Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 28 Number 14, December 2023 ISSN 1531-7714

The Development and Validation of a Survey Measuring Opportunity to Learn Spatial Reasoning Skills at Home

Sarah Wellberg, *University of Virginia*
Anthony Sparks, *American Board of Obstetrics and Gynecology*
Leanne Ketterlin-Geller, *Southern Methodist University*

The early development of spatial reasoning skills has been linked to future success in mathematics (Wai, Lubinski, & Benbow, 2009), but research to date has mainly focused on the development of these skills within classroom settings rather than at home. The home environment is often the first place students are exposed to, and develop, early mathematics skills, including spatial reasoning (Blevins-Knabe, 2016; Hart, Ganley, & Purpura, 2016). The purpose of the current study is to develop a survey instrument to better understand Kindergarten through Grade 2 students' opportunities to learn spatial reasoning skills at home. Using an argument-based approach to validation (Kane, 2013), we collected multiple sources of validity evidence, including expert review of item wording and content and pilot data from 201 parent respondents. This manuscript outlines the interpretation/use argument that guides our validation study and presents evidence collected to evaluate the scoring inferences for using the survey to measure students' opportunities to learn spatial reasoning skills at home.

Keywords: mathematics education, home mathematics environment, spatial reasoning, survey development, survey validation, interpretation/use argument

Introduction

Spatial reasoning is often defined as the ability to interact with, navigate in, and understand one's environment (NRC, 2001; 2009), and it is predictive of future success in mathematics coursework and pursuit of STEM degrees and careers (Delgado & Prieto, 2004; Uttal & Cohen, 2012; Wai et al., 2009). Some students may play with games or manipulatives at home that improve their spatial reasoning skills (e.g., building blocks, puzzles, board games; Jirout & Newcombe, 2015), and teachers and researchers who have a better understanding of how spatial reasoning skills develop outside of school can help guide instruction and intervention practices. The goal of the current study is to detail the development and initial validation of a survey to better understand Kindergarten through Grade 2 students' exposure to and development of spatial reasoning in the home environment. This survey is titled "Survey of Children's Opportunities to Learn At-home Reasoning Skills" and is abbreviated as SCOLARS. This survey is intended to be completed by parents as they reflect on or directly observe their children engaging in specified spatial reasoning activities. As such, we use the phrase "at-home" to denote this environment. The survey could be completed by caregivers in other out-of-school environments (e.g., after-care programs); however, data reported in this manuscript were obtained from parents in reference to the home environment. We also consider exposure to and engagement with the activities described in the survey to provide opportunities for children to learn spatial reasoning skills. We, therefore, define "opportunity to learn" as students' exposure to activities that involve spatial reasoning and their development of spatial reasoning skills, as observed by their parents.

Even though spatial reasoning has been identified as a crucial indicator of future mathematics success, there is a surprisingly limited number of surveys designed to measure students' opportunities to learn spatial reasoning at home. Children typically establish a foundational understanding of mathematics through informal experiences at home or in out-of-school settings that include direct (e.g., counting) or indirect (e.g., using measurements during cooking) activities (Blevins-Knabe, 2016; Hart et al., 2016). Even after accounting for other explanatory variables, such as household income and parents' anxiety around mathematics, the richness of the home mathematics environment is a significant positive predictor of children's mathematics skills, as reported by their parents (Hart et al., 2016), and can predict early mathematics performance on standardized tests (Blevins-Knabe & Musun-Miller, 1996). In a study of the contribution of home-based activities on children's performance, mathematics activities were a stronger predictor of future performance than were reading activities (Huntsinger et al., 2016). Although these studies focus on mathematics in general, similar trends may exist for spatial reasoning, in particular (Mix & Cheng, 2012). As such, information from parent-supplied surveys, such as SCOLARS, may be useful to both teachers and researchers.

For teachers, understanding students' exposure to and development of spatial reasoning skills at home may help inform their instructional design decisions, such as identifying students' prior knowledge in pre-requisite skills (e.g., block play for understanding dimensionality) or selecting/designing activities to extend students emerging knowledge (e.g., move from reading maps to drawing maps). Moreover, because parents often place greater emphasis on literacy activities at home (Cannon & Ginsberg, 2008), teachers may share resources with parents to support their engagement in mathematics-related activities with their children. For researchers, having greater insights into students' exposure to and development of spatial reasoning at home may provide unique insights into the predictive relationship between the home mathematics environment and future outcomes. For these reasons, we developed the SCOLARS survey.

Theoretical framework

There are well-documented connections between spatial reasoning skills and mathematics performance (see Mix & Cheng, 2012). Rourke (1993), for example, found that students with low spatial reasoning abilities tend to struggle with parsing mathematical symbols, switching between mathematics procedures, and remembering arithmetic facts. Despite these connections, the emphasis on number sense – the ability to work with numbers flexibly (Gersten & Chard, 1999) – and spatial reasoning skills are not equitable in most classrooms. Zippert and Rittle-Johnson (2020) found that more instruction is centered around number sense compared to spatial reasoning. Moreover, students come to school with a wide range of mathematical skills and knowledge, which suggests differences in the home mathematics environment (Zippert & Rittle-Johnson, 2020). Consequently, the goal of our research is to develop a measure that may capture variability in students' opportunities to learn spatial reasoning at home. In the future, this survey instrument may be used by researchers to determine the relationship between students' early learning opportunities and future success as well as by teachers to guide instruction.

