Assessing Large-Scale Spatial Abilities in Virtual Reality: A Pilot Study on Map–Environment Transformations

Helen Müller (helen.mueller@uni-muenster.de) & Angela Schwering
Institute for Geoinformatics, University of Münster

Abstract

Large-scale spatial abilities rely on the capacity to transform between abstract spatial representations, such as maps, and embodied experiences of space. However, assessing these abilities remains challenging due to a trade-off between experimental control and ecological validity. This challenge has gained renewed relevance in light of increasing cognitive off-loading through digital navigation systems, underscoring the need for reliable instruments to measure and support spatial abilities over time. We introduce a virtual reality-based assessment targeting spatial transformations between maps and environments across multiple task facets, including landmark, location, direction, and route processing. A pilot study with students from grades 3 and 4 examines the feasibility of the assessment and its potential to differentiate between individuals and task difficulty levels. Initial results indicate systematic variation in item difficulty and meaningful interindividual differences, providing first evidence for the suitability of the proposed instrument for assessing large-scale spatial abilities.

Keywords: spatial cognition; assessment; virtual reality; large-scale spatial abilities; map-reading

Introduction

Spatial ability – although not uniformly defined in literature – is a central construct in cognitive psychology and encompasses a broad range of mental processes involved in perceiving, visualizing, transforming, and reasoning about spatial information (Linn & Petersen, 1985; Uttal et al., 2013). These abilities support a wide range of everyday activities, from manipulating objects to navigating through complex environments, and are associated with learning and performance across multiple domains, including STEM education and wayfinding (Hawes et al., 2022; Newcombe, 2016; Zhu et al., 2023; Wai et al., 2009).

A key dimension for structuring spatial abilities concerns the scale of the spatial information involved. Montello (1993) proposed a distinction between different spatial scales, ranging from small, manipulable spaces to large environments that extend beyond the observer's immediate field of view. Building on this distinction, research has often differentiated between small-scale spatial abilities, which involve objects or layouts that can be perceived from a single viewpoint, and large-scale spatial abilities, which require navigation, orientation, and the integration of spatial information across multiple perspectives, often supported by the use of maps as external spatial representations. Empirical evidence suggests that these two classes of abilities overlap only partially. Studies with both adults and children indicate that performance on small-scale spatial tasks predicts large-scale spatial performance only to a limited extent, pointing to partially distinct underlying cognitive processes (Hegarty et al., 2006; Montello, 2005; Li et al., 2019; Lasc et al., 2023). Environmental-scale navigation remains cognitively demanding, as large spaces can only be apprehended through locomotion and require continuous updating of spatial relations and self-positioning (Ishikawa & Newcombe, 2021; Holmes et al., 2018). Although a small number of instruments address abilities related to this scale - for instance, the perspective-taking test (Kozhevnikov & Hegarty, 2001; Hegarty & Waller, 2004) – these measures focus predominantly on viewpoint transformations. Consequently, they fail to capture navigation-related skills that depend on locomotion, sequential decision-making, route learning, and the dynamic updating of spatial representations during movement.

The majority of established spatial ability assessments focus on small-scale spatial abilities. Widely used instruments include paper-and-pencil tests such as the Mental Rotations Test (Vandenberg & Kuse, 1987), the Cards Rotation Test (French et al., 1963), or the Paper Folding Test (Ekstrom et al., 1976). These measures offer high experimental control and ease of administration, but they capture spatial reasoning at a scale that differs fundamentally from the demands of real-world navigation and orientation, especially when maps are used as spatial representations. One reason for the comparatively few instruments for assessing large-scale spatial abilities might be methodological: large-scale spatial tasks typically require interaction with extended environments, which are difficult to control, replicate, and standardize. Field-based navigation tasks often depend on specific locations, are sensitive to environmental variability, and are therefore not easily transferable across contexts or suitable for repeated measurements. This lack of location-independent assessment tools might have limited systematic research on large-scale spatial abilities and map-environment transformations.

Recent advances in immersive technologies offer new opportunities to address these challenges. Virtual reality (VR) allows users to navigate extended environments under controlled and repeatable conditions while preserving key characteristics of large-scale spatial behavior, such as movement, perspective changes, and self-referenced orientation. Moreover, VR environments can be coupled with digital spatial representations, enabling the systematic study of how individuals relate maps to the environments they represent. Building on these possibilities, the present paper introduces and pilot-evaluates a VR-based assessment

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

designed to capture a central but under-assessed component of large-scale spatial ability: the ability to transform spatial information between maps and environments.

