

Evaluating the Geospatial Thinking of Pre-Service Geography Teachers Through an Orienteering Design Task: Analysis Using the Many-Faceted Rasch Model

Yanhua Xu, Ziqing Ou, Zhiting Wu, Yating Lin, Wei Zeng, Jiayan Yang, Jialu Li, Mengfan Shan & Yunqin Li

To cite this article: Yanhua Xu, Ziqing Ou, Zhiting Wu, Yating Lin, Wei Zeng, Jiayan Yang, Jialu Li, Mengfan Shan & Yunqin Li (2024) Evaluating the Geospatial Thinking of Pre-Service Geography Teachers Through an Orienteering Design Task: Analysis Using the Many-Faceted Rasch Model, *Journal of Geography*, 123:6, 153-164, DOI: [10.1080/00221341.2024.2394894](https://doi.org/10.1080/00221341.2024.2394894)

To link to this article: <https://doi.org/10.1080/00221341.2024.2394894>



Published online: 17 Sep 2024.



Submit your article to this journal [↗](#)



Article views: 286



View related articles [↗](#)



View Crossmark data [↗](#)



Evaluating the Geospatial Thinking of Pre-Service Geography Teachers Through an Orienteering Design Task: Analysis Using the Many-Faceted Rasch Model

Yanhua Xu^a , Ziqing Ou^b , Zhiting Wu^b , Yating Lin^b , Wei Zeng^b , Jiayan Yang^b , Jialu Li^b , Mengfan Shan^b , and Yunqin Li^b 

^aSchool of Geography and Environment, Jiangxi Normal University, JiangXi, China; ^bSchool of Geography, South China Normal University, Guangzhou, China

ABSTRACT

Measurement is significant in geospatial thinking research. This study evaluated the geospatial thinking of pre-service geography teachers through an orienteering design tasks and subjective evaluation methods. Moreover, it analyzed their level of geospatial thinking, difficulty of evaluation indicators and rater severity with the Many-Faceted Rasch Model. The results showed the geospatial thinking of participating pre-service geography teachers was at a moderate level. The estimation results of the model were well fitted. The evaluation indicators had a good degree of differentiation. Therefore, using the MFRM helps interpret the results of a geospatial thinking assessment in a more comprehensive way.

KEYWORDS

Geospatial thinking; many-facet Rasch model; pre-service geography teachers

Introduction

Since the National Research Council (2006) published *Learning to Think Spatially*, geospatial thinking has attracted attention from countries and regions around the world and has been incorporated into school curricula. Phillip Davis, director of the National Geospatial Technology Center, stated that geospatial thinking is a socially necessary skill. Zwartjes et al. (2017) pointed out that geospatial thinking, like other skills such as language, mathematics, and science, can and should be learned during one's schooling. Therefore, the development of geospatial thinking should be included in the field of basic education, as it is in the geography curriculum standards of some countries (Solari 2015). For example, the European Commission has attempted to encourage the inclusion of geospatial thinking in the curriculum by supporting the GI-Learner project, the goal of which is to develop an approach to learning spatial thinking that meets the needs of students (Zwartjes et al. 2017). The American the *National Geography Standards* (Heffron and Downs 2012) and current Chinese *General High School Geography Curriculum Standards* both emphasized that geospatial thinking plays an important role in fostering talents. In addition, a 1996 report by the French School Commission listed “understanding space and time” as one of six mandatory school outcomes.

There is a rising understanding in the geography-education community of the importance and usefulness of spatial thinking, as well as geography as a tool of developing spatial

thinking in the classroom (Jo and Bednarz 2014). The study of geospatial thinking is currently attracting the interest of geography-education researchers around the world, and there is a rich literature that explores connotations, components (Zwartjes et al. 2017) and factors associated with geospatial thinking (Low, Boger, and Mandryk 2014). Researchers have investigated the discipline (Liu and Guo 2021), instructional interventions (Collins 2018; Perugini and Bodzin 2020), and assessment tools (Lee and Bednarz 2009). The role of GIS (Geographic Information System) learning in the development of spatial thinking has also been explored (Perugini and Bodzin 2020); for example, Bodzin (2011) examined the contribution of GIS learning to the development of geospatial thinking. Moreover, scholars have focused on the impact of training in geospatial thinking on other skills (e.g., Hegarty et al. 2010).

The measurement of geospatial thinking is a key issue in geospatial thinking research. Early research on this domain was mainly in the field of psychology (Goldstein, Haldane, and Mitchell 1990), using, for example, spatial tests from the Kit of Factor Referenced Cognitive Tests (Ekstrom et al. 1976). On this basis, geographers and geography educators have continued to search for appropriate tools for assessing geospatial thinking (Charcharos, Tomai, and Kokla 2015). They have created some assessment tools that combine the characteristics of geography and geographic information technology to improve the accuracy of assessment (Huynh and Sharpe 2013; Kali, Orion, and Mazor 1997; Kerski 2003). However, few researchers have used hands-on methods, such

as the results of performance tasks, to collect data about levels of geospatial thinking.

Most scholars have used classical test theory to evaluate geospatial thinking. Few scholars apply the Multi-Faceted Rasch Model (MFRM), which is based on item-response theory to evaluation tasks. Compared to traditional knowledge-testing methods, MFRM is able to eliminate error caused by subjective factors such as raters and difficulty of test questions. For this reason, MFRM has been used in the field of education to analyze the effect of task difficulty of tasks, rater severity and rater scoring style on the scores received by test takers (Wu and Tan 2016).

The geospatial thinking of pre-service geography teachers plays a crucial role in the future teaching of geography and geospatial thinking. Therefore, in this paper, MFRM is used to analyze scores of pre-service geography teachers in a task assessment, with the goal of taking the task difficulty and rater severity into consideration to provide researchers with a guidance for using MFRM in the area of assessment of geospatial thinking.

Literature review and hypotheses

Geospatial thinking

The report of the National Research Council, “Learning to think spatially” defines “spatial thinking” as a kind of thinking related to the use of spatial concepts, representational tools and reasoning processes, which can use space to deconstruct problems, find answers and propose solutions. The “space” here refers to the subject matter of geography and earth sciences. The report emphasizes the important role of geospatial thinking in the discipline of geography (National Research Council 2006, Ishikawa 2015). Therefore, geospatial thinking can be seen as part of spatial thinking, a subset of spatial thinking concerned with the Earth’s surface and its representation (Huynh and Sharpe 2013). From this perspective, geospatial thinking is a collection of knowledge or skills that are learnable and can be improved through training.

Geospatial thinking plays an active role in many fields of science and is an important part of learning in the twenty first century (Azevedo, Osorio, and Ribeiro 2019). It is often used in the sciences, including in mathematics (Delgado and Prieto 2004), physics (Keehner et al. 2004), surgical training (Anastakis, Hamstra, and Matsumoto 2000), engineering (Hegarty and Waller 2006) and art and architecture (Chan 2008). Certain aspects are part of liberal arts courses, such as social sciences (Lobao 2003) and history (Knowles 2000). Numerous studies have confirmed that geospatial thinking is crucial for success in STEM fields (Wai, Lubinski, and Benbow 2009; Lubinski 2010; Cohen and Hegarty 2012), and spatial thinking increases the likelihood of students earning a degree or career in STEM (Keehner et al. 2006; Shea, Lubinski, and Benbow 2001). It can be argued that geospatial thinking is one of the key competencies for talent from now and into the future.

Individual neural differences, experience and the clarity of objective information can influence geospatial thinking.

Studies have been conducted to examine the effects (Casey, Nuttall, and Pezaris 1997), age (Solem et al. 2021), education, occupation, general intelligence, geographic expertise, map use habits, and interest in geographic learning on geospatial thinking.

Due to the plasticity of geospatial thinking and its confirmed extensibility (Uttal, Miller, and Newcombe 2013; Newcombe 2017), there is a growing interest in teaching geospatial thinking. Numerous empirical studies have reported on educational interventions to improve geospatial thinking, including (a) attention to the acquisition and use of spatial vocabulary (Gentner 2007); (b) the use of body language to complement mental maps (Newcombe 2010); (c) the use of teaching aids, mental maps, and virtual space software (Hauptman 2010); (d) extensive use of GIS (Manson et al. 2014); (e) geography games (Feulner and Kremer 2014); (f) Design interesting and challenging geography courses using a variety of geospatial techniques (Favier and van der Schee 2014, Carbonell-Carrera, Saorin, and Hess-Medler 2020).

