Exploring the Relationship of Spatial Visualization and Scientific Modeling in Grades 4 and 7 Students Based on Cognitive Assessment Data

Jing Lin | ORCID: 0000-0003-3721-710X
Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, 100875, China
linjing@bnu.edu.cn; wzlinjing@163.com

Letong Zhang
Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, 100875, China
17860717936@163.com

Knut Neumann | ORCID: 0000-0002-4391-7308
Department of Physics Education, IPN – Leibniz Institute for Science and Mathematics Education, 24118 Kiel, Germany
neumann@leibniz-ipn.de

Ping-Han Cheng
Department of Science Education, National Taipei University of Education, Taipei City 106, Taiwan
hcheng@mail.ntue.edu.tw

Wenting Wei
Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, 100875, China
578948472@qq.com

Chun-Yen Chang | ORCID: 0000-0003-2372-2004
Science Education Center, Graduate Institute of Science Education and the Department of Earth Sciences, National Taiwan Normal University Taipei City 116, Taiwan
chunyenc@gmail.com

Received: 29 June 2022 | Revised: 23 October 2022 | Accepted: 6 November 2022
Abstract

Scientific modeling (SM) is a core practice of science and an important component of scientific literacy. Supporting students in developing the competence to construct, use, evaluate, and revise models is hence of particular relevance. While research has shown that spatial visualization (SV), a core component of spatial ability, is correlated with students’ SM performance, it is unclear which role SV plays in students performing the four elements of SM: the construction, the use, the evaluation, and the revision of models. This study analyzes the role of SV in the performance of a series of SM tasks by 279 students in Grades 4 and 7. The findings indicate that SV affects students’ performance in SM but that the effect is more significant in Grade 4 than in Grade 7. More specifically, SV is significantly predictive for model revision in Grade 4, but significant for model evaluation in Grade 7. However, there was no gender difference in the effect of SV on SM. The implications are that science teaching and learning must better attend to supporting younger students through visual aids when engaging them in SM. The study also suggests that further studies are needed to understand the different cognitive processes involved in students’ SM and their complex interplay.

Keywords

scientific modeling – spatial visualization – mental model – interactive computer assessment
Introduction

Scientific modeling (SM) is a scientific practice that is core to the development of scientific literacy (NGSS Lead States, 2013). In fact, some researchers view science as an iterative process of constructing, using, evaluating, and refining models with the purpose of developing explanations to macroscopic and microscopic phenomena and predicting their future behavior (e.g., Gilbert & Justi, 2016; Harrison & Treagust, 1996). As a consequence, it is believed that engaging students in SM will help them develop a deeper understanding of scientific concepts, principles, and laws and develop positive attitudes towards and values related to science (NGSS Lead States, 2013). Studies have also shown that SM is rather challenging for students (Schwarz et al., 2009). The national science education quality monitoring in China from 2017 revealed, for example, a poor performance of Grades 4 and 8 students with respect to the construction and use of scientific models for explaining phenomena and solving problems (MOE, 2018). To further facilitate science education reform moving from examination-centered to literacy-oriented science education, the recent update of the Chinese science standards highlight SM as a core practice of science education (MOE, 2022). Teaching students SM, however, poses a huge challenge for primary and middle school science teachers as it is a completely new requirement for teachers in the standards.

The practice of SM broadly refers to the development and use of models to provide explanatory accounts of scientific phenomena (National Research Council, 2012). The complex and higher-order cognitive process of SM involves four elements: the construction, use, evaluation, and refinement of a model (Schwarz et al., 2009). The construction of a model begins with the creation of a mental model of the phenomenon in question (Hestenes, 1995). Spatial visualization (SV) ability is considered to be involved in the establishment of mental models (Gobert, 2005). Students with high levels of SV ability were found to be able to construct quality mental models, while students with low level of SV ability could only form lower quality ones (Wang & Barrow, 2011; Briggs & Bodner, 2005). However, despite ample evidence of the role that SV plays in the creation of mental models, the role that SV plays in the process of SM is largely unclear to date.

This study uses data from a field test in which students respond to a series of SM tasks requiring the construction, use, evaluation, and refinement of a model and a test assessing students' SV ability. SM tasks were interactive, computer-based tasks using visual representations to test students' ability to construct, use, evaluate, and refine models, that is, to assess students' SM abilities. Students' SV abilities were assessed as a general cognitive ability using a specifically designed instrument: the paper-folding test (Ekstrom et al., 1976).
Using hierarchical regression analysis, we extend prior approaches investigating correlative patterns between SM and other constructs with the purpose of facilitating the teaching and learning of the practice of SM in China and other countries.

2 Theoretical Background

2.1 Mental Models in Scientific Modeling

The practice of scientific modeling (SM) involves a range of knowledge, abilities, and skills (Schwarz & White, 2005). Amongst different delineations of the practice of SM (e.g., Hestenes, 1987; Halloun, 1996; Chang & Chiu, 2009; Justi & Gilbert, 2002), the one proposed by Schwarz et al. (2009) has become the one most widely employed. According to Schwarz et al. (2009), SM involves four elements: the construction, use, evaluation, and revision of models, each combining a range of cognitive processes. The creation of a mental model from reality is the foundation of the SM process (Hestenes, 1995). A mental model usually includes two components: a visual-pictorial one and a propositional one (Greca & Moreira, 2002; Mayer, 2001). In theory, each element in the process of SM requires coordinating these two components (NRC, 2012).