Spatial reasoning is commonly thought to include two main aspects: spatial visualization and spatial orientation (Bishop, 1980; NRC, 2009; Sarama & Clements, 2009a). *Spatial visualization* involves identifying and manipulating objects mentally, such as when one imagines the result of mentally rotating an object. *Spatial orientation* involves imagining objects or settings from other perspectives (e.g., seeing an object from a different viewpoint) and may include visual representations, such as maps. Combined, these spatial reasoning skills support students' overall mathematics proficiency, as well as their understanding of specific concepts like place value, relations between numbers (including the number line), and operations (Battista, 1990; Cheng & Mix, 2014; Newcombe, 2010; NRC, 2001).

As part of a larger research project, our research team articulated fine-grained descriptions of these two dimensions of spatial reasoning, which serve as the construct for classroom-based assessment resources for students in Kindergarten through Grade 2. Using an empirically-based and iterative process (see Perry et al., 2020), spatial visualization was articulated as three interconnected core concepts, each with four to ten subcomponents that specify the knowledge, skills, and abilities students develop as they increase their facility with mentally manipulating and transforming objects. These three core concepts include: (A1) understanding shape attributes and properties for two- and three-dimensional shapes and figures, (A2) understanding and applying transformations (rotation, translation, and reflection), and (A3) composing and decomposing shapes and figures. Spatial orientation was also divided into three core concepts with three to eight subcomponents each. The core concepts include: (B1) understanding and using spatial language to describe position and perspective, (B2) understanding maps and models as representations of objects and space, and (B3) recognizing, taking, and constructing perspectives from multiple viewpoints. (See Table 1 for examples of these core concepts.) The subcomponents within each core concept increase in complexity to reflect students' development of sophistication in knowledge, skills, and abilities across the grades. The core concepts and subcomponents were used as the basis for the SCOLARS instrument to understand students' opportunities to learn all aspects of spatial reasoning outside of the classroom.

Most of the previous studies that have investigated the impact of students' home mathematics environments have primarily focused on children before they enter Kindergarten (Ferrara et al., 2011; Ho et al., 2017; Levine, Ratliff, Huttenlocher, & Cannon, 2011; Purden & Levine, 2017; Verdine et al., 2014). Although this research provides an important context for students' opportunities to learn spatial reasoning prior to entering formal school settings, our research focuses on students in Kindergarten through Grade 2; as such, we examined research on children who have already begun their formal schooling. We identified two surveys that investigated the home mathematics environment for school-aged children (c.f., Hart et al., 2016; Jirout & Newcombe, 2015), but these surveys included both spatial reasoning and numeracy skills. Furthermore, the surveys did not cover all of the hypothesized aspects of spatial reasoning. Jirout and Newcombe (2015) examined parent-reported aspects of spatial play, including blocks, puzzles, and board games. They correlated these findings with subsequent

Table 1. Hypothesized aspects of spatial reasoning

Label	Description	Example
Spatial visualization		
A1	Knowledge of shape attributes and properties for two- and three-dimensional shapes and figures	Identifying a basketball as a sphere
A2	Knowledge and application of transformations (rotation, translation, and reflection)	Rotating a puzzle piece so that it fits in an open space
A3	Composition and decomposition of shapes and figures	Using two congruent triangular prism blocks to form a rectangular prism
Spatial orientation		
B1	Knowledge and use of spatial language to describe position and perspective	Describing an object's position in relation to another object (e.g., the sock is *under* the bed)
B2	Interpretation of maps and models as representations of objects and space	Using a map to figure out where to go next in a video game
B3	Recognition and construction of perspectives from multiple viewpoints	Understanding that if you and another person are facing each other, something that is on your right is on their left

spatial reasoning skills. Hart and colleagues (2016) included a sample of parents with children ages 3-8, and focused on both early numeracy skills and early spatial skills. Given this notable gap, we developed SCOLARS to better understand Kindergarten through Grade 2 students' opportunities to learn the full range of spatial reasoning skills at home. The remainder of this article describes the development and the initial validation of the scoring inferences for this survey.

Initial survey development

We used an iterative approach that aligned with the steps described in the *Test Standards* (AERA et al., 2014) to develop the SCOLARS instrument. Multiple sources of data were included as part of this development process, including theoretical evidence from extant literature, feedback from experts, and pilot test data. Initially, SCOLARS was developed to help explain nuances in student behavior we observed during cognitive interviews that were designed to measure Kindergarten through Grade 2 students' spatial reasoning. Two members of the research team developed an initial survey that consisted of 17 items that focused exclusively on the frequency with which students engaged in spatial reasoning activities (e.g., "about how often does your child play with blocks?"). The 4-point scale ranged from "never" to "almost daily." Items generally aligned with the core concepts of spatial visualization and spatial orientation, but the subcomponents were not systematically sampled. We also included some items relating to numeric relational reasoning. This initial survey was distributed to 55 parents of students in Kindergarten through Grade 2, 16 of whom also participated in the cognitive interviews. After initial analyses, which included descriptive statistics and factor analyses, we decided to incorporate SCOLARS more systematically into the overall research goals for the larger research project. To increase the relevance of the data from SCOLARS beyond the cognitive interviews, we needed to redesign the survey.

Three members of the research team redesigned SCOLARS with the goal of measuring students' exposure to and development of spatial reasoning skills at home. The aspects of spatial reasoning underlying the construct were defined by the core concepts and subcomponents of spatial visualization and spatial orientation we articulated for the larger project. We operationalized students' opportunities to learn as including two components: (1) exposure to spatial reasoning activities, which we evaluated using the frequency with which students engaged in various activities, and (2) development of spatial-reasoning skills, which we evaluated using the level of independence with which students could engage in various tasks.