Measurement of geospatial thinking

The measurement of geospatial thinking is an important topic in research related to geospatial thinking. To date, however, few assessment tools for spatial thinking have been produced, and measures and methodologies differ by discipline and study, such as testing spatial thinking on a geographic scale in geoscience contexts (Lee and Jo 2022). Researchers first focused on the development of psychometric scales (Havelkova and Hanus 2021; Hegarty and Waller 2006), such as psychometric scales and intelligence tests (Liben, Kastens, and Stevenson 2002), cognitive-ability tests (Battersby, Reginald, and Meredith 2006), and mental rotation tests (Pertusic, Varro, and Jamieson 1978). However, these paper-and-pencil psychometric scales can only measure small-scale geospatial ability (Hegarty et al. 2006), and geospatial ability cannot be equated with geospatial thinking (Wakabayashi and Ishikawa 2011). Studies have confirmed that there is not a perfect match between students’ scores on psychometric scales and their geospatial thinking performance (Ishikawa 2013). There has been a demand for further development of instruments that go beyond the psychological domain and that better represent geospatial thinking. Two such instruments are the Attitude toward Spatial Thinking Inventory (Milson and Alibrandi 2008) and the Self-Assessment of Spatial Habits of Mind (Kim and Bednarz 2013). In 2005, the SST (Spatial-Skills Test) was created to examine the effects of GIS learning on the spatial-thinking skills of university students (Lee 2005). Later, Lee and Bednarz (2009) developed and iteratively validated the STAT through a more rigorous and standardized research process, and it remains one of the most widely used measurement tools used. The STAT was used to investigate the effectiveness of a web-based GIS application in improving students’ geospatial thinking (Jo, Hong, and Verma 2016) and to identify the factors that influence geospatial thinking (Uttal and Cohen 2012). However, STAT does possess the drawbacks of paper-and-pencil tests. A hands-on approach

to assessment was used in creating a “tool to assess learners’ spatial thinking about the enhanced greenhouse effect” (Skaza 2016). Compared with STAT, this spatial thinking assessment project has the characteristics of contextualization, overcomes the shortcomings of generality, and focuses more on the environmental science content domain. Researchers in various fields have attempted to create similar assessments, for example, by measuring spatial thinking in a geographical context through a 30-question test that identifies characteristics of geospatial thinkers at different levels (Huynh and Sharpe 2013).

In summary, a number of geospatial thinking assessment tools have been developed, but all have their own drawbacks, and more valid and reliable measures of geospatial thinking continue to be investigated. In order to better analyze and interpret the evaluation scores of geospatial thinking, a multi-factor comprehensive evaluation method is adopted.

Many-faceted Rasch model

The Rasch model is one of the earliest models proposed by item response theory. It estimates individual’s potential ability or trait by analyzing the response to a series of items, which has a profound influence on subsequent psychometric models. However, the Rasch model also has some drawbacks: it only considers the difficulty of the project as the only project characteristic parameter, which may not be sufficient to describe all types of test items. MFRM is an extension of Linacre (1994) to the one-parameter Rasch model based on Rasch measurement theory. Compared with the Rasch model, this model extends the factors that affect the test result to multiple dimensions, including raters, grading standards, etc., so as to measure the real ability level of candidates more accurately. It has been used to analyze the many factors that influence test results including in language assessment (Batty 2015). In recent years, the MFRM has been used quite frequently in the fields of language assessment (Sims et al. 2020; Fan and Bond 2016; Hang 2011), pedagogy (Basturk 2008) and psychometrics and medical technology (Lawson and Brailovsky 2006).

Rater-mediated performance assessment is playing an increasingly important role in education and the behavioral sciences (Seo and Husein 2013). In assessments and examinations involving raters, there is often a need to face the challenge of scoring consistency (Wolfe 2004); that is, that scores not be affected by the difficulty of test questions or personal characteristics of raters. There is a growing interest in the potential of the MFRM to handle the many-faceted data that is produced through rater-mediated assessment. Related research has focused both on the mathematical and statistical foundations of the MFRM (Chalmers and Andrich 1991; Fischer and Molenaar 1995) and on its application (Bond and Fox 2001; Royal 2014).

However, few studies have applied MFRM to the assessment of geospatial thinking. In the past, Classical Testing Theory (CTT) was widely used to evaluate geospatial thinking to ensure the reliability and validity of the measured data. For example, De Miguel and De Lazaro tested the

spatial thinking level of students through the Learning Progression Scale, the reliability and validity of which were tested by Cronbach’s α quantization index, but the study ignored the impact of subjective differences of raters on the scoring results (De Miguel González and De Lázaro Torres 2020). Toru Ishikawa examined the relationships between geospatial thinking and spatial ability. In this study, two independent raters scored the geospatial thinking test, and the study conducted a scorer reliability analysis on the test data to prove the reliability of the score (Ishikawa 2013). The correlation coefficient obtained by this kind of traditional assessment method based on CTT only means that the assessment is consistent, but it cannot judge whether the score of the rater can accurately reflect the real level of the subject.

As one of the models of Item Response Theory, the MFRM can effectively make up for the shortcomings of traditional assessment methods. The MFRM parameterizes the various measurement dimensions of the scoring process and presents that these dimensions will collectively influence the score that testees receive. In the field of education, researchers have focused on the effects of test question difficulty, task difficulty, rater severity and the influence of dimensions like scoring style (Peeters, Sahloff, and Stone 2010). The statistical framework provided by MFRM can eliminate the influence of various factors in the subjective score on the score result and improve the reliability of the subjective score result (Linacre 1994). Therefore, the use of MFRM statistical tools can provide more information about the test and the candidate’s performance on the test to better analyze and interpret geospatial thinking assessment scores.

Methodology

Participants

This study was conducted in the School of Geographical Sciences at a teacher-training university in Guangdong Province, China, where the quality of pre-service geography teachers is among the highest in China. The study used a cluster sampling method, and 94 full-time, undergraduate, junior geography-science (teacher-training) majors participated, of whom 18 (19.00%) were male and 76 (81.00%) were female. The study took place during a compulsory professional course, Geography Pedagogy, and the grade for the performance task was part of the grade for the course. Because the course was important to the students, most of them worked hard to achieve high marks. Prior to this, the study participants had taken more than 10 compulsory courses in their major, such as physical geography, geological geomorphology, human geography, regional geography, cartography and GIS, and so had a certain foundation in geography.

Prior to the finalization of the research design, an exploratory focus interview was conducted to understand pre-service geography teachers’ perceptions of geospatial thinking. Five researchers in geospatial thinking participated in this exploratory focus interview. Three researchers believed that Lee et al.’s geospatial thinking evaluation index is

applicable to pre-service geography teachers. Some respondents believed that the difficulty of their test questions may be lower than the specific level of pre-service geography teachers, and suggested using subjective evaluation methods to complete the pre-service geography teacher's spatial thinking ability test.

Orienteering design task: A task-based assessment of pre-service teachers' geospatial thinking

In order to address the pre-service geography teachers' professional attributes to geospatial thinking, the pre-service geography teachers were asked to complete an orienteering design task, which included designing a one-day practical geography activity and recording it on a designated campus map. This task requires pre-service geography teachers to consult relevant materials and use the LocalSpace Viewer software to create a directional off-road map of Cunjinqiao Park in Zhanjiang City. The map needs to include 7 passing points (including starting point, ending point, and check-in point). Pre-service geography teachers need to choose one of the layout methods, such as closed, semi closed, open, and intersecting, based on factors such as terrain, personnel level, venue, time, and season. Moreover, the map needs to accurately indicate the starting and ending points, check-in points, and walking routes. The route design should consider accessibility and rationality, and be explained using text and pictures.