In the study by Pierson et al. (2017), for example, eighth grade students are asked to create mechanistic models to explain and predict the ecosystem of their school garden. First, students construct diagrammatic models about the phenomenon they observed involving the soil in the garden being moister in some places than in others. Students then gradually use, evaluate, and revise models from their literal observation. Some students conceptualize the function of the plant roots as a water pump (i.e., the propositional component of mental model), and further create physical microcosms and computational models by visualizing the interaction of the roots and the environment (i.e., the visual-pictorial one). Some students pay attention to the different leaves of the plants. They note transpiration (i.e., the propositional one) and draw different soil moisture around different plants with branches and leaves (i.e., the visual-pictorial one). Some students revise their models considering the changes of the air humidity in the garden ecosystem for a more accurate prediction of plant growth. By framing three model types (i.e., diagrammatic, physical, and computational) students are induced to continually process and refine their mental models to create mechanistic rather than descriptive models.

Taking the study of Pierson et al. (2017) as an example, it can be seen that the essence of students’ SM practice is to create abstract and generative models that can explain or predict the relevant phenomena. The process of
constructing, using, evaluating, and revising models is the process in which students continuously process and improve their internal mental models according to external phenomena or representations (Gobert, 2005; Chang, 2007). It has been pointed out that SV guides students in visualizing the structure and features of models in their minds to create the mental models (Fiorella et al., 2003; Barak & Hussein-Farraj, 2013) with the visual-pictorial and propositional components, especially in visualizing those that cannot be seen or that are difficult to explain in the classroom (Hoffler & Leutner, 2007; Chittleborough & Treagust, 2008). It has also been argued that students’ difficulty in SM is mainly due to the difficulty in visually transforming relevant phenomena into mental models (Al-Balushi, 2013; Sins et al., 2005). Hence, SV plays a role in the internal cognitive processing process of the visualization and is thus involved in the four elements of SM.

2.2 Scientific Modeling and Spatial Visualization

Spatial visualization (SV) has received increasing attention in science education as an ability relevant to addressing many of the issues of the digital age (Krell et al., 2015; Kell et al., 2013). As a core component of spatial ability (McGee, 1979; Morris, 2018; Mix et al., 2021), SV is defined as an ability to manipulate, rotate, and twist two- or three-dimensional objects in a multi-step psychological process and to transform representations of spatial arrangements of objects into representations of other arrangements (Linn & Petersen, 1985; Ekstrom et al., 1976).

In science education, SV is regarded as the meta-cognitive ability to understand three-dimensional phenomena from two-dimensional representations of them (Barnea, 2000; Gilbert, 2005). Studies have shown that SV plays a central role in students’ efforts to create scientific models. In the context of biology learning, the ability to derive information about the spatial configuration of phenomena is found to be linked to the depth of students’ conceptual understanding of the models such as the double helical structure of DNA (Al-Balushi, 2013). In the teaching and learning of physics, students with high levels of SV ability are observed to work with graphs or visual representations of physical models more readily in problem solving; for example, in coordinating system changes and performing vector transformations in solving motion problems (Kozhevnikov et al., 2007). In geology learning, students with high levels of SV ability tend to give more “penetrative answers” in a geologic test (Kali & Orion, 1996), which means that they are able to grasp the deeper structure of geologic phenomena beyond superficial features. In interdisciplinary STEM education, SV has also been found to be connected to students’ understanding of models representing related phenomena (e.g., Sorby et al., 2018; Stieff & Uttal, 2015).
However, despite ample evidence of a positive association between SV and SM, the role that SV plays in the process of SM, in particular with respect to the four elements, is unclear to date.

With the point of view that SV engages in creating the mental model in the process of SM, this study connects SV and SM to test the relationship between them. The assumed role of SV in the process of SM, especially with respect to the four elements, is: (i) When students are asked to construct models about a phenomenon, SV should help students visualize the structure and characteristics they already have about the phenomenon in their minds to construct initial mental models; (ii) during the model use process, SV should help students convert symbolic representations into 3D structures so that they can use their initial mental models to explain or predict relevant phenomena; (iii) further, SV should help students visually match their initial mental models with the structure and characteristics of more relevant phenomena or alternative models, which can also include scientific models, to evaluate the interpretable or predictable function of their mental models; and, (iv) finally, SV should help students to refine their mental models to explain or predict relevant phenomena more accurately and scientifically, in other words, to create mechanistic models. Thus, SV is critical to engaging in the four elements of SM by transforming between symbolic representations and 3D structures to iteratively refine mental models.