The iterative development process began by removing the items that did not measure spatial reasoning. Next, we mapped the original spatial reasoning items to the subcomponents of spatial visualization and orientation that they represented and identified the gaps in content coverage that needed to be filled. Then, we conducted a review of current research and available survey items pertaining to spatial reasoning in order to identify exemplar items for measuring spatial reasoning skills that were inadequately represented in the survey. If we were unable to find existing survey items that addressed certain subcomponents, we used the findings of our literature review to write additional items meant to assess those subcomponents. We ended up with a total of 25 items that covered all the core concepts and associated subcomponents. The items were organized into sections that would allow parents to focus their attention on one aspect of home-based activities, such as playing games or using technology. (See Table A-1 in Appendix A for the full text and response options provided for each of the original items.)

Validation framework

Validity is defined as the trustworthiness and meaningfulness of the interpretations and the uses of test scores (AERA et al., 2014). Evidence to evaluate validity can be organized into a structured argument in which claims about the meaning of the scores are proposed and empirically tested (Kane, 2013). The claims can be organized as a series of cascading inferences that link observed performances with the intended interpretations and uses; Kane (2013) refers to this chain of inferences as the interpretation and use argument (IUA).

Kane (2013) specified three types of inferences: scoring, generalization, and extrapolation. *Scoring inferences* are most closely associated with observed performance and examine claims about the technical quality of the scores in relation to the intended construct and purpose. *Generalization inferences* help evaluate the connection between tested and untested content that underlie the intended construct. Often, these inferences examine whether the content that was sampled on the test would yield the same outcomes as would another sample of content or content that was sampled using another item format. Finally, *extrapolation inferences* extend beyond the intended construct to focus on broader claims associated with the intended interpretations and uses. For example, if scores on a test are intended to indicate whether students need additional instruction, an extrapolation inference might examine whether students identified as needing additional instruction do, in fact, benefit from additional instruction (Ketterlin-Geller, Perry, & Adams, 2019).

Each inference is tied to assumptions that can be evaluated based on the collected evidence. The *Test Standards* (AERA et al., 2014) organize sources of validity evidence into five categories: content, response processes, internal structure, relation to other variables, and consequences of testing. Briefly defined, evidence based on the test content examines the relationship between a test's content and the intended construct (e.g., content alignment). Evidence based on response processes examines whether the items are eliciting responses that align with the intended construct. Evidence based on internal structure looks at the association between the test items and the organization of the construct (e.g., dimensionality, reliability). To examine the relations with other variables, evidence is collected that allows comparisons between the test scores and external variables. Finally, evidence based on test consequences examines the intended and unintended outcomes associated with the decisions made using the test scores.

We applied the argument-based approach to collect initial validity evidence about the SCOLARS instrument. In order to fully evaluate the trustworthiness and meaningfulness of the scores generated by SCOLARS, we would need to examine all three types of inferences. For example, before recommending that teachers use the results of SCOLARS to guide instruction, it would be important to evaluate whether the scores are interpretable by teachers, provide useful information, and are sensitive enough to help teachers determine their next steps in instruction. Possible sources of evidence to evaluate these inferences could include interviews with teachers or classroom observations. However, before we can evaluate whether teachers can use the scores produced by this survey to inform their actions, initial validation of the scoring inferences are needed to examine the technical quality of the survey and to ensure that it appropriately measures students' opportunities to learn spatial reasoning skills at home. We, therefore, focus on scoring inferences in the present study. These inferences, and the sources of evidence we used to test them, are articulated in Table 2. We leave the evaluation of generalization and extrapolation inferences to future studies.

The remainder of this manuscript describes two studies conducted to gather initial validity evidence about the trustworthiness and meaningfulness of the scores produced by SCOLARS for measuring Kindergarten through Grade 2 students' opportunities to learn spatial reasoning at home. The studies contribute evidence to evaluate the inferences specified in Table 2 by addressing the following research questions:

Study 1: Expert review before the pilot survey was administered

1. Do the items that comprise the SCOLARS instrument adequately reflect students' exposure to and development of all aspects of spatial reasoning skills at home?
2. Do the items have appropriate response options?

Study 2: Analysis of pilot survey data

1. Do respondents use the full range of response options for each item?
2. Are there items that do not cohere well with the others and need to be removed?
3. Does SCOLARS measure the hypothesized dimensions of spatial reasoning?
4. Do the items fit the scoring model?
5. Does SCOLARS produce reliable scores?

Table 2. Scoring inferences for the SCOLARS instrument

Inference	Evidence	Assumption	Analysis
SCOLARS adequately measures students' exposure to and development of all aspects of spatial reasoning skills at home	Content	Theory and existing literature support the items on SCOLARS	Literature review
		Items align with the hypothesized core concepts for spatial visualization and spatial orientation	Expert feedback (Study 1)
The response options are appropriate for each item	Response processes	Parents will use the full range of response options	Expert feedback (Study 1)
			Descriptive statistics for response options selected (Study 2)
SCOLARS data fit the scoring model	Internal structure	Items cohere well together	Item discrimination analyses (Study 2)
		Items measure the hypothesized dimensions of spatial reasoning	Confirmatory factor analysis (Study 2)
		Items demonstrate acceptable fit with the scoring model	Item response theory model and item fit (Study 2)
		SCOLARS yields reliable results.	Cronbach's alpha (Study 2)

Study 1: Expert Review

Methods

This study was designed to collect evidence needed to examine two inferences: (1) SCOLARS adequately measures students' exposure to and development of all aspects of spatial reasoning skills at home, and (2) the response options are appropriate for each item.