During the process of completing directional orienteering design tasks, pre-service geography teachers may face challenges that require the use of specific geospatial thinking abilities to solve problems. For example, when the

pre-service teachers choose waypoints, they need to identify point data in spatial information and further determine the positional relationship between waypoints. This requires pre-service geography teachers to have the ability to understand the depiction of geographic features and search for spatial advantage. Afterwards, the pre-service teachers may experience difficulty in positioning during the process of drawing maps, which requires them to understand orientation and positioning, such as front to back, east to west, north to south, east to west, horizontal and vertical.

Evaluation criteria

This study used the STAT (Lee and Bednarz 2012), as shown in Table 1. Based on the types of geospatial thinking ability tested by the questions in the STAT test, this paper analyzes the content of 16 questions and obtains 8 dimensions of geospatial thinking ability indicators: (a) understanding direction, (b) comparing map information, (c) choosing the best location based on geospatial elements, (d) visualizing a profile from a topographic map, (e) searching for spatial relevance, (f) visualizing a 3-D image from 2-D information, (g) layer overlays, and (h) understanding the depiction of geographical features. In order to enhance the differentiation of orienteering design task scoring and to measure pre-service teachers' geospatial thinking, the 0–1 scoring of the original scale was abandoned in favor of a 5-point Likert scale, with 5 = excellent and 1 = poor.

To assess the design proposals, six researchers in the field of geospatial thinking were selected as scorers, and they anonymously scored the designs according to the evaluation criteria.

Taking the activity design of Orienteering in Cunjinqiao Park in Zhanjiang City as an example, this study gathered 6 scholars with research experience in geospatial thinking. Based on 17 secondary indexes of three first-level indexes, namely, geographical practice ability, geospatial thinking and creativity, the Orienteering scheme designed by pre-service geography teachers is graded, and finally the average score of scholars is divided into data to measure the completion of pre-service teachers works. Among them, the secondary indicators of "geographical practical ability" include "spatial orientation," "access to information," "scientific demonstration" and "sharing and communication." The secondary indicators of "geospatial thinking" include "understanding of direction," "comparing map information," "choosing the best location based on geospatial elements," "visualizing a profile from a topographic map," "searching for spatial relevance," "visualizing a 3-D image from 2-D information," "layer overlays," and "understanding the depiction of geographical features." The secondary indicators of "creativity" include "novelty," "fluency," "flexibility," "precision" and "usefulness."

Taking the evaluation results of the orienteering activities in Cunjinqiao Park as an example, Figure 1 shows the Orienteering Map which was designed by student NO.20182621002. This scheme designs Orienteering map,

Table 1. Indicators of geospatial thinking.

Indicator	Contents
Understanding direction	Understand orientation and positioning, such as front to back, east to west, north to south, east to west, horizontal and vertical.
Comparing map information	Identify, reason about and graph spatial relationships for a single variable.
Choosing the best location based on geospatial elements	Using spatial reasoning, visualize, superimpose and manipulate spatial objects mentally without physically superimposing them on a map.
Visualizing a profile from a topographic map	Recognize spatial forms such as cross-sections, 3D boxplots; transform perceptions, representations and images from one dimension to another.
Searching for spatial relevance	Recognize and reason about the spatial relationships between groups of maps, including correlating and estimating distributional phenomena, and then draw spatial representations.
Visualizing a 3-D image from 2-D information	Reason spatially, orient oneself in a realistic environment, and visualize real-world, three-dimensional terrain from a two-dimensional topographic map.
Layer overlays	Reason spatially about maplayer Boolean logic (overlay vs. superposition).
Understanding the depiction of geographical features	Identify and visualize spatial data (points, lines, surfaces) and their spatial patterns from spatial information expressed visually or verbally; understand spatial shapes and patterns.

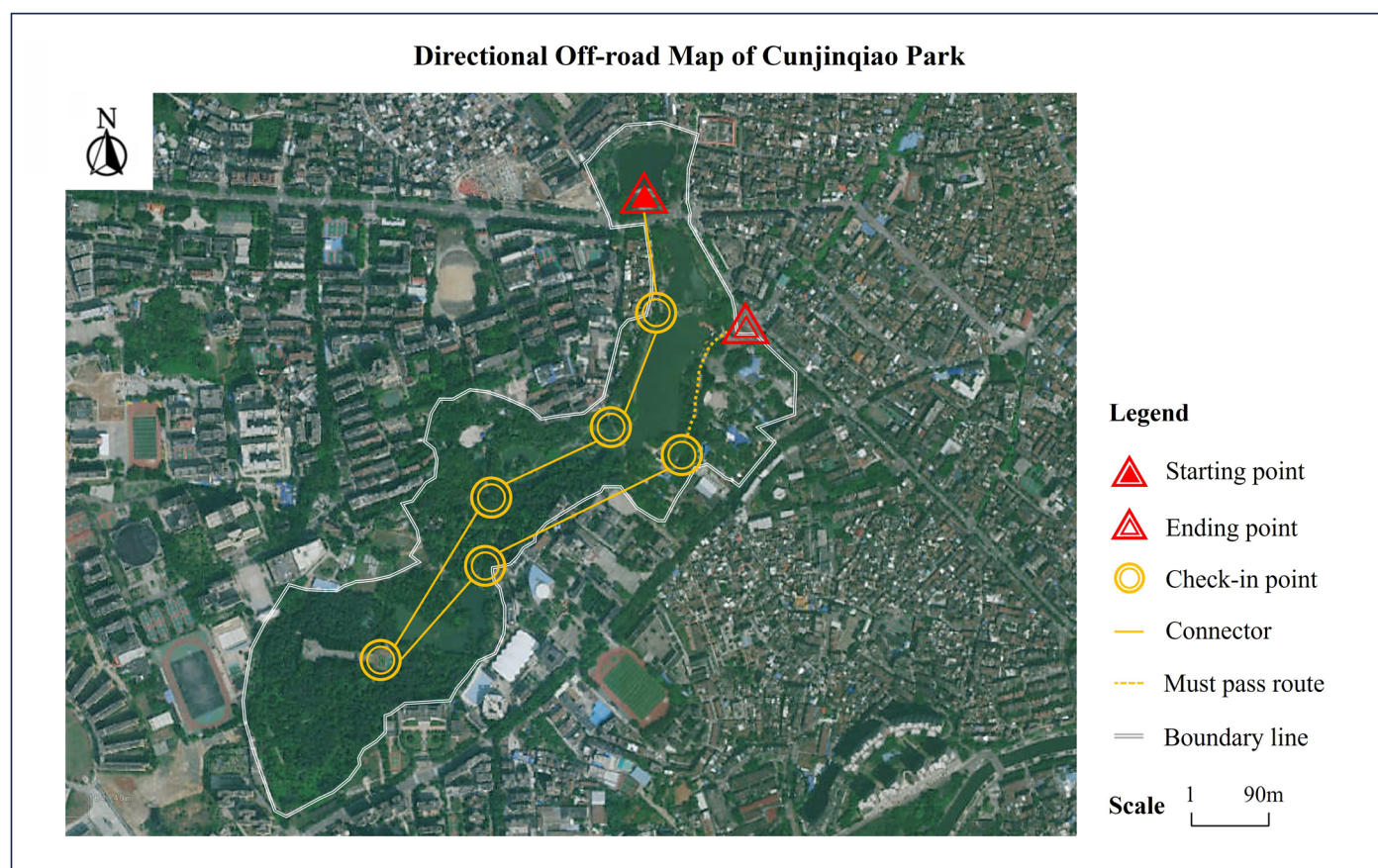


Figure 1. An example of orienteering design task.

and excavates the activity sites with local characteristics in the area for contestants to carry out landscape appreciation, landscape identification and other activities. It can well reflect the expected goal of “geographical practice power” and the goals of “understanding direction,” “comparing map information,” “choosing the best location based on geospatial elements” in “geospatial thinking.” At the same time, according to the local actual situation, the activity design also pays attention to the coverage of the area, excavates the natural and cultural landscape of CunJinqiao Park, the purpose is strong and the scheme design has good innovation. However, in the scheme design, there is a lack of geospatial thinking ability, such as “visualizing a profile from a topographic map,” “searching for spatial relevance,” “visualizing a 3-D image from 2-D information,” “layer overlays” and so on. The design of these contents needs to be strengthened in the scheme.