2.3 Individual Differences in Scientific Modeling and Spatial Visualization

Existing studies have indicated that both SM and SV develop with the individual maturity. For SV, Piaget (1954) believed that children can develop spatial-related abilities until school age (about 6 years old). More recent psychological research has shown that individuals already have visual-spatial awareness of objects at 6 months of infancy (Lauer et al., 2015). At the age of 1 to 2 years old, children’s visual-spatial awareness develops rapidly and becomes more sophisticated (Lauer & Lourencó, 2016). The longitudinal study of Geer et al. (2019) found that older students scored higher in SV tasks. Grade 2 students exhibited a significant improved performance over Grade 1 students. Students from Grade 3 also showed an improvement that was smaller. Morris (2018) suggested that the SV performance of students from Grades 2 to 8 increased with each grade and that there were significant differences between these three groups: Grades 2 and 3, Grades 4 and 5, and Grades 6–8. Ogunkola and Knight (2019) argued that SV’s development gradually slows down after the age of 14, although the development for some individuals can last until after the age of 20. Most current studies have supported the conclusion that SV ability continues to develop as the individual matures. But the stagnation and
rapid development stages in SV ability have not yet been made explicit. For SM, in the perspective of learning progression, older students have commonly been classified as higher level (Schwarz et al., 2009, 2012; Bambeger & Davis, 2013). There has also been evidence that students from Grades 5 and 6 cannot reach the higher level of SM that students from Grade 8 can demonstrate (Pierson et al., 2017), and high school students can exhibit a level of performance that is not achieved by K-9 students (Fortus et al., 2016).

The gender difference in SV has always been one of the important foci for research. Some early studies found no significant gender difference in SV (e.g., Voyer et al., 1995; Linn & Petersen, 1985), whereas others found that males frequently scored higher than females in SV (e.g., Macoby & Jacklin, 1974). Battista’s (1990) research on high school students found that boys had significantly higher SV ability than girls. However, Caplan et al. (1985, 1986) believed that there was currently no evidence to prove gender differences in SV from a biological perspective. Many other empirical studies for the K-12 students have also shown that the difference in SV between boys and girls was not significant (e.g., Tartre, 1990; Manger & Eikeland, 1996; Capraro, 2001). In subsequent studies on SV ability, a lack of discernible difference between genders has gradually become mainstream. Morris’s (2018) study of students from Grades 2 to 8 found no significant difference in SV test scores between boys and girls in most age groups. Only in Grades 4 and 5, girls’ performance in the visuospatial working memory task scores was significantly higher than boys’. Morris believed that this may be related to the fact that girls are the first to enter puberty at this stage, and their brains develop rapidly. Ogunkola and Knigt’s (2019) research on 420 middle school students aged 13–15 also showed no significant difference in the performance of male and female students in SV. As it can be seen here, the above different results may occur in specific age groups. Therefore, the issue of gender differences in SV is still worth exploring – especially since no gender difference has been found in students’ SM performance in previous studies. Only recently one study suggested girls construct higher quality models and demonstrate better learning gains with respect to SM (Mierdel & Bogner, 2019). This study will hence further examine gender difference in SV and its role in SM. More specifically, this study will examine whether the effect of SV on SM differs by gender.

3 Research Questions

Previous research has suggested that SV plays an essential role in SM. However, there has been no evidence about the extent to which SV is correlated with SM, which is an important requisite for developing students’ SM practice. In
addition, since there have been shown to be age and gender differences in SV and age differences in SM, the relationship between SV and SM may also vary by age or gender. The present study aims to address this gap in the current literature. The respective research questions are:

1. To which extent does SV affect students’ performance in constructing, using, evaluating, and revising models?
2. How does the effect of SV on students’ SM practice depend on student characteristics such as gender and age?

4 Methods

In order to address the research questions, a set of computer-based interactive tasks were designed to assess students’ performance in the four-element process of SM practice. These computer-based interactive tasks were administered to students from Grades 4 and 7 in primary and middle schools, together with a paper-folding test to assess students’ SV abilities. The students took the tests in their classrooms using computers to individually respond to the tests. The collected data were analyzed using hierarchical linear regression modeling to account for different types of variables.

4.1 Measurement Instruments

4.1.1 Interactive Computer Assessment of Scientific Modeling

Scientific modeling has been investigated through interviews (e.g., Grosslight et al., 1991), questionnaires (e.g., Lin & Chiu, 2008), and observations (e.g., Pierson et al., 2017), including students’ meta-knowledge about SM (e.g., Fortus et al., 2016) and practice of SM (e.g., Schwarz et al., 2012; Pierson et al., 2017). However, none of the instruments used has specifically assessed primary and middle school students’ performance in SM as a four-element process. Hence, we designed a new instrument to assess the four-element process of SM.

Research has shown that interactive computer assessment (ICA) can provide students with an immersive environment and allow students to fully demonstrate their cognitive abilities (Dede, 2009). ICA is not supposed to simply digitize the paper-and-pencil test items nor take the place of performance-based assessments, but to take advantage of human computer interactive (HCI) to effectively assess students’ higher-order thinking and complex scientific practices (Kuo et al., 2019; Pritchett & Sakrzewski, 2000). Hence, we authored interactive computer-based tasks as part of our newly designed instrument to assess SM.