Participants. Five mathematics-education researchers were recruited to provide expert feedback about the SCOLARS instrument. Four experts were professors at universities who specialize in early mathematics education with an emphasis on spatial reasoning and/or parental engagement in mathematics. The fifth expert was a research scientist at a non-profit organization who has conducted research in early mathematics learning contexts. Their years of experience working in mathematics education ranged from 12 to more than 30. All actively conduct research and publish in mathematics education journals.

Instrument. Reviewers received the survey and a form asking them to rate their level of agreement with the following statements using a four-point Likert scale from "strongly agree" to "strongly disagree":

1. The item is well aligned with the associated subcomponent(s)
2. The item is accessible and understandable
3. The item is free of bias
4. The item is culturally appropriate for potential respondents
5. The response options are appropriate for the item

We must note that we did not define any of these terms for the expert reviewers, so they evaluated the items based on what accessibility, bias, and cultural appropriateness meant to them. Finally, the experts were asked to consider the set of items for each core concept to determine whether it was well represented by the collection of items in the survey. Experts were asked to provide an explanation for any unsatisfactory ratings. Space was also provided to make suggestions for assessing any sub-components that they felt were inadequately represented on the survey.

Results

Overall, the expert reviewers agreed or strongly agreed that most items aligned with the stated subcomponents and were accessible and understandable, free from bias, and culturally appropriate. Table 3 presents a summary of the frequency of affirmative ("agree" or "strongly agree") responses across the 25 items by indicator. For example, for 22 items, 100% of the experts (5 out of 5) agreed or strongly agreed that the content was well aligned with the subcomponents of spatial visualization and spatial orientation that we had specified. For two items, 80% of the experts agreed or strongly agreed that the content was well aligned. For one item, 60% of the experts agreed or strongly agreed that the content was well aligned.

The modal value for all indicators except the appropriateness of the response options was 100%, indicating that for most items, all the experts agreed or strongly agreed with the statement. For two items, some concerns were raised about the accessibility and understandability of the language. Moreover, for five items, 60% of the experts agreed or strongly agreed that the item and/or examples were culturally appropriate, which the reviewers interpreted as whether students from all backgrounds were likely to experience the activity as described; as such, 40% disagreed or strongly disagreed that they were culturally appropriate.

The modal value for the appropriateness of the response options was 60%, indicating that for most items, only 60% of the experts (3 out of 5) agreed or strongly agreed that the response options were appropriate for the given prompt. The expert reviewers expressed concerns about the response options used to determine the frequency with which children engaged in spatial reasoning activities. The original response options included four choices: "almost never," "occasionally," "frequently," and "almost daily." Three experts noted that there should be an option for "never." Another expert suggested offering an option for "daily." Moreover, they noted that the distinction between "occasionally" and "frequently" may be unclear to some parents. For the items used to measure the level of independence with which students engaged in spatial reasoning activities, some concerns were raised about the original response options ("not yet," "sometimes," "frequently").

For the final question, which asked the experts to evaluate the content representation of the collection of items, all experts agreed or strongly agreed that the items adequately represented the six core concepts of spatial reasoning. However, written feedback suggested that two aspects of the core concepts could be better represented with additional items. The reviewers provided written feedback for indicators with which they did not agree or strongly agree. These data were summarized by item to be address as part of the iterative survey development process.

Table 3. Frequency of ratings of agree or strongly agree for all 25 items by indicator.

	Frequency of Rating for Given Percent			
Indicator	100%	80%	60%	40%
The item is well aligned with the associated subcomponent(s)	22	2	1	0
The item is accessible and understandable	13	5	5	2
The item is free of bias	21	3	1	0
The item culturally appropriate for potential respondents	16	4	5	0
The response options are appropriate for the item	0	8	17	0

Discussion

Results from the expert review contributed to the development and initial validation efforts for SCOLARS. The experts found most of the items to be well aligned with the targeted content, accessible and understandable, free of bias, and culturally appropriate. These data provide content-related validity evidence for the survey by indicating alignment with the assessed content and acceptability of the item format. In instances where the experts did not agree with a statement, we revised the item using the written feedback provided by the reviewers. These changes included revising the wording of items to make them more interpretable by parents or providing more examples that were culturally relevant. We added two items that had been suggested by the reviewers to enhance areas of the core concepts that were identified as being inadequately represented. Because we revised the items using the wordings suggested by the reviewers, we did not ask them to reevaluate the updated set of items.

Feedback regarding the response options was used to evaluate score-based inferences about the response processes. For all items, at least 60% (or 3 out of 5) of the experts agreed or strongly agreed that the response options were appropriate. However, because two experts disagreed or strongly disagreed for 17 items, we carefully considered the adequacy of the options. To address concerns raised by the expert reviewers, we revised the prompts and response options based on their feedback. For the frequency items, reviewers were concerned that parents might misunderstand or have different definitions for non-specific terms such as "occasionally." To address this feedback, we used specific time references to clarify and standardize the response options. The revised set of response options for frequency-related items included five categories: "never," "1-2 times/year," "1-2 times/month," "1-2 times/week," and "almost daily." To address concerns about the response options used to measure students' independence in doing spatial reasoning tasks, we changed the prompt to state, "Does your child do any of the following activities on their own?" The response options were changed to "not yet," "yes, with a lot of help," "yes, with a little bit of help," and "yes, without help." The revised set of items, along with the item stem and response options for each item group, are provided in Table A-1 in Appendix A. This table also specifies the shortened label we use to refer to each item throughout the manuscript. Table A-2 in Appendix A details the components of spatial visualization and orientation that each of the revised items represent.