Rater training

Before the raters graded the orienteering design tasks, the research team explained the purpose of the study and the scoring criteria. Then, raters were asked to rate three randomly selected tasks according to the scoring criteria. Items with large differences in scoring and parts that the raters did not understand were discussed. Finally, all raters graded independently.

Data analysis

In this study, 94 pre-service teachers were organized to complete “orienteering design tasks,” 94 questionnaires were distributed, and 94 questionnaires were returned, with a recovery rate of 100%. This study used FACETS (version 3.83.6) software to create the MFRM and to analyze the data. The MFRM is a measurement model that Linacre extended from the Rasch model to analyze multiple variables with potential impact on measurement outcomes. FACETS software, which is commonly used to analyze MFRM data, is able to include the rater, the assessment task and the scoring criteria as three separate facets, so it was appropriate for use in this study (Linacre and Wright 2006). In this study, three dimensions were created using the FACETS software: the geospatial thinking of pre-service geography teachers, the difficulty of the evaluation indicators, and the severity of the raters. Raters were coded as judge 1, judge 2, ... judge 6. The 94 pre-service geography teachers were coded as student 1, student 2, ... student 94. Eight indicators of geospatial thinking were coded as item 1, item 2, ... item 8.

The dimensions considered in this study when evaluating geospatial thinking were both the difficulty of evaluation indicators and the severity of raters. For the n th pre-service geography teacher, with a geospatial thinking score of k , an evaluation-indicator score of l and a rater-severity score of j , the Logit for a rating of k by this rater for this indicator can be expressed as

$$\ln\left(\frac{P_{nljk}}{P_{nljk-1}}\right) = \theta_n - \delta_l - \alpha_j - \tau_k$$

where P_{nljk} is the probability that the n th student, with item l and rater j , will score k when the student's geospatial thinking score is k ; and P_{nljk-1} is the probability that the n th student, with item l and rater j , will score $k-1$ when the student's geospatial thinking score is k .

θ_n is the score for geospatial thinking of pupil n .

δ_l is the difficulty of evaluating indicator l .

α_j is the severity of rater j .

τ_k refers to the difficulty boundary between k and $k-1$ on the evaluation scale also known as the threshold difficulty.

The model shows that the pre-service geography teachers score for geospatial thinking, the difficulty of the assessment indicator, and the severity of the rater are the dimensions considered. In terms of assessing model fit, infit MnSq and outfit MnSq can be used to measure how well each individual's actual score fits the model predictions.

Separation and *reliability* are used to measure whether the difference between individuals for each facet is greater than the measurement error. Higher values of separation and reliability indicate greater confidence that the values for that facet are significantly different and that there are significant differences between individuals. Separation coefficients greater than 2 are generally considered to be significantly different, and a confidence level closer to 1 means that the data are more stable.

Results

This study assessed the geospatial thinking exhibited in 94 "orienteering design tasks" completed by participating pre-service geography teachers. A three-dimensional Rasch model was used to calculate students' geospatial thinking and rater severity, using the approach of Zhang and Wu (2008), which anchored the difficulty of evaluation indicators at 0 logit.

Summary statistics for the three-dimensional Rasch model analysis

The overall estimation results and fit of the model at each level are shown in Table 2. Scores for rater severity was much lower than those for difficulty of the evaluation indicators, showing that raters' scores were relatively lenient. The performance scores of the pre-service geography teachers tested were higher than the difficulty of the evaluation indicators, indicating that most performed quite well. They demonstrated a moderate level of geospatial thinking with a left-skewed distribution, and the results of the calibration were as expected.

In terms of fit, the mean square values for infit and outfit ranged from 0.7 to 1.3, which indicated suitability for analysis using the MFRM and the overall fit of the model was good. Reliability values for all dimensions are very close to 1, indicating a fairly high level of stability for this

assessment of geospatial thinking. Finally, the significance of the overall model was checked with a chi-square statistic (Hambleton and Swaminathan 1985), which showed that all levels of the model for this geospatial-thinking assessment were significant.

The Rasch model also provided the root mean square error (RMSE) for each dimension with RMSE values of 0.06, 0.06 and 0.21 for the evaluation scale, rater severity and the level of geospatial thinking, respectively, indicating very low measurement error for the three dimensions.

Analysis of three dimensions

Table 2 also presents the difficulty parameter estimates for each dimension, including the mean, logical measurement value, and degree of fit and separation.

"Measure" indicates the difficulty of the evaluation indicators, severity of the raters, and pre-service geography teachers ability on the logit scale, with higher scores indicating greater degree of each dimension. For the evaluation indicators, the highest value of 0.91 for "4: Visualize a profile from a topographic map" indicates that the most difficult task for pre-service geography teachers was to visualize three-dimensional topography in the real world from a two-dimensional topographic map. Item "6: Visualizing a 3-D image from 2-D information" was also not easy for the pre-service teachers to master. In contrast, indicator "1 Understanding direction" was the easiest for pre-service geography teachers to achieve, presenting lower parameter values. As for raters, all raters measured less than 0, indicating that raters as a whole maintained a consistent level of severity, they were all lenient. However, certain raters were relatively stricter or more lenient, with rater J2 being more lenient and raters J5 and J6 being stricter. Table 2 also presents the data for the 10 pre-service geography teachers with the highest scores for geospatial thinking. S74 and S82 have the highest values, indicating that they are strong geospatial thinkers. At the same time, all the pre-service teachers had measurements above 1 logit, indicating that they exhibited a high level of geospatial thinking.

The results of separation and chi-square tests indicated that there were significant differences in the evaluation indicators and geospatial thinking of the pre-service geography teachers. The separation between raters was 17.94, with significant differences between raters, meaning that if there had been a normally distributed parent group of similarly composed judges to this study, the differences in the severity of their ratings could have been divided into at least 17 tiers, which implies that even with trained and discussed judges, it is still difficult to disentangle the subjectivity of raters when rating. Meanwhile, estimate discrimination, one of the measures of model fitness, should reasonably range between 0.5 and 1.5 (Linacre and Wright 2006), with all ratings within an acceptable range of discrimination.

A review of the calculations of level of geospatial thinking in Table 3 shows that the mean progressively increases across the different evaluation indicators, that the

Table 2. Summary statistics for the three-dimensional rasch model analysis.

Facet	Reliability	Chi-square(df)	RMSE	Separation	Strata	Measure	Model SE	Infit MnSq	Outfit MnSq	Obsvd Average	Fair(M) Average	Estim. Discrm
Evaluation indicator	0.99	709.3(7)***	0.06	9.55	13.07	0.00	0.06	1.00	1.04			
1 Understanding direction						−0.98	0.06	1.01	1.14	2.51	2.68	0.84
2 Comparing map information						−0.32	0.06	0.91	0.95	2.17	2.21	1.10
3 Choosing the best location based on geospatial elements						−0.29	0.06	1.03	1.09	2.16	2.18	0.90
4 Visualizing a profile from a topographic map						0.91	0.07	0.89	0.76	1.63	1.47	1.19
5 Searching for spatial relevance						−0.27	0.06	1.03	1.00	2.15	2.17	1.05
6 Visualizing a 3-D image from 2-D information						0.76	0.07	1.14	1.42	1.68	1.53	0.90
7 Layer overlays						0.42	0.06	0.95	0.89	1.82	1.71	1.07
8 Understanding the depiction of geographical features						−0.24	0.06	1.06	1.09	2.13	2.15	0.95
Rater	1.00	1759.4(5)***	0.06	17.94	24.26	−1.70	0.05	1.01	1.04			
j1						−2.05	0.05	0.95	0.96	1.82	1.75	1.01
j2						−3.31	0.07	1.11	1.46	1.35	1.24	0.81
j3						−1.45	0.05	0.99	0.96	2.12	2.16	1.09
j4						−2.28	0.06	1.03	0.9	1.71	1.62	1.17
j5						−0.84	0.05	0.95	0.96	2.44	2.59	1.06
j6						−0.26	0.05	1.03	1.03	2.75	2.99	0.91
Pre-service teacher	0.96	1796.6(93)***	0.21	4.93	6.90	0.56	0.36	1.00	1.04			
s74						2.87	0.42	1.09	0.71	1.13	1.07	1.03
s82						2.70	0.39	0.93	1.62	1.15	1.08	0.89
s80						2.56	0.37	0.47	0.56	1.17	1.10	1.16
s68						2.43	0.35	1.03	0.72	1.19	1.11	1.01
s71						2.01	0.30	0.72	0.44	1.27	1.16	1.19
s57						1.77	0.28	0.70	0.43	1.33	1.21	1.34
s72						1.77	0.28	0.90	0.68	1.33	1.21	1.13
s13						1.42	0.25	1.13	1.03	1.44	1.29	1.00
s53						1.42	0.25	0.72	1.63	1.44	1.29	0.88
s55						1.42	0.25	0.92	1.71	1.44	1.29	0.94

Note. *** $p < 0.001$.