Conceptual Framework

Large-scale spatial abilities operate at what Montello (1993) termed the *environmental scale*, referring to spaces that extend beyond the observer's immediate field of view and cannot be apprehended from a single perspective. At this scale, spatial cognition relies on navigation, orientation, and the integration of spatial information across movement and time. Successful behavior in environmental-scale spaces requires individuals to continuously update their position and orientation while relating local perceptual information to more global spatial representations. These demands distinguish large-scale spatial abilities from small-scale spatial skills and make them particularly sensitive to representational format and learning context.

Within this domain, maps constitute a central representational tool. As symbolic representations, maps depict selected features of a referent space in a structured and scaled format, allowing users to access spatial information that may not be directly perceivable in the environment (Blades & Spencer, 1987; Davies & Uttal, 2007). Using a map for navigation requires more than interpreting information within the map itself. Rather, effective map use depends on transforming spatial information between representations of the same space.

Lobben (2007) defined the concept *navigational map reading* as a person´s ability to use map-based information to reach an unfamiliar destination in an unknown environment. This skill involves both cognitive skills and representational transformations that enable individuals to decode map information, relate them to spatial structures in the environment, and translate this information into navigational actions. Accordingly, navigational map reading is not understood as a single, isolated skill but as a competence arising from the coordinated interaction of several spatial processes (Lobben, 2007).

A central characteristic of navigational map reading is the need to transform spatial information between representations. Maps and environments differ fundamentally in their representational properties: maps provide a stable, allocentric overview of space, whereas environmental experience is inherently egocentric, sequential, and action-based. Liben and Downs (1989) proposed that map-use performance emerges from the interplay between (a) characteristics of the map itself, including symbolic features such as legends and geometric properties such as scale, orientation, and dimensionality; (b) properties of the reference environment, such as familiarity, size, and structural complexity; and (c) characteristics of the map reader, including developmental level, prior knowledge, and spatial experience. The transformation between the two representations occurs in two directions: from map to environment, such as when planning or following a route based on map information, and from environment to map, such as when determining one's current position or encoding experienced landmarks onto a map.

Empirical research supports the distinction between these transformation directions. Thorndyke and Hayes-Roth (1982) demonstrated that spatial knowledge acquired through direct navigation differs systematically from knowledge acquired from maps, with navigation experience supporting accurate local judgments and pointing, whereas map-based learning supports more configurational, survey-like representations. Similar dissociations have been observed in virtual environments, indicating that these effects reflect general properties of spatial representation rather than characteristics of physical space alone (Ruddle et al., 1997).

A systematic review of map-related research further indicates that, despite considerable variation in terminology, studies have relied on a relatively stable set of task demands to operationalize navigational map use. Arthurs et al. (2023) distinguished in a literature review between within-map tasks and map-in-environment tasks, with the latter requiring users to actively relate a map to the environment it represents. The reviewed map-in-environment tasks include locating places/objects, determining direction and distance, finding an object in a real space from a map, self-location, orienting one´s self in the correct direction, determining a route, giving directions, and navigation.

Taken together, these task types share a common structural requirement: they all depend on the availability of both a map and a corresponding referent environment to which the map must be related. Unlike within-map tasks, map-in-environment tasks are inherently situated, as they require interaction with a specific spatial context. This reliance on real-world environments poses substantial methodological challenges. Outdoor spaces differ in their spatial structure, landmark configuration, and scale, and even the same location may vary over time due to changing environmental conditions, visibility, or contextual cues. As a result, map-in-environment tasks are difficult to replicate across studies, hard to standardize, and often limited in their comparability across participants and testing contexts. In this context, virtual reality (VR) offers a promising methodological alternative for studying navigational map use and large-scale spatial abilities. VR allows users to move through and interact with simulated spaces while preserving essential characteristics of real-world spatial behavior - such as embodied movement, changing viewpoints, and continuous perception-action coupling - while ensuring standardization and experimental control (Pastel et al., 2022; Dong et al., 2021).

Building on these methodological advantages, VR has been increasingly employed in spatial learning research. A growing body of evidence suggests that immersive virtual environments can support the acquisition of spatial knowledge, perspective taking, and navigation-related skills (Lin et al., 2014; Carbonell-Carrera & Saorin, 2017; Molina-Carmona et al., 2018). In addition, VR has also been adopted as a platform for assessing spatial abilities. Prior work has demonstrated that navigation behavior in virtual environments reflects meaningful individual differences in real-world wayfinding, supporting the ecological validity of VR-based spatial assessments (Coutrot et al., 2019). VR has also been used to measure small-scale spatial abilities such as mental rotation and spatial visualization under immersive conditions (Hartman et al., 2006), as well as large-scale spatial abilities involving navigation, orientation, and spatial updating in extended virtual environments (Shimada et al., 2016).