Situated in the context of the solar system, each task assessed one element of SM (see Table 1). In the national science standards of China, the solar system
is a standard topic to be covered. Students are expected to learn about the relative position and motion of the Earth, the Sun, and the Moon in primary school (MOE, 2017), and about these astronomical bodies’ relative motion in middle school (MOE, 2011). Hence, students in Grades 4 and 7 can safely be expected to have sufficient knowledge about the solar system to be able to respond to SM tasks.

### Table 1: The structure of the solar system interactive computer assessment

<table>
<thead>
<tr>
<th>Elements</th>
<th>Requirements</th>
<th>Item (score)</th>
<th>EFA rotated component matrixa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model construction</td>
<td>Constructing a 2D model to represent the given planets' positions in the solar system</td>
<td>1 (8)</td>
<td>.838</td>
</tr>
<tr>
<td></td>
<td>Constructing a 2D model to represent the given planets’ sizes in the solar system</td>
<td>2 (8)</td>
<td>.824</td>
</tr>
<tr>
<td>Model use</td>
<td>Explaining the brightness of each planet with its distance from the Sun according to your model</td>
<td>4 (8)</td>
<td>.852</td>
</tr>
<tr>
<td></td>
<td>Combining the brightness and distance information of each planet to explain the phenomena according to your model</td>
<td>6 (8)</td>
<td>.849</td>
</tr>
<tr>
<td></td>
<td>Explaining the brightness of each planet observed from the Earth according to your model</td>
<td>3 (8)</td>
<td>.647 .402</td>
</tr>
<tr>
<td>Model evaluation</td>
<td>Evaluating your own model based on the feedback towards your response in Item 4</td>
<td>5 (8)</td>
<td>.883</td>
</tr>
<tr>
<td></td>
<td>Further evaluating your model based on the information provided about an unknown planet</td>
<td>7 (8)</td>
<td>.849</td>
</tr>
<tr>
<td>Model revision</td>
<td>Revising your model based on the feedback towards your response in Item 9</td>
<td>10 (8)</td>
<td>.859</td>
</tr>
<tr>
<td></td>
<td>Reflecting on your responses in the previous steps to further revise your model</td>
<td>11 (9)</td>
<td>.829</td>
</tr>
<tr>
<td></td>
<td>Considering the size of different planets to revise your model</td>
<td>9 (8)</td>
<td>.804</td>
</tr>
<tr>
<td></td>
<td>Further considering the given different types of planets to revise your model</td>
<td>8 (8)</td>
<td>.616</td>
</tr>
</tbody>
</table>

a Note: The results of principal component analysis indicated four components were divided based on the eigenvalues (> 1). The four components explained 68.97% of the total variance. Normalized maximum variance method was used. Rotation converged after four iterations.
The tasks designed addressed students’ SM practices as follows: First, students were asked to create their own initial solar system model based on their prior knowledge in response to Tasks 1 and 2 (i.e., model construction). Second, students used their models to explain the given phenomena in response to Tasks 3, 4, and 6 (i.e., model use). Third, students evaluated their own models according to the new information shown in Tasks 5 and 7 (i.e., model evaluation). Fourth, students further revised their models according to the new information attached in Tasks 8, 9, 10, and 11 (i.e., model revision).

A range of techniques for “task analysis” (Crystal & Ellington, 2004) were employed to design and refine these tasks. Hierarchical task analysis (Fyiaz et al., 2018) was used to categorize tasks into four-element subtasks from the behavioral perspective. Activity analysis (Hashim & Jones, 2007) was used to organize the entire activity process to create the most economical and friendly context. Furthermore, the information of the planets involved in the tasks was given in the task text (see an example from Table 2) to control the interference of relevant knowledge on the test. The computer-human interaction was displayed in three ways: One was providing task information to guide students’ SM process, the second was providing timely feedback when students used their models, and the third was providing students a re-answer opportunity when they received feedback. The interface form was concise and clear. The requirement for students’ computer operation was simple: mainly clicking or dragging the mouse (see Figure 1), which reduced factors affecting ICA such as computer familiarity, computer anxiety, and computer hardware features (Bahar & Asil, 2018; Skryabin et al., 2015). By observing, recording, analyzing, and judging the users’ effectiveness, efficiency, and satisfaction in responding, three rounds of “usability test” (Albert et al., 2010) detected the tasks’ reliability and validity. As shown in Table 1, 11 interactive tasks examined the four elements of students’ SM, which were validated using exploratory factor analysis (KMO = .708; c² = 831.205, df = 55, p < .001). The results of subsequent confirmatory factor analysis (χ² = 61.444, df = 38, RMSEA = .047, CFI = .970, TLI = .957) showed a very good model fit (Hu & Bentler, 1999).