Table 3. Results of MFRM – level of geospatial thinking of pre-service geography teachers.

Level of geospatial thinking	M	Outfit MnSq	Rasch-Andrich thresholds		
			Measure	SE	Measure at −0.5
Poor (1)	−2.98	1.1	(none)	—	—
Minimal (2)	−1.68	1	−1.91	0.04	−2.52
Satisfactory (3)	−0.80	1.1	−1.22	0.04	−0.97
Above average (4)	0.18	1.1	0.52	0.06	0.62
Excellent (5)	1.20	1.1	2.61	0.14	2.86

measurements of category structure (Rasch-Andrich thresholds) for the levels were progressively higher, and that the adjacent measurements of category structure were less than 5 logits in magnitude, which means that the use of five-point Likert scales to differentiate between levels of geospatial thinking is appropriate (Linacre and Wright 2006). This indicates that five-point Likert scales is appropriate for applying the evaluation indicators.

Visualization of analysis results

Figure 2 presents three levels of internal variation from which the distribution of all objects can be seen. The “Measure” column indicates logit metric values ranging from

−4 to 3, spanning a total of 9 logits. Higher values signify a more relaxed set of requirements. The second column, “judge,” shows rater severity distributed roughly from 0 to −4 logits, with the most lenient rater, j6, at the top. The third column, “students,” shows the score for geospatial thinking of the pre-service geography teachers, with the most geospatially minded at the top of the table. The fourth column, “Scale,” presents the difficulty of the evaluation indicators, with the easiest being indicator 4.

A concentration trend in scoring can be seen in a probability plot, which is likely to exist if each band line for scores is widely spaced on the plot and has independent spikes (Myford and Wolfe 2003). As can be seen in Figure 3, although there are three separate spikes, the band lines

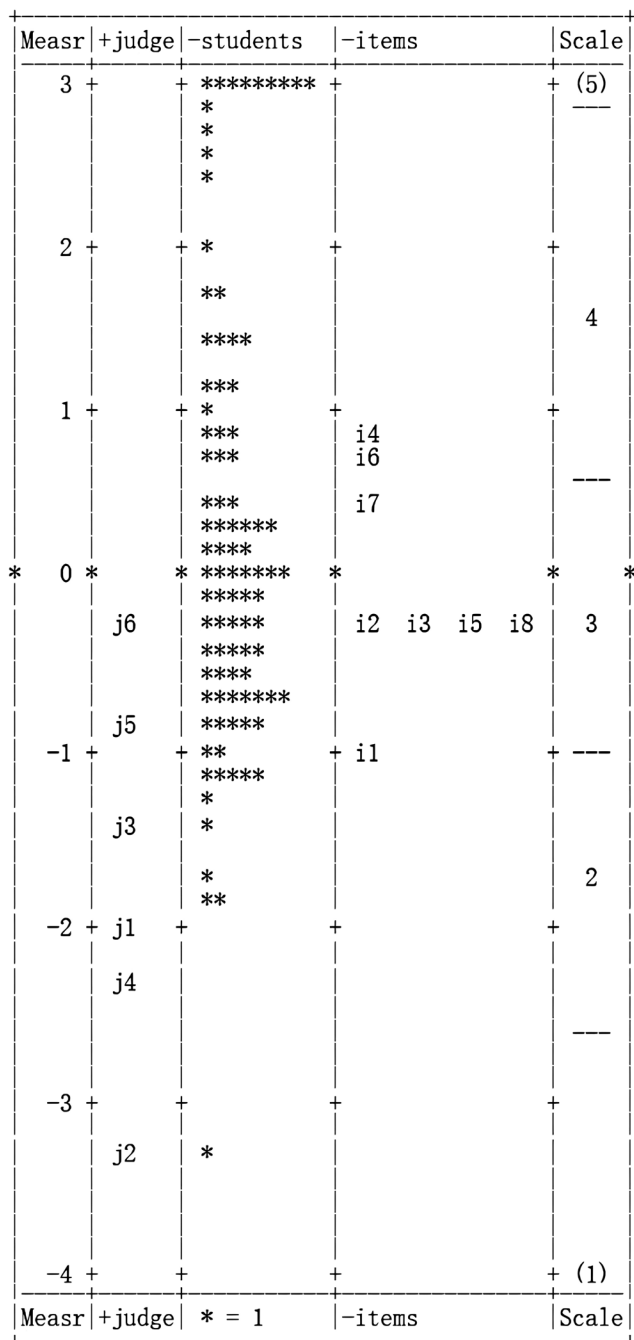


Figure 2. Vertical “rulers” for all facets of the MFRM. Measr=score on common scale, * = pre-service geography teacher; j=rater; i=indicator of geospatial thinking.

for the scores are not very widely spaced, thus again showing no significant concentration trend in scoring.

Discussion

We used the MFRM to measure the geospatial thinking of 94 pre-service geography teachers. From the results for the orienteering design tasks they had completed and the scoring data from six professional raters, we used a comprehensive method to analyze the scores of pre-service geography teachers and identify those pre-service geography teachers who were more geospatially minded.

As an assessment method for geospatial thinking for pre-service geography teachers, the MFRM is reliable, objective and innovative

Firstly, the MFRM is a reliable analysis method for pre-service geography teachers of geospatial thinking assessment data. The results showed that the model fit, reliability, and separation of the data in this study adequately met model requirements (Lee and Bednarz 2009); the stability of the information used met the assumptions of the Rasch model's single-item degree and so was suitable for the MFRM analysis. Secondly, analyzing the pre-service geography teachers of geospatial thinking assessment data by the MFRM made the results more objective because facets other than the ability of the participant were measured: the rater, the assessment indicator, and the task (Peeters, Sahloff, and Stone 2010). The MFRM was able to analyze at least three aspects of performance (the severity of the raters, the difficulty of the evaluation indicators, and the geospatial thinking of the participating pre-service geography teachers) in an integrated manner, as opposed to separate, single-facet assessments of geospatial thinking. Taking into account a wider range of information makes the analysis results more objective. Finally, the use of the MFRM to analyze the geospatial thinking of pre-service geography teachers is a methodological innovation. In the past, geospatial thinking was generally assessed through psychometric tests (Peters et al. 1995), cognitive-ability tests, and other means. The results of these assessment methods can be influenced by factors such as the reliability of questionnaires and measurement methods. In contrast, a literature search has revealed that the MFRM has been applied to such topics as the assessment of language tests (Fan, Knoch, and Bond 2019), doctor-patient performance (Zhang and Roberts 2012), public speaking (Wang et al. 2020), and travel competitiveness (Parra and Oreja 2014), and has been shown to be an operational and valid method. Therefore, the application of the MFRM to the analysis of geospatial thinking of pre-service geography teachers is available, and the findings of the study are meaningful.