Assessment Design & Tasks

Conceptual Structure of the Assessment

In the present work, the theoretically grounded task types were systematically integrated into a coherent assessment framework and six games, each corresponding to one task category, were derived (Table 1): Landmarks, Locations, Directions, Routes, Landmark Piloting, and Distances. Each game targets a distinct but related component of navigational map reading. All games are implemented symmetrically in two transformation directions: from environment to map (the first half of each game) and from map to environment (the second half of each game). This bidirectional design explicitly addresses spatial transformations between representations, which have been shown to rely on partially distinct cognitive processes (Thorndyke & Hayes-Roth, 1982). Within each game, multiple tasks with increasing difficulty are included, allowing sensitivity to individual differences across a broad ability range.

Task Categories and Difficulty Manipulations

Landmarks (Game A). Landmark tasks assess the identification and relational use of salient environmental features. Landmarks are designed to be unique and configurationally salient (von Stülpnagel & Frankenstein, 2015), with increasing difficulty based on their spatial embedding and relationships to surrounding elements (Caduff & Timpf, 2008). More difficult tasks require processing configurations of landmarks rather than single landmark cues.

Locations (Game B). Location tasks focus on representing positions in space, including self-location and target-location judgments. Tasks rely on both proximal and distal landmarks, with increasing environmental complexity operationalized through larger environments, a greater number of objects, and more structured spatial layouts (Liben, 1997).

Directions (Game C). Direction tasks require relating facing direction in the environment to directional information on the map. Difficulty is manipulated via systematic map–environment misalignment (aligned, 90°, 180°; Presson, 1982; May et al., 1995) and by varying whether orientation judgments are anchored to street directions or to more distant landmarks.

Routes (Game D). Route tasks assess the ability to follow and reconstruct paths through the environment. Difficulty increases through variations in street geometry (orthogonal vs. oblique angles), the number of decision points, and the availability of alternative routes at intersections (Gärling et al., 1986).

Landmark Piloting (Game E). Landmark Piloting tasks target navigation based on sequences of landmarks rather than metric or survey knowledge (Gallistel, 1990). Difficulty is manipulated by increasing the number of landmarks involved and by embedding them in more complex environmental configurations, following principles of spatial complexity described by Liben (1997).

Distances (Game F). Distance tasks assess the ability to compare and interpret relative distances between objects. In simpler tasks, one object differs in distance from a set of otherwise equidistant objects. More complex tasks require processing configurations of multiple distance relations, such as identifying an object based on its position relative to several nearby objects.

The present pilot study reports analyses for the categories Landmarks, Locations, and Directions; Routes, Landmark Piloting, and Distances are part of the assessment framework but will be examined in subsequent validation studies.

Assessment Environment and Platform

The assessment was implemented using GeoGami, an app for creating and playing map-based geogames in real and virtual environments that combine exploration, problem-solving, and spatial reasoning (Bartoschek et al., 2010). GeoGami supports the systematic investigation of orientation and

Table 1: Conceptual structure of the assessment and associated tasks.

Real world → map		Map → real world
Mark an object from reality on the map	**Landmarks**	Identify an object marked on the map in the real world
Self-location	**Locations**	Navigation to a place marked on the map
Mark viewing direction on the map	**Directions**	Turn into a direction marked on the map
Map a route walked	**Routes**	Follow a marked route
Map a route walked	**Landmark Piloting**	Follow a marked route
Map an object with relation to other objects	**Distances**	Navigation to a place in relation to distances of objects

navigation skills through customizable task types, adjustable difficulty levels, and configurable map functions such as zoom, rotation, and position display. For the present study, a VR-based version of GeoGami was used, allowing participants to navigate immersive environments while interacting with a dynamically linked digital map. All user interactions - including navigation trajectories, orientation changes, map interaction, responses, and task durations - are automatically recorded, enabling detailed behavioral analyses in addition to accuracy-based scoring.