### Table 2 Information in Item 3

<table>
<thead>
<tr>
<th></th>
<th>Information in Item 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>There are eight planets in the solar system including the Earth.</td>
</tr>
<tr>
<td>2</td>
<td>The Earth is the third planet from the Sun.</td>
</tr>
<tr>
<td>3</td>
<td>The brightness of the planets observed from the Earth is: Venus &gt; Jupiter &gt; Mars</td>
</tr>
<tr>
<td></td>
<td>&gt; Mercury &gt; Saturn &gt; Uranus &gt; Neptune.</td>
</tr>
<tr>
<td>4</td>
<td>The farther away the planet is, the dimmer the planet looks.</td>
</tr>
</tbody>
</table>
4.1.2 The Paper-Folding Test of Spatial Visualization

In SV assessments, participants commonly must understand one spatial form or shape and match it with another, which involves complex and multi-step operations on spatial representation (Carroll, 1992; Linn & Petersen, 1985). SV assessments mainly include paper folding, embedded figure matching, and puzzle tasks. Amongst SV assessments the ones based on paper folding tasks are the most popular, including the mental folding test for children (MFTC; Harris et al., 2013; Kalayci, 2019), the mental paper-folding test (MPFT; Bennett et al., 1973; Morris, 2018), Dat5-SR (Bennett et al., 1973; Gómez-Tone et al., 2020), and the paper-folding test (PFT; Ekstrom et al., 1976; Lamb et al., 2015). The mental folding test for children and mental paper-folding test have roughly the same tasks, in which some colored arrows and dotted lines are used to indicate how the paper will be folded, and participants are then asked to select the folded paper from four options. Dat5-SR requires participants to imagine a folded three-dimensional image based on a flattened two-dimensional graph. The paper-folding test can be seen as a summary of all the above origami test requirements. It requires participants to imagine the folded paper and imagine the holes' distribution on the paper after the folded paper is punched, which involves more steps and more complicated spatial information operations.

Embedded figure matching assessments include the embedded figures test (Witkin, 1950) and the hidden figures task (Walter & Dassonville, 2011; Muenks et al., 2020). Participants are required to select a matching figure for a specific figure from the four options within the prescribed time frame. The puzzle task is a test tool suitable for children, adapted from the jigsaw-puzzle imagery task.
(Richardson & Vecchi, 2002) for adults by Casey et al. (2011). The assessment represents a comprehensive assessment of individual spv abilities, mental rotation abilities, and spatial working memory. The test items typically show some disarranged jigsaw pieces, among which some are rotated by 180 degrees. Participants are asked to imagine moving them to piece together a specific figure and fill in the corresponding box with each piece's number.

Given that participants for this study included both primary and middle school students, the relatively complex paper-folding test was selected for an optimal distinction of spv ability levels within the sample. The paper-folding test (PFT) employed in the study was developed by Ekstrom et al. (1976) and is still widely used today (e.g., Lamb et al., 2015; Sudatha et al., 2018). Twenty tasks were included. In these tasks, students were provided with a schematic diagram of a paper folded two to three times, punched with holes. Students were asked to identify the holes' position once the paper was fully unfolded. More specifically, students were supposed to choose one correct response from five different options (see Figure 2). Each correct response scored 1 point, a wrong response 0 points. The reliability of the test was good (Cronbach’s α = .810).

4.2 Participants

For this study, students from Grades 4 and 7 were sampled. Grade 4 students were 9 or 10 years old, generally entering the cognitive development period of concrete operation according to the cognitive development theory (Piaget, 1954). Grade 7 students were 12 or 13 years old, generally entering the formal operation period. The participants of this sample (N = 390) were students in Grades 4 (n = 191) and 7 (n = 199) from four primary schools and five middle schools in the Chinese provinces of Shandong, Shanxi, and Zhejiang. The sample was a convenience sample. A call for participants was sent out to 20 partner schools of the corresponding author's university, asking for voluntary
participants for a research study on ICA. Due to the highlighting requirements in terms of computer and network facilities and test time, the nine schools mentioned above finally participated. However, in some schools, the network was disconnected or the interface failed to load during the test, which resulted in some students being unable to complete the test. After excluding vacancies and repeated submissions, 279 valid responses were obtained, with a reliability of $\alpha = .713$. In the valid sample, 49.1% of the students were in Grade 4 ($n_4 = 137$), and 50.9% were in Grade 7 ($n_7 = 142$). The students identified themselves as 58.1% male ($n_m = 162$) and 41.9% female ($n_f = 117$).

4.3 Data Collection and Analysis

The SM ICA and the SV test were conducted online. The ICA tasks were presented through CCR,1 with a time limit of 40 minutes. The SV test was administered using So Jump Survey,2 which lasted 10 minutes. Students sat in their classrooms equipped with computers and responded to the tests individually. The information technology teachers administered the tests. Students finished SM ICA then answered SV items. Teachers opened each computer to present the task interface before the test. Students were asked to raise their hands when they finished the two tests and waited for the teachers to check their submissions before they left their seats.

In terms of descriptive statistics, mean scores and standard deviations were calculated. Student $t$-tests were conducted to analyze differences in gender and grades with respect to students’ SM and SV. For Research Question 1, hierarchical regression was used to identify the predictive function of SV, and path analysis was conducted to further explore the effect of SV on the four elements of SM. For Research Question 2, linear regression was conducted to analyze the relationship of SV and SM in both grades. Further path analysis was also conducted from the perspective of the four-element process of SM. All the data analysis was finished by SPSS 25.0.