The revised survey was translated into Spanish using a popular vendor. A university professor and expert in bilingual education whose research focuses on developing school-based partnerships with Latinx families conducted a review of the Spanish translation. Each item was reviewed to verify that the English and Spanish versions had the same meaning. Moreover, each item was evaluated for cultural bias. Thirty-six instances were identified in which the proposed translation changed the meaning of the prompt. In most cases, the concerns were related to the verb form and some verb conjugations. Three researchers, including two Spanish-speakers, reviewed the proposed changes and made all the suggested revisions prior to pilot testing the SCOLARS instrument.

Study 2: Pilot Study

We conducted a pilot study to collect additional validity evidence for SCOLARS. Data collected from the pilot study were used to further evaluate the adequacy of the response processes and the internal structure of the survey.

Methods

We designed the SCOLARS instrument to be administered via Qualtrics, which is an online survey delivery program. Prior to receiving the link for the full survey, we deployed an eligibility survey to verify that parents had a child in Kindergarten through Grade 2, and that the responses were not generated by computerized bots. This survey also served as our informed consent form, as it provided potential participants with information about the study and their rights. The survey included a "reCAPTCHA" verification against robots, an acknowledgement of informed consent, a question to determine whether parents had a child in Kindergarten through Grade 2, the country and state in which the respondent lived, the respondents' name, and two spaces in which to enter and verify an email address. (See Appendix B for the wording of the items included in the eligibility survey.) Respondents were considered eligible if both emails matched and the parent confirmed that they lived in the United States and had a child who would be in Kindergarten through Grade 2 during the 2020-2021 school year.

Participants. The eligibility survey was sent through the researchers' professional networks, an email distribution list with over 3,000 mathematics educators in a southern state, and via social media. We acknowledge that this method of distribution does bias the results towards respondents who have internet access. Over 3,000 respondents completed the eligibility survey. After removing ineligible participants based on the criteria described above, duplicate IP addresses, and duplicate email addresses, 752 respondents were determined to be eligible to take the full survey. Using this final list of eligible respondents, 250 people were randomly selected to complete the survey. The sample was limited to 250 due to budgetary constraints associated with the participation incentive. Participants had five days to respond to the survey, after which time another set of 70 participants was randomly selected. A total of 325 people received the survey link, and 273 responses were received.

After removing duplicate IP and email addresses, incomplete responses, potentially spurious responses, and IP addresses originating outside of the United States, we retained a total of 201 respondents, or 62% of the invited list. Table 4 describes the characteristics of the parents in the sample. The sample was evenly split by gender. Most parents had earned associates degrees (38%) or undergraduate degrees (39%). If a parent respondent had multiple children in Kindergarten through Grade 2, they were asked to complete the survey with only one child in mind. Tables 5 and 6 describe the children of the parents who completed the survey and their schooling experiences, respectively. Most of the children of the parents sampled were in first grade at the time of the survey (53%), were male (60%), and identified as white (71%). This sample included high percentages of students enrolled in bilingual education programs, in which students are taught using a combination of English and Spanish (80%), students with a 504/IEP (55%), and students who had been enrolled in a pre-K program (89%). We asked questions related to internet access due to the possible interaction between internet access and response to some digital experience items; most of our sample had access to the internet at home (70%).

Table 4. Demographics of parents sampled

Characteristic	Count (%)
Gender	
Male	102 (51%)
Female	99 (49%)
Level of Education	
Some high school	0 (0%)
High school diploma	10 (5%)
Vocational certification	18 (9%)
Associates degree	76 (38%)
Undergraduate degree	79 (39%)
Graduate degree	18 (9%)

Analytic approach. To evaluate the adequacy of our response options, we calculated the percentage of parents who selected each option for each item and collapsed lesser-used response options as needed. We then calculated Cronbach's alpha to get a sense of the internal consistency of our survey items and used the item discrimination values to determine if there were any items that did not cohere well with the rest of the items. We used the *alpha* function in the **`psych`** package for the R statistical computing environment (Revelle, 2018) to make these calculations.

Because we had specifically designed the items to represent either the visualization or orientation aspects of spatial reasoning, we conducted confirmatory factor analysis (CFA) via the **`lavaan`** package for R (Rosseel, 2012) to evaluate how well the items cohered within these two hypothesize components. We then used modification indices to improve the fit by removing items that **`lavaan`** indicated loaded heavily on both components.

Table 5. Characteristics of the children and their home life

Characteristic	Count (%)
Gender	
Male	121 (60%)
Female	79 (39%)
Prefer not to answer	1 (1%)
Race	
Asian	2 (1%)
American Indian/Alaskan Native	10 (5%)
Black/African American	35 (17%)
Native Hawaiian/Other Pacific Islander	3 (2%)
White	142 (71%)
Two or More Races	6 (3%)
Prefer not to answer	3 (2%)
Ethnicity	
Hispanic	63 (31%)
Non-Hispanic	132 (66%)
Prefer not to answer	6 (3%)
Home Languages	
English	198 (98%)
Spanish	3 (2%)
Internet Access	
Home	140 (70%)
School	42 (21%)
Community hotspot	10 (5%)
Cellular device	9 (5%)

To understand item parameters, we conducted a series of item response theory (IRT) models. Due to the ordinal nature of the responses, we used the **`mirt`** package for R (Chalmers, 2012) to fit both the generalized partial credit model (GPCM) and the graded response model (GRM). We conducted significance tests between the different models to ascertain which was most appropriate for the data. Once we had determined which model provided a better fit to the data, we analyzed the individual item fit statistics for the retained model using Chalmers' PV-Q1. If an item was flagged for significant item misfit $(p < 0.05)$, we removed the item and ran the model again until we arrived at a model in which all items were above the 0.05 threshold of significance. We then generated item characteristic surfaces and the test information surface for the retained model. These plots allowed us to understand the functional relationship between latent abilities and item features.