Method

Participants

Participants were 60 students from grades 3 and 4 recruited from a single primary school. The sample comprised 35 boys and 25 girls, with ages ranging from 8 to 10 years ($M = 8.73$, $SD = 0.60$). 22 students attended grade 4 and 38 attended grade 3. This age group was selected because grades 3 and 4 represent a developmentally relevant stage for spatial orientation and perspective-taking (Piaget & Inhelder, 1948; Schäfer, 1984) and, as the youngest target group of the larger study, were especially informative for testing task and technology feasibility. Participation was voluntary and took place during an after-school program. Written informed consent was obtained from parents, and verbal assent was obtained from all children prior to participation.

Apparatus and Materials

The assessment was administered using a tablet-based virtual version of the GeoGami platform (Fig. 1). Tasks were presented on iPads, and interaction was realized via touch input, as pilot testing indicated that children handled tablets more easily than laptops due to greater familiarity. The digital map was dynamically linked to the virtual environment, allowing participants to switch between map-based and environment-based representations. All task interactions and navigation behavior were automatically logged.

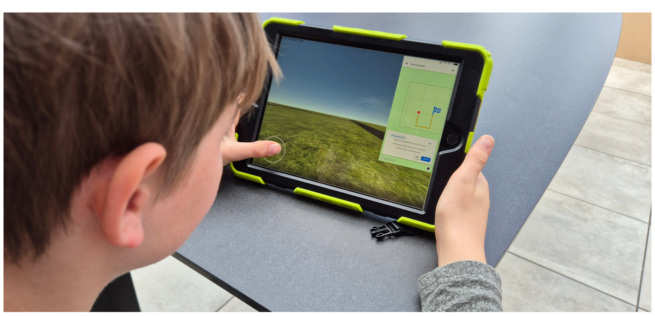

Figure 1: Tablet-based test environment used in the study.

Procedure

Testing was conducted in small groups of 4 to 10 students. Each session began with a joint introduction explaining the general task format and controls, followed by a short practice game that allowed participants to familiarize themselves with the interaction and task structure. The main assessment consisted of a series of map–environment transformation tasks ordered in six games (Games A to F). Game order was randomized for each participant. All students received the same set of tasks; however, due to the fixed testing time of approximately 45 minutes per session, not all participants completed all games.

Data Collection

All interaction data were stored in a GeoJSON format and analyzed using the platform's internal analysis tools. Logged data included task responses, task duration, and navigation trajectories. Navigation paths were additionally inspected manually. The present analyses focus on the task categories Landmarks (A), Locations (B), and Directions (C), which were fully implemented in both transformation directions at the time of data collection.

Data Analysis & Results

Task performance was analyzed using both classical test theory and item response theory (IRT) to examine item difficulty, discrimination, and measurement precision. All analyses were conducted in R (Version 4.5.2; R Core Team, 2025), using the packages tidyverse (Wickham et al., 2019) for data handling and visualization and TAM (Robitzsch et al., 2025) for Rasch modeling. Task responses were scored dichotomously (0 = incorrect, 1 = correct). Within each task category, items were divided equally between the two transformation directions; for example, in Game A, items A1–A2 assessed real world to map transformations and A3–A4 assessed map to real world transformations, while in Game B, items B1–B3 and B4–B6 corresponded to these two directions, respectively. Item difficulty was examined using solution rates (p), and item discrimination was assessed via corrected item–total correlations within task categories. Internal consistency was estimated using KR-20, which is mathematically equivalent to Cronbach's α for dichotomous items. Given the relatively small number of items per category, reliability estimates are reported descriptively. Separate Rasch models were fitted for each task category and transformation direction. Item fit was evaluated using infit and outfit statistics, which showed no pronounced deviations from model expectations. Item information functions and test information functions were derived to examine measurement precision across the ability continuum.

Items A1–A4 in Game A (Landmarks) showed substantial variability in difficulty, with solution rates ranging from low to high (Tab. 2). Item–total correlations were moderate overall. Conditional reliability varied across the ability continuum and remained comparatively low across most ability levels, with localized increases in the lower to mid-

[1] p = proportion correct; r_it = corrected item–total correlation; Rasch b = item difficulty (logits); KR-20 = internal consistency; r = correlation between transformation directions; ladder violations = proportion of Guttman order violations.

range of θ. Item information curves indicated that individual items contributed information in relatively narrow ability regions (Fig. 2).