Based on the MFRM, this study provides a more comprehensive and objective analysis of geospatial thinking among pre-service geography teachers

The results of the study show that there are differences in the level of geospatial thinking of pre-service geography teachers. On the one hand, most are better at perceiving, observing, understanding and comparing spatial elements. They received their highest scores for “understanding directions,” “comparing map information,” while “choosing the best location based on spatial elements,” “finding spatial relevance” and “understanding the representation of geographical features” received higher scores. These results may have been influenced by factors such as the characteristics of the environment in which the pre-service geography teacher lived and their degree of directional sensitivity. The ability to identify directions had been developed over the years in a

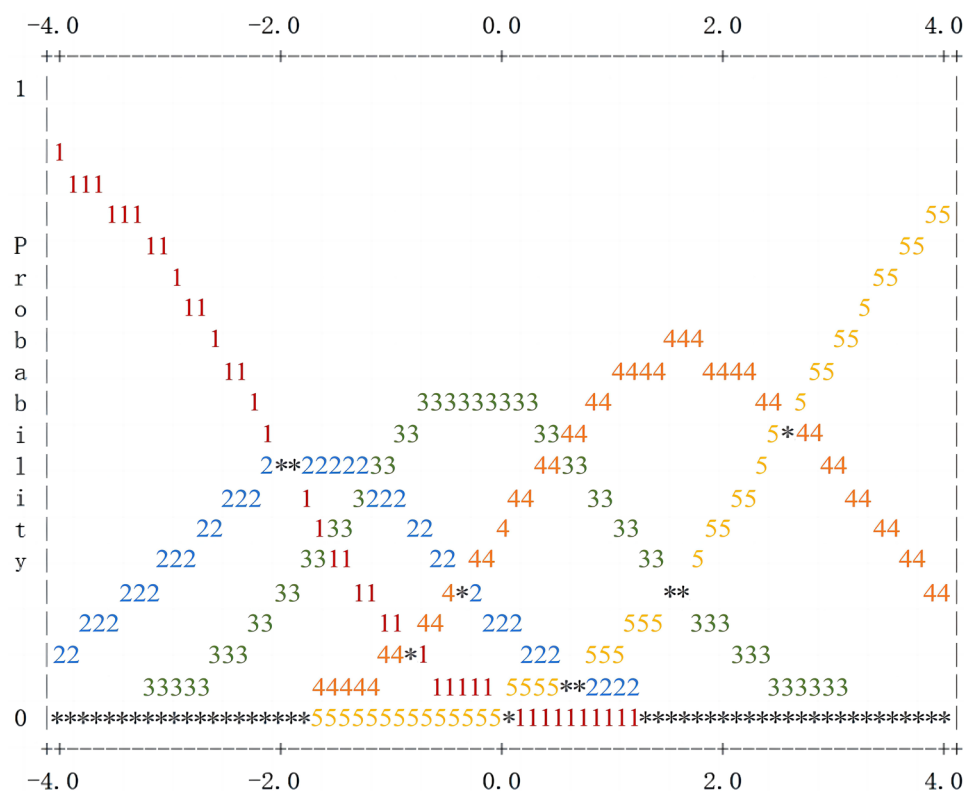


Figure 3. Probability curves.

learning context where “understanding directions” is required at the primary-school level (Wan et al. 2017) and has been transformed into an essential life skill, which is consistent with the findings of the study. Pre-service geography teachers generally have a certain level of geographical expertise, a strong interest in learning geography, better map-using habits and geographical problem-solving habits, which can facilitate the development of spatial thinking, and therefore they have a basic ability to observe, understand and compare spatial elements. On the other hand, the pre-service geography teachers had difficulties with spatial visualization involving three-dimensional space, and they scored low on the indicators of “visualizing a profile from a topographic map” and “visualizing a three-dimensional image based on two-dimensional information.” Spatial imagination is the ability to mentally create three-dimensional images, which is a type of abstract-thinking ability. Abstract thinking develops late and with difficulty in humans (Marini and Case 1994), and progressive training is often required to improve it. Research has demonstrated the role of GIS in enhancing spatial visualization (Lee and Bednarz 2009), which indicates that enhancing the geospatial thinking skills of pre-service geography teachers through the development of the spatial visualization skills is a direction worthy of attention. Therefore, in the process of cultivating pre-service geography teachers, geographical landscape images, maps, and other spatial imagery can be utilized as much as possible to develop the spatial imagination of pre-service geography teachers. Meanwhile, teaching pre-service geography teachers how to use GIS in daily teaching can cultivate their geospatial thinking skills.

Conclusion

In this study, an analysis of the geospatial thinking of pre-service geography teachers was conducted by using the MFRM to take into account the impact on the results of rater severity and of the difficulty of evaluation criteria. Firstly, the study demonstrates the reliability, objectivity and innovative nature of the MFRM as a method of analyzing the geospatial thinking of pre-service geography teachers, which can guide assessment creators in selecting appropriate raters and assessment criteria that promote fairness and accuracy. Secondly, the MFRM analysis was used to understand the actual geospatial thinking of pre-service geography teachers. The results indicate that they were more likely to have moderate geospatial abilities, finding skills related to spatial visualization and spatial analysis more challenging, which provides direction for those who train pre-service geography teachers and for geospatial thinking development.

There are, of course, certain limitations to this study. The pre-service geography teachers completed only one geospatial thinking task within this study, and a single task may not fully demonstrate all aspects of the geospatial thinking of pre-service geography teachers. Supplementing the study data with more work or exploring performance tasks that better reflect all aspects of geospatial thinking would help to improve the reliability and stability of the results of the study. Further, the raters chosen were trained to fulfill the fundamental criteria of the research in all indicators, so the data given were reliable. However, in future studies, more effective methods of rater training could be explored, or a pool of expert raters could be created, which would make the assessment more objective and valid.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This work was supported by the Guangdong Planning Office of Philosophy and Social Science under Grant number GD20CJY19.

Notes on contributors

Yanhua Xu is now working in School of Geography and Environment, Jiangxi Normal University. His main research interest is Geography education promotes students' lifelong development and educational measurement.

Ziqing Ou has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2022 to present.

Zhiting Wu has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2022 to present.

Yating Lin has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2023 to present.

Wei Zeng is now working in the School of Geography, South China Normal University. Her main research interest is geography curriculum and teaching in middle school.

Jiayan Yang has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2023 to present.

Jialu Li has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2023 to present.

Mengfan Shan has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2022 to present.

Yunqin Li has been enrolled in the School of Geography in South China Normal University, majoring in Teaching of Geography Discipline from September 2022 to present.