Results

Response processes evidence. Upon inspecting the response frequencies for each of the items, we decided to collapse some categories before proceeding with our analyses. The original frequencies for the independence and frequency item groups are presented in Tables B-1 and B-2, respectively, in Appendix B. We saw that the "never" option was selected very rarely, with fewer than 10 of the 201 parents making this selection for any given item. For this reason, we decided to collapse the "never" and "1-2 times/year" categories into an "almost never" group. Additionally, the parents had only been directed to answer the Digital Experiences items if they had previously indicated that their child had access to the associated technology. As a result, there are large percentages of missing data for these four items. As our interest is in the frequency with which children engage in these activities, we recoded these missing values as "almost never" on the assumption that a child who does not have access to a certain technology likely never uses it.

Table 6. Characteristics of the children's schooling experiences

Characteristic	Count (%)
Grade	
Kindergarten	43 (21%)
Frist	107 (53%)
Second	51 (25%)
School Type	
Public	132 (66%)
Private – Non-religious	24 (12%)
Private – Religious	38 (19%)
Charter	6 (3%)
Other: Homeschool	1 (1%)
Bilingual education	
Yes	160 (80%)
No	41 (20%)
Attended pre-K	
Yes	179 (89%)
No	21 (10%)
Prefer not to answer	1 (1%)
After-school program	
Yes	151 (75%)
No	50 (25%)
504/IEP	
Yes	110 (55%)
No	86 (43%)
Prefer not to answer	5 (3%)

The percentages of responses for each category after these adjustments were made are presented in Figures 1 and 2 for the independence and frequency items, respectively.

Internal structure evidence. In this section, we present the results of the internal structure evidence obtained via CTT, CFA, and IRT analyses.

Classical test theory (CTT). Once we had recoded our response options, we calculated Cronbach's alpha for the full set of items to get a sense of the internal consistency of our survey. The original set of 26 items had an alpha value of 0.88. We then examined the discrimination values of the items and iteratively removed the item with the lowest value until all items had a discrimination value of 0.3 or higher (see Table 7 for original and final values and Table B-3 in Appendix B for the results of each round of analyses). This process left us with a set of 22 items (12 visualization and 10 orientation) that had adequate discrimination. The Cronbach's alpha value for this filtered set of items was 0.89.

We then found the polychoric correlations between each item pair. We opted for a polychoric, rather than Pearson, correlation because our items are on an ordinal scale instead of an interval scale. We can see in Table 8 that most item pairs had moderate positive correlations. Three item pairs had correlations larger than 0.60, which has been suggested as a threshold for which items may be considered psychometrically synonymous (Meade & Craig, 2012). Two of these pairs had correlations of 0.62, which was close enough to the suggested threshold that we were not overly concerned. The third item pair, however, had a correlation of 0.66, which was large enough that we decided to watch for any possible issues that may arise with those two items as we proceeded with our analyses. We also see an item pair with small a negative correlation, though it was not statistically distinguishable from zero.

Figure 1. Item response distributions (independence items)

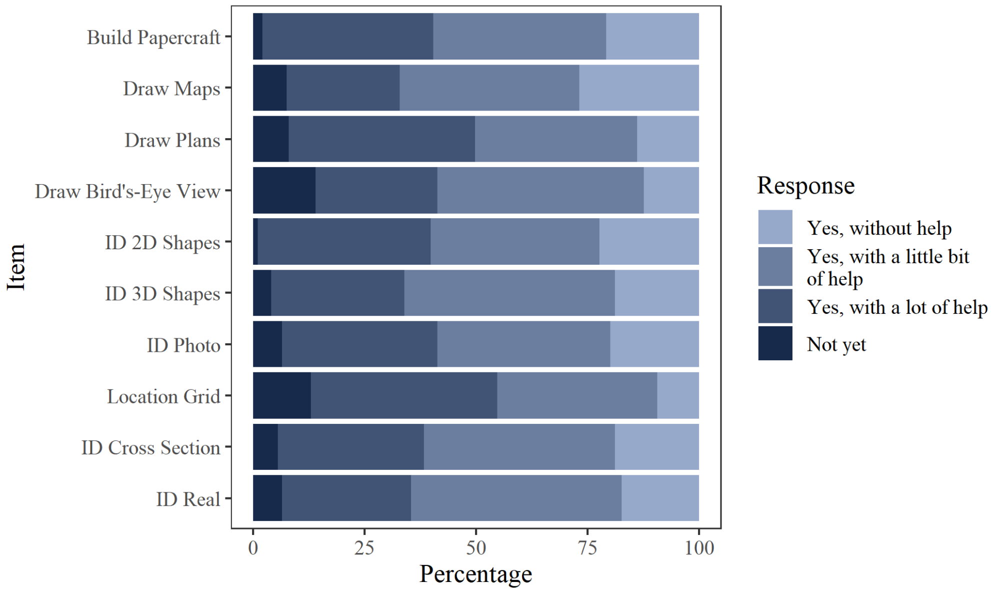


Figure 2. Item response distributions (frequency items)

