), 201–204. https://doi.org/10.1007/s10339-013-0542-8

Nori, R., Grandicelli, S., & Giusberti, F. (2006). Visuo-spatial ability and wayfinding performance in real-world. *Cognitive Processing*, *7*(1), S135–S137. https://doi.org/10.1007/s10339-006-0104-4

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Parsons, T. D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, G., & Rizzo, A. A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropscyhologia*, *42*(4), 555–562. https://doi.org/10.1016/j.neuropsychologia.2003.08.014

Patkin, D., & Dayan, E. (2013). The intelligence of observation: Improving high school students' spatial ability by means of intervention unit. *International Journal of Mathematical Education in Science and Technology*, *44*(2), 179–195. https://doi.org/10.1080/0020739X.2012.703335

Rahe, M., & Quaiser-Pohl, C. (2021). Can (perceived) mental-rotation performance mediate gender differences in math anxiety in adolescents and young adults? *Mathematics Education Research Journal*, *35*, 255–279. https://doi.org/10.1007/s13394-021-00387-6

Reinhold, F., Hofer, S., Berkowitz, M., Strohmaier, A., Scheuerer, S., Loch, F., Vogel-Heuser, B., & Reiss, K. (2020). The role of spatial, verbal, numerical, and general reasoning abilities in complex word problem solving for young female and male adults. *Mathematics Education Research Journal*, *32*(2), 189–211. https://doi.org/10.1007/s13394-020-00331-0

Resnick, I., Harris, D., Logan, T., & Lowrie, T. (2020). The relation between mathematics achievement and spatial reasoning. *Mathematics Education Research Journal*, *32*(2), 171–174. https://doi.org/10.1007/s13394-020-00338-7

Rodán, A., Contreras, M. J., Elosúa, M. R., & Gimeno, P. (2016). Experimental but not sex differences of a mental rotation training program on adolescents. *Frontiers in Psychology*, *7*, Article 1050. https://doi.org/10.3389/fpsyg.2016.01050

Safadel, P., White, D., & Kia, A. (2023). Spatial self-efficacy and spatial ability: An analysis of their relationship. *New Review of Hypermedia and Multimedia*, *29*(2), 114–150. https://doi.org/10.1080/13614568.2023.2248057

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604–614. https://doi.org/10.1037/0022-0663.93.3.604

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Sorby, S. (2009). Educational research in developing 3-D spatial kkills for engineering students. *International Journal of Science Education*, *31*(3), 459–480. https://doi.org/10.1080/09500690802595839

Stylianides, A. J., & Stylianides, G. J. (2013). Seeking research-grounded solutions to problems of practice: Classroom-based interventions in mathematics education. *ZDM—The International Journal on Mathematics Education*, *45*(3), 333–340. https://doi.org/10.1007/s11858-013-0501-y

Tam, Y. P., Wong, T. T. Y., & Chan, W. W. L. (2019). The relation between spatial skills and mathematical abilities: The mediating role of mental number line representation. *Contemporary Educational Psychology*, *56*, 14–24. https://doi.org/10.1016/j.cedpsych.2018.10.007

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, *22*(7), 996–1013. https://doi.org/10.1002/acp.1420

Titze, C., Jansen, P., & Heil, M. (2010). Mental rotation performance and the effect of gender in fourth graders and adults. *European Journal of Developmental Psychology*, *7*(4), 432–444. https://doi.org/10.1080/17405620802548214

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

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

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 521–525). Springer. https://doi.org/10.1007/978-94-007-4978-8_170

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *24*(2), 307–334. https://doi.org/10.3758/s13423-016-1085-7

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology*, *82*(1), 157–181. https://doi.org/10.1111/j.2044-8279.2011.02049.x

Widder, M. (2018). What does study-time reveal about perceptual visual difficulty? Debating two theories about study-time allocation in spatial geometry tasks. *Educational Alternatives*, *16*, 132–141. https://www.scientific-publications.net/en/article/1001791/.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*(4), 763–771. https://doi.org/10.3758/PBR.15.4.763

Xie, F., Zhang, L., Chen, X., & Xin, Z. (2019). Is spatial ability related to mathematical ability: A meta-analysis. *Educational Psychology Review*, *32*(1), 113–155. https://doi.org/10.1007/s10648-019-09496-y

Young, C. (2016). Wisweb applets. [Blog post]. https://colleenyoung.wordpress.com/2016/10/29/wisweb-applets-4/.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19. https://doi.org/10.1162/jocn.2008.20013