5 Results

For the convenience of reporting the results more intensively and clearly for the two research questions, an overview of the participants’ performance on the

---

1 CCR is an online free interactive platform for teaching, learning, and assessment. Its full name is Cloud Classroom. Its link address is https://ccrr.cn.
2 So Jump Survey is an online platform for questionnaire survey, evaluation, and voting. The link address is https://www.wjx.cn.
SM interactive tasks and the SV paper-folding test is presented first. Therefore, there are three parts in the presentation of the results.

5.1 Students’ Performance on Spatial Visualization and Scientific Modeling

The mean scores and respective standard deviation (SD) for students’ SV and SM are shown in Table 3. A t-test suggested that there was a significant difference of SV and SM in mean score between Grade 4 and Grade 7 \((t = 5.762, p < .001; t = 4.366, p < .001)\). However, there were no gender differences in SV and SM. The percentage correct reflected the difficulty sequence of the four elements (see Figure 3). For both grades, the lowest percentage correct was found for model evaluation and the highest one was found for model use. It is important to note that other than for the other three elements, model evaluation tasks involved open-response format items. Requiring students to enter their own text may have increased the difficulty. All in all, model use tasks appeared to be relatively easy, whereas model evaluation and model revision tasks appeared to be relatively difficult for participants.

The results of the \(t\)-test on the gender differences in the four elements of SM revealed no significant differences \((t_{MC} = 1.088, p = .277; t_{MU} = 1.007, p = .315; t_{ME} = -.664, p = .507; t_{MR} = -1.790, p = .075)\). The results of the \(t\)-test on the grade differences in the four elements of SM showed that there were significant differences in model construction \((t = 6.559, p < .001)\), model use \((t = 2.742, p = .007)\), and model evaluation \((t = 5.276, p < .001)\), with Grade 7 students performing significantly better than Grade 4 students in the three elements. However, there was no significant difference in model revision between the two groups \((t = .235, p = .814)\).

TABLE 3 Descriptive statistics of SM and SV and \(t\)-test results

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th></th>
<th>SM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>(t)</td>
<td>(p)</td>
</tr>
<tr>
<td>Grade 4</td>
<td>7.49 (3.352)</td>
<td>5.762</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Grade 7</td>
<td>10.23 (4.495)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8.94 (4.164)</td>
<td>.247</td>
<td>.805</td>
</tr>
<tr>
<td>Female</td>
<td>8.81 (4.265)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Total score of SM ICA = 89; total score of PFT = 20.
5.2 Predictive Function of Spatial Visualization on Scientific Modeling

Hierarchical regression analysis was used to investigate the predictive function of SV on SM. For this purpose, the test scores were transformed into Z scores first. Gender and grade were transformed into dummy codes (0 for female or Grade 4, 1 for male or Grade 7) and were entered into Model 1 as control variables (i.e., covariates). A Z score of SV was entered into Model 2. The hierarchical model was convincing because the residuals were independent (Durbin-Watson value = 1.847). As shown in Table 4, the inclusion of SV improved the explanatory power by 1.8% (ΔR² = .018), and all three variables explained 11.3% of the variance of SM (R²_model2 = .113). The effect size of the predictive power of SV was calculated using the formula $f^2 = \Delta R^2 / R^2_{model} = .02$. According to Cohen (1988), this compares to a small effect size. Coefficient results indicated that there was no multicollinearity among the three variables (VIF < 5). Student grade positively predicted SM, whereas student gender did not. After controlling for the effect of both gender and grade, SV still presented a significant predictor of SM (see Table 4).

Taking the Z score of SV as the independent variable, and the Z score of the average score of the four elements of SM as the dependent variable, the path analysis was conducted. As shown in Table 5, SV significantly predicted model construction ($\beta = .253, p < .001$) and model evaluation ($\beta = .228, p < .001$). Increasing each unit in SV score, the average score of model construction and

![Percentage correct on the four elements of SM](image-url)
model evaluation increased by 0.253 and 0.228 units, respectively. But SV was not significantly predictive for model use ($\hat{\beta} = .083, p = .164$) and model revision ($\hat{\beta} = .060, p = .319$).

5.3 The Role of Grade in the Relationship of Spatial Visualization and Scientific Modeling

Regression analysis was conducted to answer Research Question 2. For each grade, one regression analysis was performed. In both analyses SV was treated as
the independent and SM as the dependent variable (still using z-standardized scores). The results indicated that SV was able to significantly predict SM in Grade 4 \((F = 8.199, p < .001)\) but not in Grade 7 \((F = .940, p = .334)\). Figure 4 shows the difference in the relationship between SV and SM in the two groups. Although the SM score of Grade 7 students was higher than that of Grade 4 students, the slope for Grade 4 students was significantly greater, indicating that SV and SM were more closely related in Grade 4 than in Grade 7. The regression equations of the two straight lines were:

\[
Z_{SM\_7th\ grade} = 0.27 + 0.08 \times Z_{SV\_7th\ grade}
\]

\[
Z_{SM\_4th\ grade} = 0.22 + 0.26 \times Z_{SV\_4th\ grade}
\]