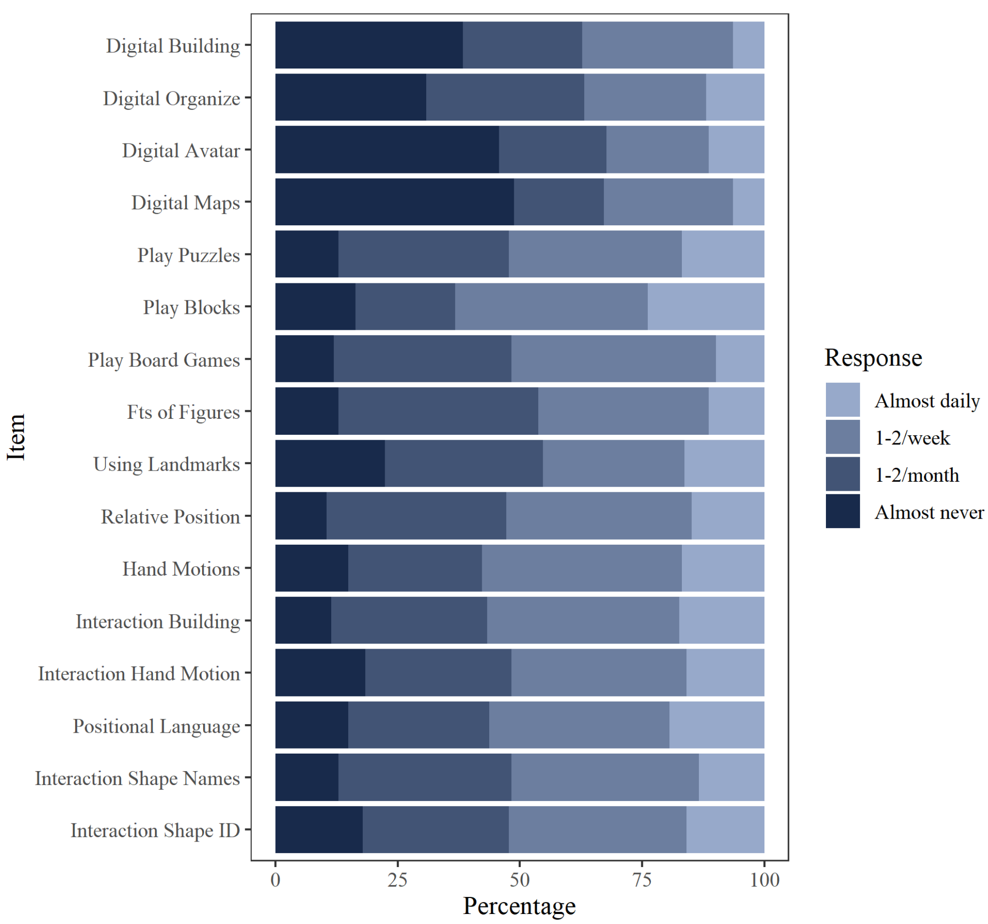


Table 7. Item discrimination values before and after removal of low-discrimination items

Item	Original	After removal
Draw Brid's-Eye	0.212	-
Location Grid	0.249	-
Digital Navigate	0.275	-
Draw Plans	0.281	-
Draw Maps	0.333	0.334
Digital Building	0.337	0.308
ID 3D Shape	0.367	0.369
Digital Avatar	0.387	0.359
ID Photos	0.402	0.363
ID Real	0.440	0.436
Positional Language	0.467	0.467
Digital Organize	0.475	0.516
Build Papercraft	0.476	0.476
Interaction Shape Names	0.494	0.504
Play Puzzles	0.495	0.529
Features of Figures	0.495	0.491
Play Board Games	0.502	0.513
ID Cross-sections	0.508	0.495
Play Blocks	0.518	0.551
Interaction Shape IDs	0.525	0.545
Hand Motions	0.530	0.545
Using Landmarks	0.547	0.562
Relative Position	0.564	0.560
Interaction Hand Motions	0.578	0.598
ID 2D Shape	0.586	0.585
Interaction Building	0.589	0.592

Confirmatory factor analysis (CFA). We used CFA to determine how well the items cohered within the hypothesized dimensions of spatial visualization and orientation. When we modeled the items associated with each dimension that remained after the CTT analyses, the fit statistics (CFI = 0.944, RMSEA = 0.092, SRMR = 0.098) were slightly outside of the conventionally-accepted ranges (Hu & Bentler, 1999). We used modification indices to get a sense of what could be causing misfit, and we found that the ID 2D Shapes, ID 3D Shapes, and ID Real items would be better suited in the spatial orientation component. Because these changes did not align with our theoretical framework, we removed the items. This alteration brought the fit to an acceptable level (CFI = 0.978, RMSEA = 0.059, SRMR = 0.078). (See Table B-4 in Appendix B for CFA factor loadings before and after modification.)

Item response theory (IRT). We conducted a series of multidimensional IRT models using the factor structure from the modified CFA. We used the multidimensional GPCM and GRM models for the 19 remaining survey items and compared the relative fit statistics to determine which model to use. Because the GPCM provided a slightly better fit (see Table B-5 in Appendix B) we proceeded with that model. Next, we investigated the item fit statistics to determine whether any items were not performing as expected. We used Chalmers' PV-Q1 statistic to identify misfitting items and iteratively removed the most misfitting item until all items had values greater than 0.05 (see Tables B-6