Table 2: Item statistics and reliability estimates for game A[1]

Item	p	r_it	Rasch b	Infit	Outfit
A1	0.804	0.34	-2.39	0.948	0.711
A2	0.321	0.34	1.30	0.935	0.781
A3	0.821	0.419	-2.66	0.977	0.656
A4	0.446	0.419	0.402	0.95	0.807

Process	n_items	KR20	r (between processes)	Ladder violations
A_RW_map	2	0.53	0.445	0.143
A_Map_RW	2	0.602	0.445	0.25

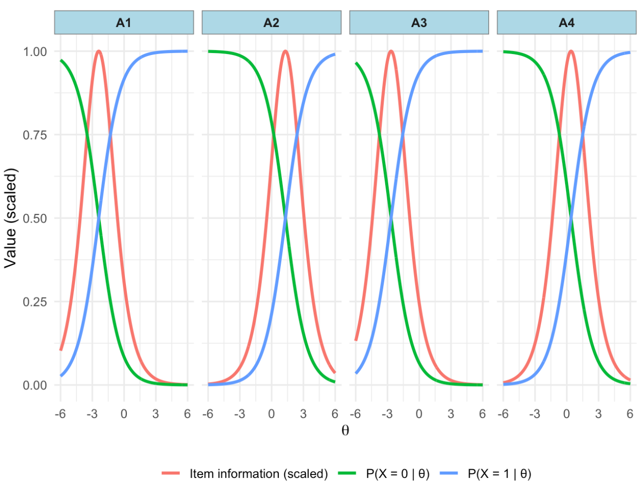

Figure 2: Item response and information curves for game A

Game B (Locations) included items B1–B6 spanning a moderate range of difficulty levels (Tab. 3). Item–total correlations varied across items, with some items contributing more strongly to total score variance than others. Conditional reliability reached higher values than in Game A in the central ability range, while decreasing toward the lower and upper ends of the continuum. Item information was unevenly distributed across items, reflecting differences in item difficulty and discrimination (Fig. 3).

Table 3: Item statistics and reliability estimates for game B

Item	p	r_it	Rasch b	Infit	Outfit
B1	0.768	0.32	-1.82	0.998	0.926
B2	0.446	0.418	0.329	0.965	0.921
B3	0.214	0.315	1.97	1.02	1.19
B4	0.714	0.495	-1.66	0.928	0.738
B5	0.518	0.615	-0.125	0.863	0.665
B6	0.179	0.313	2.70	1.07	1.80

Process	n_items	KR20	r (between processes)	Ladder violations
B_RW_map	3	0.668	0.543	0.036
B_Map_RW	3	0.553	0.543	0.089

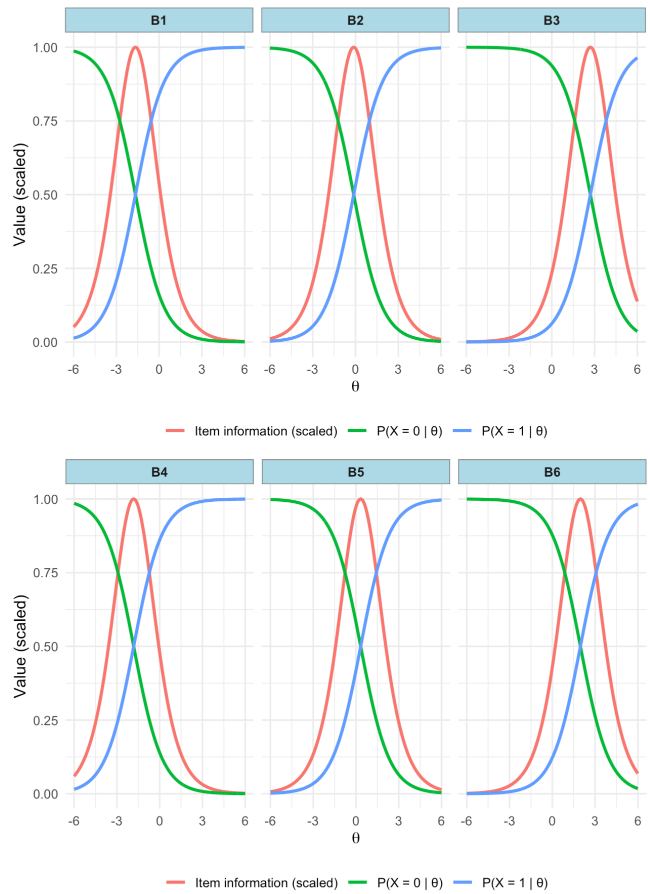

Figure 3: Item response and information curves for game B

Items C1–C8 in Game C (Directions) covered a broader range of difficulty levels and showed more homogeneous item–total correlations (Tab. 4). Conditional reliability was higher in the mid-range of ability and declined more gradually toward the extremes of θ. Item information curves suggested that multiple items contributed information across overlapping regions of the ability continuum (Fig. 4).