ORCID

Yanhua Xu  <http://orcid.org/0000-0001-8822-8891>

Ziqing Ou  <http://orcid.org/0000-0002-3701-3492>

Zhiting Wu  <http://orcid.org/0000-0002-3834-8593>

Yating Lin  <http://orcid.org/0009-0000-3393-7172>

Wei Zeng  <http://orcid.org/0000-0002-0839-6826>

Jiayan Yang  <http://orcid.org/0009-0004-8371-059X>

Jialu Li  <http://orcid.org/0009-0005-3185-9330>

Mengfan Shan  <http://orcid.org/0000-0002-2455-4812>

Yunqin Li  <http://orcid.org/0000-0002-9011-4823>

References

- Anastakis, D. J., S. J. Hamstra, and E. D. Matsumoto. 2000. Visual-spatial abilities in surgical training. *American Journal of Surgery* 179 (6):469–71. doi:10.1016/S0002-9610(00)00397-4.
- Azevedo, L., A. Osorio, and V. Ribeiro. 2019. GIT and augmented reality as tools for promotion and development of spatial thinking. In 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). New York: IEEE. doi:10.23919/CISTI.2019.8760995.
- Basturk, R. 2008. Applying the many-facet Rasch model to evaluate powerpoint presentation performance in higher education. *Assessment & Evaluation in Higher Education* 33 (4):431–44. doi:10.1080/02602930701562775.
- Battersby, S. E., G. G. Reginald, and J. M. Meredith. 2006. Incidental learning of geospatial concepts across grade levels: Map overlay. *Journal of Geography* 105 (4):139–46. doi:10.1080/00221340608978679.
- Batty, A. O. 2015. A comparison of video and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing* 32 (1):3–20. doi:10.1177/0265532214531254.
- Bodzin, A. M. 2011. The implementation of a geospatial information technology (GIT)-supported land use change curriculum with urban middle school learners to promote spatial thinking. *Journal of Research in Science Teaching* 48 (3):281–300. doi:10.1002/tea.20409.
- Bond, T. G., and C. M. Fox. 2001. Applying the Rasch model: Fundamental measurement in the human sciences. *Journal of Educational Measurement* 40 (2):185–7. doi:10.1111/j.1745-3984.2003.tb01103.x.
- Carbonell-Carrera, C., J. L. Saorin, and S. Hess-Medler. 2020. A geospatial thinking multiyear study. *Sustainability* 12 (11):4586. doi:10.3390/su12114586.
- Casey, M. B., R. L. Nuttall, and E. Pezaris. 1997. Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology* 33 (4):669–80. doi:10.1037/0012-1649.33.4.669.
- Chalmers, P. C., and D. Andrich. 1991. Rasch models for measurement. *The Statistician* 40 (1):119. doi:10.2307/2348242.
- Chan, W. D. 2008. Assessing visual-spatial talents: The use of the impossible figures task with Chinese students in Hong Kong. *High Ability Studies* 19 (2):173–87. doi:10.1080/13598130802504296.
- Charcharos, C., E. Tomai, and M. Kokla. 2015. Assessing spatial thinking ability. GEOTHNK International Closing Conference. doi:10.13140/RG.2.1.1621.0962.
- Cohen, C. A., and M. Hegarty. 2012. Inferring cross sections of 3D objects: A new spatial thinking test. *Learning and Individual Differences* 22 (6):868–74. doi:10.1016/j.lindif.2012.05.007.
- Collins, L. 2018. The impact of paper versus digital map technology on students' spatial thinking skill acquisition. *Journal of Geography* 117 (4):137–52. doi:10.1080/00221341.2017.1374990.
- Delgado, A. R., and G. Prieto. 2004. Cognitive mediators and sex-related differences in mathematics. *Intelligence* 32 (1):25–32. doi:10.1016/S0160-2896(03)00061-8.
- De Miguel González, R., and M. L. De Lázaro Torres. 2020. WebGIS implementation and effectiveness in secondary education using the digital atlas for schools. *Journal of Geography* 119 (2):74–85. doi:10.1080/00221341.2020.1726991.
- Ekstrom, R. B., J. W. French, H. H. Harman, et al. 1976. Kit of factor-referenced cognitive tests (spatial orientation). *Education testing service*, 1–314. Princeton, NJ: Office of Naval Research Contract.
- Fan, J., and T. Bond. 2016. Using MFRM and SEM in the validation of analytic rating scales of an English speaking assessment. In *Pacific rim objective measurement symposium (PROMS) 2015 conference proceedings*, ed. Quan Zhang, 29–50. Singapore: Springer Singapore. doi:10.1007/978-981-10-1687-5_3.
- Fan, J., U. Knoch, and T. Bond. 2019. Application of Rasch measurement theory in language assessment: Using measurement to enhance language assessment research and practice. *Papers in Language Testing and Assessment* 8 (2):iii–x.
- Favier, T. T., and J. A. van der Schee. 2014. The effects of geography lessons with geospatial technologies on the development of high school students' relational thinking. *Computers & Education* 76:225–36. doi:10.1016/j.compedu.2014.04.004.
- Feulner, B., and D. Kremer. 2014. Using geogames to foster spatial thinking. In *Gi Forum 2014: Geospatial innovation for society*, ed. R. Vogler, A. Car, J. Strobl, and G. Griesebner, 344–7. Vienna: Austrian Acad Science Press. doi:10.1553/giscience2014s344.
- Fischer, G. H., and I. W. Molenaar, eds. 1995. *Rasch models*. New York, NY: Springer New York. doi:10.1007/978-1-4612-4230-7.
- Gentner, D. 2007. Spatial cognition in apes and humans. *Trends in Cognitive Sciences* 11 (5):192–4. doi:10.1016/j.tics.2007.03.002.

- Goldstein, D., D. Haldane, and C. Mitchell. 1990. Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition* 18 (5):546–50. doi:10.3758/BF03198487.
- Hambleton, R. K., and H. Swaminathan. 1985. *Item response theory*. Dordrecht: Springer Netherlands. doi:10.1007/978-94-017-1988-9.
- Hang, L. I. 2011. Investigating the reliability of CET6 essay scoring: An application of generalizability theory and many-facet Rasch model. *Foreign Languages and Their Teaching*. http://en.cnki.com.cn/article_en/cjfdtotal-wywj201105013.htm.
- Hauptman, H. 2010. Enhancement of spatial thinking with virtual spaces 1.0. *Computers & Education* 54 (1):123–35. doi:10.1016/j.compedu.2009.07.013.
- Havelková, L., and M. Hanus. 2021. Upper-secondary students' strategies for spatial tasks. *Journal of Geography* 120 (5):176–90. doi:10.1080/00221341.2021.1981979.
- Hegarty, M., R. D. Crookes, D. Dara-Abrams, et al. 2010. Do all science disciplines rely on spatial abilities? Preliminary evidence from self-report questionnaires. In *Spatial cognition*, vol. VII, 85–94. Berlin: Springer. doi:10.1007/978-3-642-14749-4_10.
- Hegarty, M., D. R. Montello, A. E. Richardson, T. Ishikawa, and K. Lovelace. 2006. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence* 34 (2):151–76. doi:10.1016/j.intell.2005.09.005.
- Heffron, S. G., and R. M. Downs. 2012. *Geography for life: National geography standards 1994*. Washington, DC: National Council for Geographic Education. https://ncge.org/wp-content/uploads/2021/06/Geography_for_Life_2ndEd.pdf
- Hegarty, M., and D. Waller. 2006. Individual differences in spatial abilities. In *Handbook of visuospatial thinking*, 121–69.
- Huynh, T. N., and B. Sharpe. 2013. An assessment instrument to measure geospatial thinking expertise. *Journal of Geography* 112 (1):3–17. doi:10.1080/00221341.2012.682227.
- Ishikawa, T. 2013. Geospatial thinking and spatial ability: An empirical examination of knowledge and reasoning in geographical science. *The Professional Geographer* 65 (4):636–46. doi:10.1080/00330124.2012.724350.
- Ishikawa, T. 2015. Spatial thinking in geographic information science: Students' geospatial conceptions, map-based reasoning, and spatial visualization ability. *Annals of the American Association of Geographers* 106 (1):76–95. doi:10.1080/00045608.2015.1064342.
- Jo, I., and S. W. Bednarz. 2014. Dispositions toward teaching spatial thinking through geography: Conceptualization and an exemplar assessment. *Journal of Geography* 113 (5):198–207. doi:10.1080/00221341.2014.881409.
- Jo, I., J. E. Hong, and K. Verma. 2016. Facilitating spatial thinking in world geography using web-based GIS. *Journal of Geography in Higher Education* 40 (3):442–59. doi:10.1080/03098265.2016.1150439.
- Kali, Y., N. Orion, and E. Mazar. 1997. Software for assisting high-school students in the spatial perception of geological structures. *Journal of Geoscience Education* 45 (1):10–21. doi:10.5408/1089-9995-45.1.10.
- Keehner, M., Y. Lippa, D. R. Montello, F. Tendick, and M. Hegarty. 2006. Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Applied Cognitive Psychology* 20 (4):487–503. doi:10.1002/acp.1198.
- Keehner, M. M., F. Tendick, M. V. Meng, H. P. Anwar, M. Hegarty, M. L. Stoller, and Q.-Y. Duh. 2004. Spatial ability, experience, and skill in laparoscopic surgery. *American Journal of Surgery* 188 (1):71–5. doi:10.1016/j.amjsurg.2003.12.059.
- Kerski, J. J. 2003. The implementation and effectiveness of geographic information systems technology and methods in secondary education. *Journal of Geography* 102 (3):128–37. doi:10.1080/00221340308978534.
- Kim, M., and R. Bednarz. 2013. Effects of a GIS course on self-assessment of spatial habits of mind (SHOM). *Journal of Geography* 112 (4):165–77. doi:10.1080/00221341.2012.684356.
- Knowles, A. K. 2000. Introduction. Historical GIS: The spatial turn in social science history. *Social Science History* 24 (3):451–70. doi:10.1017/S0145553200010269.
- Lawson, D. M., and C. Brailovsky. 2006. The presence and impact of local item dependence on objective structured clinical examinations scores and the potential use of the polytomous, many-facet Rasch model. *Journal of Manipulative and Physiological Therapeutics* 29 (8):651–7. doi:10.1016/j.jmpt.2006.08.002.
- Lee, J. W. 2005. Effect of GIS learning on spatial ability. PhD diss., Texas A&M University. <https://www.proquest.com/dissertations-theses/effect-gis-learning-on-spatial-ability/docview/305374255/se-2>
- Lee, J., and R. Bednarz. 2009. Effect of GIS learning on spatial thinking. *Journal of Geography in Higher Education* 33 (2):183–98. doi:10.1080/03098260802276714.
- Lee, J., and R. Bednarz. 2012. Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography* 111 (1):15–26. doi:10.1080/00221341.2011.583262.
- Lee, J., and I. Jo. 2022. Assessing spatial skills/thinking in geography. In *Assessment in geographical education: An international perspective*, ed. T. Bourke, R. Mills and R. Lane, 77–97. Cham: Springer. doi:10.1007/978-3-030-95139-9_4.
- Liben, L. S., K. A. Kastens, and L. M. Stevenson. 2002. Real-world knowledge through real-world maps: A developmental guide for navigating the educational terrain. *Developmental Review* 22 (2):267–322. doi:10.1006/drev.2002.0545.
- Lim, K. Y. T. 2005. Augmenting spatial intelligence in the geography classroom. *International Research in Geographical and Environmental Education* 14 (3):187–99. doi:10.1080/10382040508668350.
- Linacre, J. M. 1994. *Many-Facet Rasch measurement*. 3rd ed. Chicago: Institute for Objective Measurement.
- Linacre, J. M., and B. D. Wright. 2006. A user's guide to BIGSTEPS: Rasch-model computer program, Last Modified January 2, 2006. Accessed April 30, 2024. <https://www.winsteps.com/a/bigsteps.pdf>
- Liu, Y. H. and Y. X. Guo. 2021. The current situation of geographic spatial thinking of high school students and its revelation. *Geography Teaching* 22:19–23.
- Lobao, L. 2003. Rural sociology and the “spatial turn” across the social sciences. *The Rural Sociologist* 23 (2):1–2.
- Low, R., R. A. Boger, and C. A. Mandryk. 2014. Connecting spatial literacy and climate literacy using a place-based GIS approach in a collaborative online educational setting. In *Agu Fall Meeting*. <http://adsabs.harvard.edu/abs/2014AGUFMED33B3511L>.
- Lubinski, D. 2010. Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences* 49 (4):344–51. doi:10.1016/j.paid.2010.03.022.
- Manson, S., J. Shannon, S. Eria, L. Kne, K. Dyke, S. Nelson, L. Batra, D. Bonsal, M. Kernik, J. Immich, et al. 2014. Resource needs and pedagogical value of web mapping for spatial thinking. *Journal of Geography* 113 (3):107–17. doi:10.1080/00221341.2013.790915.
- Marini, Z., and R. Case. 1994. The development of abstract reasoning about the physical and social world. *Child Development* 65 (1):147–59. doi:10.1111/j.1467-8624.1994.tb00741.x.
- Milson, A. J., and M. Alibrandi. 2008. Digital geography. https://xs.dailyheadlines.cc/books/about/Digital_Geography.html?hl=zh-CN&id=QfonDwAAQBAJ.
- Myford, C. M., and E. W. Wolfe. 2003. Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of Applied Measurement* 4 (4):386.
- National Research Council. 2006. *Learning to think spatially*. Washington, DC: National Academies Press.
- Newcombe, N. S. 2010. Picture this: Increasing math and science learning by improving spatial thinking. *American Educator* 34 (2):29.
- Newcombe, N. S. 2017. Harnessing spatial thinking to support stem learning. *OECD Education Working Paper* 161:51. doi:10.1016/b978-0-12-394293-7.00004-2.
- Pertusic, W. M., L. Varro, and D. G. Jamieson. 1978. Mental rotation validation of two spatial ability tests. *Psychological Research* 40 (2):139–48. doi:10.1007/BF00308409.
- Perugini, S., and A. M. Bodzin. 2020. Using web-based GIS to assess students' geospatial knowledge of hurricanes and spatial habits of mind. *Journal of Geography* 119 (2):63–73. doi:10.1080/00221341.2019.1710764.
- Peters, M., B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson. 1995. A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition* 28 (1):39–58. doi:10.1006/brcg.1995.1032.
- Royal, K. D. 2014. Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. *Psychometrika* 79 (4):733–5. doi:10.1007/s11336-013-9398-1.