With the same variable treatment, a path analysis was carried out for both Grade 4 and Grade 7. The results are shown in Table 6. In the fourth grade group, SV was significantly predictive for model revision \((\beta = .278, p = .011)\), but had no significant predictive effect on the other elements. In the seventh grade group, SV only significantly predicted model evaluation \((\beta = .244, p = .008)\) and had no predictive effect on the other three elements.
Discussion

This study investigated the relationship of SV and SM and analyzed individual differences in students’ grades and genders. The results of the field test suggest that SM is a challenging process for students, but also that some of the elements are more challenging than others. The comparatively difficult elements appear to be model evaluation and model revision. Model use appears to be relatively simple for students, suggesting that students are most capable with respect to that element. These findings are consistent with previous studies. Chang and Chiu (2009) found, for example, the average score of 10th grade students on an SM assessment to be about 1 point (with a full score of 5), and that most students struggled with evaluating their own models. Bamberger and Davis (2013) found only a few sixth grade students were able to evaluate models, and no student was able to adapt models to a new situation. Possibly, the evaluation and revision of models may involve more complex cognitive skills than other elements. Based on student performance on the field test, the following discussion focuses on the relationship between SV and SM, how this relationship varies by grade, and the issue of no gender differences.

6.1 The Role of Spatial Visualization in Scientific Modeling

The results of the hierarchical regression analysis show a positive relationship between SV and SM, which suggest that SV is positively involved in the process of SM practice. An explanation for this finding is that SV plays a role in the creation of mental models during the SM process and thus predicts students’ SM practice. Individuals with high SV ability are better at visualization competence (Gilbert, 2005), meaning that they are better at making sense of
phenomena with abstract, generative visual representations. Prior research has indicated that visualization competence plays an important role in the SM process (Griffard, 2013; Chang, 2022), which also reveals that SV and SM are related. The result of this study suggests we should increase the use of visual representations to scaffold students’ SM practice (Wang & Tseng, 2020; Glimcher, 2020).

The effects of SV on SM differ across the four elements, which suggests that the cognitive processes inherent in the four elements vary in the degree of their relationship to SV. SV plays a greater role when students construct their initial mental models and evaluate their own models in the model construction and model evaluation of the SM process, respectively. The findings may prompt us to reflect on and improve science curricula and the teaching and assessment of SM. A few studies on student SM practice have presented SM as a single task requiring students to build models and have tended to address students’ SM products (e.g., Ruppert et al., 2019; Bamberger & Davis, 2013). This result suggests that students may need more or better guidance and visual representations to support them in constructing their initial mental models and iteratively evaluating and revising their mental models into scientific models (Plummer et al., 2022). Science teaching should focus on creating a visualized learning environment to guide and facilitate students in visualization of their cognitive process of SM, especially of model construction and model evaluation.

It is worth noting that the results of the regression analysis in this study show that the effect size of SV on SM is small ($f^2 = .02$), and the contribution rate of the three variables (i.e., gender, grade, and SV) is 11.3%. This result of the study suggests that SM is a complex process involving many potential influencing variables that contain SV as just one of them. Further study worth considering could include more variables such as students’ relevant knowledge and other cognitive skills to systematically explain the role of SV in students’ SM practice.

6.2 Difference in the Relationship of Spatial Visualization and Scientific Modeling

The outcomes of the regression analysis show that SV significantly predicts SM in Grade 4 but not in Grade 7 which suggests that the two groups differ in their reliance on SV during the SM process. The effects of SV on SM differ amongst the four elements in the two grades, which confirm the significant difference in the relationship of SV and SM in the two grades. A possible explanation is that many of the higher-order scientific thinking engaging in SM practice also vary in grade. For example, students at the formal reasoning stage (i.e., about 12 years old; Inhelder & Piaget, 1964) can better use their scientific reasoning
abilities to solve problems. Therefore, for students from Grade 7, in addition to SV processing mental models, other cognitive skills play a key role in the SM process too. But the situation is different for students from Grade 4 (i.e., about 9 years old). Their higher-order abstract thinking skills are usually in their infancy stage, and they thus rely more on SV which has been developed (e.g., Piaget, 1954; Morris, 2018; Ogunkola & Knight, 2019) to deal with the SM practice. This result of the study suggests that we should make the best use of the critical role of SV in SM to further improve primary students’ SM ability.

6.3 No Difference between Genders in Spatial Visualization and Scientific Modeling

The finding that there were no gender differences in SM is consistent with previous research. The finding that there were also no gender differences in students’ SV abilities is noteworthy. For one, it aligns with the assumption that SV abilities play a major role in SM practice. In the STEM field, SV has been recognized as being associated with high academic achievement and a major predictor for a successful STEM career (Wai et al., 2009; Lubinski, 2010), fields which are currently dominated by males. Had there been a gender difference in SV abilities, we would have expected respective gender difference in students’ SM practice. Now the question is why there were no gender difference in either SV or SM, and to what extent the lack of gender differences may play out in female students’ STEM achievement and future high-tech careers. Would it be further conjectured that, physiologically, boys and girls are equal in the potential success in STEM careers since there are no gender differences between boys and girls in relevant higher-order cognitive abilities such as SV and SM? The findings of this study may suggest that science teachers should remove gender bias and give male and female students equal encouragement and opportunities in science learning so as to enable both groups equal chances of a successful STEM career and being competent for a high-tech future. This is a hypothesis worthy of being explored in future research.