Table 4: Item statistics and reliability estimates for game C

Item	p	r_it	Rasch b	Infit	Outfit
C1	0.946	0.13	-4.56	1.25	1.38
C2	0.536	0.511	-0.332	1.01	1.09
C3	0.518	0.575	-0.202	0.931	0.913
C4	0.25	0.592	1.93	0.83	0.559
C5	0.571	0.306	-0.446	1.10	1.12
C6	0.464	0.36	0.195	1.03	1.01
C7	0.232	0.59	1.75	0.845	0.686
C8	0.143	0.335	2.58	1.04	0.957

Process	n_items	KR20	r (between processes)	Ladder violations
C_RW _map	4	0.679	0.539	0.143
C_Map _RW	4	0.619	0.539	0.25

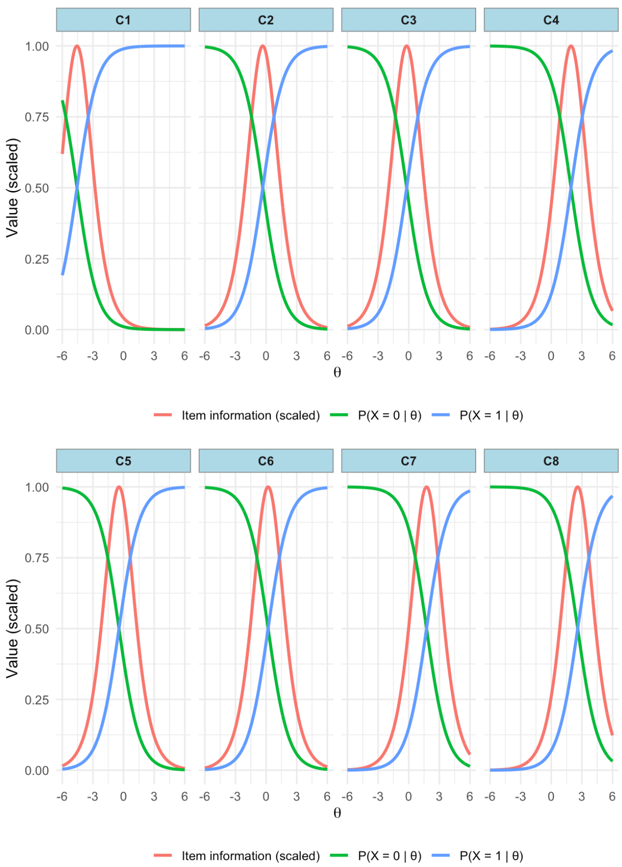

Figure 4: Item response and information curves for game C

When all tasks were combined, test information increased across the ability continuum (Fig. 5). The resulting conditional reliability curve showed higher reliability across a broader range of ability levels (approximately $\theta = -3$ to $+3$) than the curves for the individual games.

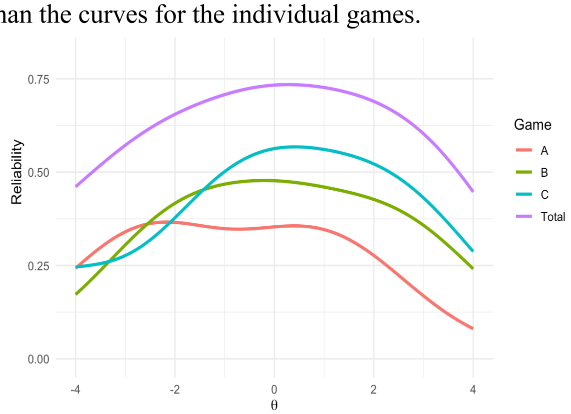

Figure 5: Conditional reliability across ability levels.

Discussion

The present pilot study examined a virtual reality–based assessment designed to capture spatial transformations between maps and environments in children from grades 3 and 4. The results show systematic variation in item difficulty and discrimination across task categories and transformation directions, indicating that the assessment differentiates between tasks of varying cognitive demands. Conditional reliability differed across games and ability levels, suggesting that measurement precision depends on both task type and participant ability.

Across task categories, aggregating items increased test information and resulted in higher conditional reliability across a broader range of ability levels. This pattern is consistent with the expectation that combining items with different difficulty profiles yields more stable measurement across the latent ability continuum. Rasch analyses further provided initial evidence of acceptable item fit, supporting the use of an IRT framework for modeling performance in this assessment.