- Seo, D., and T. Husein. 2013. Introduction to Many-Facet Rasch measurement: Analyzing and evaluating rater-mediated assessments. *Applied Psychological Measurement* 37 (2):173–5. doi:10.1177/0146621612469721.
- Shea, D. L., D. Lubinski, and C. P. Benbow. 2001. Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology* 93 (3):604–14. doi:10.1037/0022-0663.93.3.604.
- Sims, M. E., T. L. Cox, G. T. Eckstein, K. J. Hartshorn, M. P. Wilcox, and J. M. Hart. 2020. Rubric rating with MFRM versus randomly distributed comparative judgment: A comparison of two approaches to second-language writing assessment. *Educational Measurement: Issues and Practice* 39 (4):30–40. doi:10.1111/emip.12329.
- Skaza, H. 2016. Development and testing of an assessment to measure spatial thinking about enhanced greenhouse effect. UNLV Theses, Dissertations, Professional Papers, and Capstones, August. doi:10.34917/9302964.
- Solari, O. M. 2015. *Geospatial technologies and geography education in a changing world: Geospatial practices and lessons learned*, ed. A. Demirci and J. van der Schee. 1st ed. Tokyo: Springer.
- Solem, M., P. Vaughan, C. Savage, and A. S. De Nadai. 2021. Student- and school-level predictors of geography achievement in the United States, 1994–2018. *Journal of Geography* 120 (6):201–11. doi:10.1080/00221341.2021.2000009.
- Uttal, D. H., and C. A. Cohen. 2012. Spatial thinking and STEM education: When, why, and how? *Psychology of Learning and Motivation* 57:147–81. doi:10.1016/B978-0-12-394293-7.00004-2.
- Uttal, D. H., D. I. Miller, and N. S. Newcombe. 2013. Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science* 22 (5):367–73. doi:10.1177/0963721413484756.
- Wai, J., D. Lubinski, and C. P. Benbow. 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101 (4):817–35. doi:10.1037/a0016127.
- Wakabayashi, Y., and T. Ishikawa. 2011. Spatial thinking in geographic information science: A review of past studies and prospects for the future. *Procedia - Social and Behavioral Sciences* 21:304–13. doi:10.1016/j.sbspro.2011.07.031.
- Wan, J. Y., X. X. Lu, Y. Q. Lu, et al. 2017. Factors influencing the improvement of secondary school students geographic spatial thinking ability: An empirical study based on the senior students of the first middle school in Baiyin City, Gansu Province. *Progress in Geography* 36 (7):853–63. doi:10.18306/dlkxjz.2017.07.007.
- Wang, P., K. Coetzee, A. Strachan, S. Monteiro, and L. Cheng. 2020. Examining rater performance on the CELBAN speaking: A many-facets Rasch measurement analysis. *Canadian Journal of Applied Linguistics* 23 (2):73–95. doi:10.37213/cjal.2020.30436.
- Wolfe, E. 2004. Identifying rater effects using latent trait models. *Psychology Science* 46 (January):35–51.
- Wu, M. S., and S. Tan. 2016. Managing rater effects through the use of FACETS analysis: The case of a university placement test. *Higher Education Research & Development* 35 (2):380–94. doi:10.1080/07294360.2015.1087381.
- Zhang, X., and W. L. Roberts. 2012. Investigation of standardized patient ratings of humanistic competence on a medical licensure examination using Many-Facet Rasch Measurement and generalizability theory. *Advances in Health Sciences Education: Theory and Practice* 18 (5):929–44. doi:10.1007/s10459-012-9433-5.
- Zhang, X. L., and S. C. Wu. 2008. A Multi-Facet Rasch analysis on rating the academic scientific papers. *Psychological Testing* 551:105–28.
- Zwartjes, L., M. Lazaro, K. Donert, et al. 2017. *GI learner: Literature review on spatial thinking*. The European Commission. Europe. Accessed August 22, 2024. <https://www.gilearner.ugent.be/wp-content/uploads/GI-Learner-SpatialThinkingReview-3.pdf>.