7 Limitations, Conclusions, and Implications

The present study, despite providing important insights into the role of SV in SM, is subject to limitations. For one, the study employs convenience sampling. Hence the results need to be generalized with care, in particular to other samples (i.e., students from other grades but also students from the same grades but other societal or cultural backgrounds). The sampling of students from Grades 4 and 7 also renders the findings on individual differences in maturity preliminary. The question in fact is whether the differential role of SV in SM is
indeed a question of maturation or whether it is an issue that can be addressed through instruction. This is an open question that requires further research. Based on the results obtained in this study, the conclusion that can be drawn is that SV plays a substantial role in SM, and that this role is more substantial for Grade 4 than Grade 7 students.

The results of students' percentage correct across the four elements of SM suggest that students from Grades 4 and 7 struggle the most with the evaluation of models, although model use was relatively easy for them. To alleviate the challenges students face in the practice of SM and to better help students develop respective abilities, science teachers in China and the rest of the world should engage students more in the evaluation and revision of their models to rather than focusing on using models from teachers or textbooks. The detailed analyses of the role of SV in SM, which reflect these results, indicate that SV is an important ability for students to master in developing competence in SM. These conclusions imply that in primary science education more attention may need to be directed at supporting students in mastering this important ability, for example, through embedding training in this area within the curriculum and the use of visual representations to scaffold students' SM practices, especially to facilitate students' cognitive processes in the evaluation and revision of their mental models. In middle school science education, the specific role of the skills needed in model evaluation needs to be focused on, for example, through specific cues or scaffolds.

Further research will be needed to understand how science teachers can eliminate gender differences and hence create equal learning opportunities in the future for both male and female students. In summary, this study provides important evidence for the role of SV in SM and insights into where SV plays a specific role. This evidence will need to be corroborated and expanded by further research, in particular with respect to the mechanisms underlying this role and how it can be addressed in science teaching. The findings of this study may guide future improvement of science education in China and, potentially, other countries, in particular, with respect to attending to students' SV abilities as an important pre-requisite to the development of students' SM practices and hence the development of science literacy – effectively supporting ongoing science education reform efforts.

Abbreviations

ICA  Interactive computer assessment
SM  Scientific modeling
SV  Spatial visualization
Acknowledgements

We acknowledge the students, teachers, and principals involved in this study. We would achieve nothing without their collaboration. We would like to thank the support from Collaborative Innovation Center of Assessment for Basic Education Quality Foundation at Beijing Normal University (China) and National Natural Science Foundation of China.

Funding

This work was supported by the National Natural Science Foundation of China with a grant awarded to Jing Lin (Grant No. 62077008) and the Collaborative Innovation Center of Assessment for Basic Education Quality Foundation (China) with a grant awarded to Jing Lin (Grant No. 202101-103-BZK01).

Ethical Consideration

This study has gone through ethical approval by the research ethics committee at the corresponding authors’ university. The data reported in this study have received human subjects’ approval.

About the Authors

**Jing Lin** is an Associate Professor in Beijing Normal University (China), Director of Science Education Quality Development Department at Collaborative Innovation Center of Assessment for Basic Education Quality. Dr Lin is Executive Director of Alliance for Improving Scientific Literacy (AISL) for ALL, a committee member of NARST and a session editor of the *Eurasia Journal of Mathematics, Science and Technology Education*. Her research focuses on the assessment and improvement for K-12 students’ scientific literacy and teachers' professional competency, SSI learning, and STEM education.

**Letong Zhang** is a graduate student at the Collaborative Innovation Center of Assessment for Basic Education Quality at Beijing Normal University. His research focuses on scientific literacy assessment and development.
Knut Neumann is Director of the Department of Physics Education at the Leibniz-Institute for Science and Mathematics Education (IPN) and professor of physics education at the Christian-Albrechts-University of Kiel. His research interests include how to assess student competence and the development of student competence in science at various levels of education, how to support students in developing such competence and how to provide teachers with professional competence, in particular pedagogical content knowledge (PCK), to best support students in developing competence in science.

Ping-Han Cheng is an Associate Professor in the Department of Science Education, National Taipei University of Education (NTUE), Taiwan. His research focuses on science education, game-based learning, and ICT-assisted teaching for student learning and teacher training. Dr Cheng’s major research areas are earth science and sustainable development goals.

Wenting Wei is a graduate student at the Collaborative Innovation Center of Assessment for Basic Education Quality at Beijing Normal University. Her research focuses on scientific literacy assessment and development.

Chun-Yen Chang is a science education scholar in Taiwan. Currently, he serves as National Taiwan Normal University (NTNU) Chair Professor, Director of Science Education Center (NTNU), Professor at the Graduate Institute of Science Education and the Department of Earth Sciences (NTNU). Dr Chang now is the editor-in-chief of the *Eurasia Journal of Mathematics, Science and Technology Education*. He is also on the editorial board of several SSCI-level journals. His major research interests include science education, e-learning, interdisciplinary science learning, and science communication.

References