Several limitations must be acknowledged. The number of items per task category was relatively small, which restricts the stability of reliability estimates and limits the precision with which differences within categories can be assessed. A more robust instrument will require a larger pool of tasks per category in future work in order to capture the intended constructs more comprehensively and to improve measurement precision across the ability continuum. At the same time, expanding the item pool creates a practical challenge, as assessment time is necessarily limited when testing children. One possible solution would be to distribute the assessment across multiple sessions, which is conceptually feasible given that it consists of six separate games corresponding to the six task categories. Another limitation is that due to time constraints not all participants completed all tasks, which may affect the stability of reliability estimates. Moreover, the present analyses focused on three task categories; additional categories included in the assessment framework remain to be evaluated. Future studies should therefore examine larger item pools, include all task categories, and assess measurement invariance across age groups.

Conclusion

The present findings provide initial evidence for the feasibility of a VR-based approach to assessing large-scale spatial transformation abilities in children. By integrating immersive navigation with map-based representations, the assessment captures key aspects of navigational map reading that are difficult to measure using traditional small-scale spatial tests. While the results are preliminary, they suggest that the proposed framework can support the systematic investigation of individual differences in large-scale spatial abilities. Further validation studies are needed to refine the assessment and to establish its reliability and validity across developmental stages and contexts.

References

Arthurs, L. A., Baumann, S. P., Rice, J. M., & Litton, S. D. (2023). *The development of individuals' map-reading skill: What research and theory tell us*. International Journal of Cartography, 9(1), 3–28.

Bartoschek, T., Bredel, H., & Forster, M. (2010). *GeospatialLearning@PrimarySchool: A minimal GIS approach.* In Proceedings of the Sixth International Conference on Geographic Information Science (Extended abstracts).

Blades, M., & Spencer, C. (1987). *Young children's strategies when using maps with landmarks.* Journal of Environmental Psychology, 7(3), 201–217.

Caduff, D., & Timpf, S. (2008). *On the assessment of landmark salience for human navigation*. Cognitive Processing, 9(4), 249–267.

Carbonell-Carrera, C., & Saorin, J. L. (2018). *Virtual learning environments to enhance spatial orientation*. Eurasia Journal of Mathematics, Science and Technology Education, 14(3), 709–719.

Coutrot, A., Schmidt, S., Coutrot, L., Pittman, J., Hong, L., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2019). *Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance*. PLOS ONE, 14(3).

Davies, C., & Uttal, D. H. (2007). *Map use and the development of spatial cognition.* In J. M. Plumert & J. P. Spencer (Eds.), The emerging spatial mind (pp. 219–247). Oxford University Press.

Dong, W., Qin, T., Yang, T., Liao, H., Liu, B., Meng, L., & Liu, Y. (2021). *Wayfinding behavior and spatial knowledge acquisition: Are they the same in virtual reality and in real-world environments*? Annals of the American Association of Geographers.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Educational Testing Service.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors.* Educational Testing Service.

Gallistel, C. R. (1990). *The organization of learning*. MIT Press.

Gärling, T., Säisä, J., Böök, A., & Lindberg, E. (1986). *The spatiotemporal sequencing of everyday activities in the large-scale environment.* Journal of Environmental Psychology, 6(4), 261–280.

Hartman, N., Connolly, P., Gilger, J., & Bertoline, G. (2006). *Developing A Virtual Reality Based Spatial Visualization Assessment Instrument.* 2006 Annual Conference & Exposition Proceedings.

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

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). *Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning.* Intelligence, 34(2), 151–176.

Hegarty, M., & Waller, D. (2004). *A dissociation between mental rotation and perspective-taking spatial abilities*. Intelligence, 32(2), 175–191.

Holmes, C. A., Newcombe, N. S., & Shipley, T. F. (2018). *Move to learn: Integrating spatial information from multiple viewpoints*. Cognition, 178, 7–25.

Ishikawa, T., & Newcombe, N. S. (2021). *Why spatial is special in education, learning, and everyday activities*. Cognitive Research: Principles and Implications, 6(1), Article 20.

Kozhevnikov, M., & Hegarty, M. (2001). *A dissociation between object manipulation spatial ability and spatial orientation ability*. Memory & Cognition, 29(5), 745–756.

Lasc, D., Grinshpun, S., Bixter, M. T., & Yang, Y. (2025). *Predicting large-scale spatial ability from small-scale spatial abilities in children: An application of the double-dimension framework*. Cognition, 254, 105982.

Liben, L. S. (1997). *Children's understanding of spatial representations of place*. In N. Foreman & R. Gillett (Eds.), Handbook of spatial research paradigms and methodologies (pp. 41–82). Psychology Press.

Liben, L. S., & Downs, R. M. (1989). *Understanding maps as symbols: The development of map concepts in children*. Advances in Child Development and Behavior, 22, 145–201.

Li, Y., Kong, F., Ji, M., Luo, Y., Lan, J., & You, X. (2019). *Shared and distinct neural bases of large- and small-scale spatial ability: A coordinate-based activation likelihood estimation meta-analysis*. Frontiers in Neuroscience, 12, 1021.

Lin, C.-H., Chen, C.-M., & Lou, Y.-C. (2014). *Developing spatial orientation and spatial memory with a treasure hunting game*. Educational Technology & Society*, 17*(3), 79–92.

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

Lobben, A. K. (2007). *Navigational map reading: Predicting performance and identifying relative influence of map-related abilities*. Annals of the Association of American Geographers, 97(1), 64–85.

May, M., Péruch, P., & Savoyant, A. (1995). *Navigating in a virtual environment with map-acquired knowledge: Encoding and alignment effects*. Ecological Psychology, 7(1), 21–36.

Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A., & Mora-Mora, H. (2018). *Virtual Reality Learning Activities for Multimedia Students to Enhance Spatial Ability*. Sustainability, 10(4), 1074.

Montello, D. R. (1993). *Scale and multiple psychologies of space*. In A. U. Frank & I. Campari (Eds.), Spatial information theory: A theoretical basis for GIS (pp. 312–321). Springer.

Montello, D. R. (2005). *Navigation*. In P. Shah & A. Miyake (Eds.), The Cambridge handbook of visuospatial thinking (pp. 257–294). Cambridge University Press.

Newcombe, N. S. (2016). *Thinking spatially in the science classroom*. Current Opinion in Behavioral Sciences, 10, 1–6.

Pastel, S., Bürger, D., Chen, C. H., Petri, K., & Witte, K. (2022). *Comparison of spatial orientation skill between real and virtual environment*. Virtual Reality, 26(1), 91–104.

Piaget, J., & Inhelder, B. (1948). *La représentation de l'espace chez l'enfant*. Presses Universitaires de France.

Presson, C. C. (1982). *The development of map-reading skills*. Child Development, 53(1), 196–199.

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Robitzsch, A., Kiefer, T., & Wu, M. (2025). *TAM: Test Analysis Modules* (Version 4.3-25). CRAN. https://CRAN.R-project.org/package=TAM

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). *Navigating buildings in desktop virtual environments: Experimental investigations using extended navigational experience*. Journal of Experimental Psychology: Applied, 3(2), 143–159.

Schäfer, G. (1984). *Die Entwicklung des geographischen Raumverständnisses im Grundschulalter: Ein Beitrag zur Curriculumdiskussion.* Geographiedidaktische Forschungen, (9), Dietrich Reimer Verlag.

Shimada, K., Hiroi, K., Kawaguchi, N., & Kaji, K. (2016). *Measurement methods of spatial ability using a virtual reality system*. In Proceedings of the International Conference on Mobile Computing and Ubiquitous Networking (pp. 1–6).

Thorndyke, P. W., & Hayes-Roth, B. (1982). *Differences in spatial knowledge acquired from maps and navigation*. Cognitive Psychology, 14(4), 560–589.

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

Vandenberg, S. G., & Kuse, A. R. (1978). *Mental rotations: A group test of three-dimensional spatial visualization*. Perceptual and Motor Skills, 47, 599–604.

von Stülpnagel, R., & Frankenstein, J. (2015). *Configurational salience of landmarks: An analysis of sketch maps using space syntax*. Cognitive Processing, 16(Suppl. 1), 437–441.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). *Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance*. Journal of Educational Psychology, 101(4), 817–835.

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). *Welcome to the tidyverse*. Journal of Open Source Software*, 4*(43), 1686.

Zhu, C., Leung, C. O.-Y., Lagoudaki, E., Velho, M., Segura-Caballero, N., Jolles, D., Duffy, G., Maresch, G., Pagkratidou, M., & Klapwijk, R. (2023). *Fostering spatial ability development in and for authentic STEM learning*. Frontiers in Education, 8, 1138607.