Table 8. Polychoric interitem correlations

		Build/Draw		ID					Digital			Play			Language				Interactions				
		Paper	Maps	2D	3D	Photos	Cross	Real	Build	Org	Avatar	Puzzles	Blocks	Board	Fts. Figs	Land	Rel. Pos.	Hands	Build	Hands	Pos. Lang.	Names	Shape IDs
B/D	Paper	1.00	0.26	0.51	0.28	0.20	0.37	0.23	0.24	0.33	0.31	0.34	0.24	0.30	0.29	0.27	0.20	0.17	0.41	0.30	0.22	0.28	0.36
	Maps		1.00	0.42	0.30	0.21	0.32	0.28	-0.15	0.10	0.20	0.21	0.28	0.09	0.17	0.40	0.29	0.19	0.29	0.33	0.20	0.24	0.08
ID	2D			1.00	0.28	0.49	0.57	0.59	0.26	0.27	0.16	0.30	0.35	0.32	0.36	0.47	0.46	0.35	0.39	0.50	0.31	0.26	0.32
	3D				1.00	0.20	0.44	0.35	0.07	0.11	0.18	0.04	0.25	0.24	0.04	0.41	0.24	0.37	0.25	0.43	0.29	0.15	0.17
	Photos					1.00	0.28	0.47	0.14	0.18	0.26	0.05	0.18	0.26	0.23	0.25	0.29	0.22	0.20	0.26	0.28	0.25	0.08
	Cross						1.00	0.42	0.31	0.26	0.13	0.19	0.24	0.29	0.23	0.33	0.40	0.37	0.30	0.40	0.32	0.23	0.31
	Real							1.00	0.13	0.18	0.08	0.09	0.21	0.32	0.17	0.42	0.28	0.38	0.20	0.30	0.53	0.18	0.22
Digital	Build								1.00	0.18	0.38	0.16	0.14	0.22	0.26	0.16	0.23	0.22	0.22	0.31	0.29	0.04	0.33
	Org									1.00	0.12	0.58	0.43	0.50	0.37	0.25	0.34	0.34	0.62	0.36	0.27	0.42	0.57
	Avatar										1.00	0.29	0.26	0.17	0.19	0.31	0.24	0.30	0.20	0.34	0.22	0.28	0.24
Play	Puzzles											1.00	0.51	0.50	0.52	0.35	0.40	0.27	0.62	0.31	0.24	0.38	0.47
	Blocks												1.00	0.30	0.38	0.49	0.36	0.36	0.50	0.53	0.22	0.54	0.35
	Board													1.00	0.35	0.36	0.37	0.38	0.53	0.31	0.33	0.29	0.42
Language	Fts. Figs														1.00	0.25	0.41	0.31	0.58	0.26	0.23	0.40	0.45
	Land															1.00	0.37	0.46	0.34	0.66	0.31	0.28	0.33
	Rel. Pos.																1.00	0.46	0.26	0.43	0.44	0.46	0.45
	Hands																	1.00	0.30	0.49	0.45	0.41	0.37
Interactions	Build																		1.00	0.25	0.29	0.36	0.54
	Hands																			1.00	0.25	0.41	0.34
	Pos. Lang.																				1.00	0.32	0.31
	Names																					1.00	0.40
	Shape IDs																						1.00

and B-7 in Appendix B for item fit details and item parameters, respectively). This led to the removal of two items (Relative Position and Using Landmarks), one of which was part of the item pair with a high correlation in Table 8 (Using Landmarks).

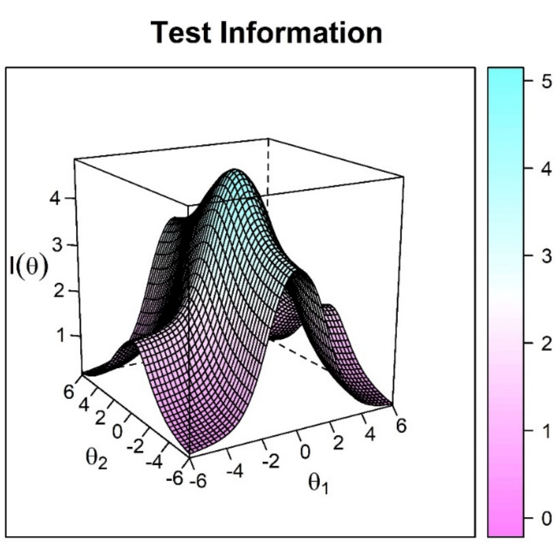
We noticed that more of the remaining items represented the spatial visualization subcomponents than represented the spatial orientation subcomponents, and we wanted to try to even out the representation. Several items associated with spatial visualization were coded as representing aspects A2 and A3 (see Table 1), so we thought removing one or more of these items would help balance the survey. Given that the Digital Build and Play Blocks items measure roughly the same type of activity (one digital and one hands-on), and the digital items had imputed data, we removed the Digital Build item to reduce the number of spatial visualization subcomponents represented. (See Table B-8 in Appendix B for a full list of items retained and removed and their associated subcomponent representations.) We checked the item fit statistics with this item removed and found them to be acceptable (see the Balance column in Table B-6 in Appendix B). We then checked the coherence of the remaining items using another round of CFA (factor loadings are shown in the Final Items column of Table B-4 in Appendix B), and found the fit to be improved (CFI = 0.985, RMSEA = 0.050, SRMR = 0.072). The reliability estimate was also still fairly high at 0.86.

Figure 3 provides the test information surface for the updated survey instrument. We can glean the most information from the survey at the intersection of the latent traits. Conversely, we glean less information on the ends of the latent trait spectrums, though the information does drop off more steeply for spatial visualization (θ_1) than for spatial orientation (θ_2). Figure 4 highlights two of the item characteristic surfaces from the survey[1]. For the Interaction Shape Names item, we notice that the thresholds are illustrated only for spatial visualization (θ_1), while the curve is flat along the spatial orientation (θ_2) dimension. The opposite holds for the other selected item (Play Board Games). This is due to the simple structure of the instrument.

Discussion

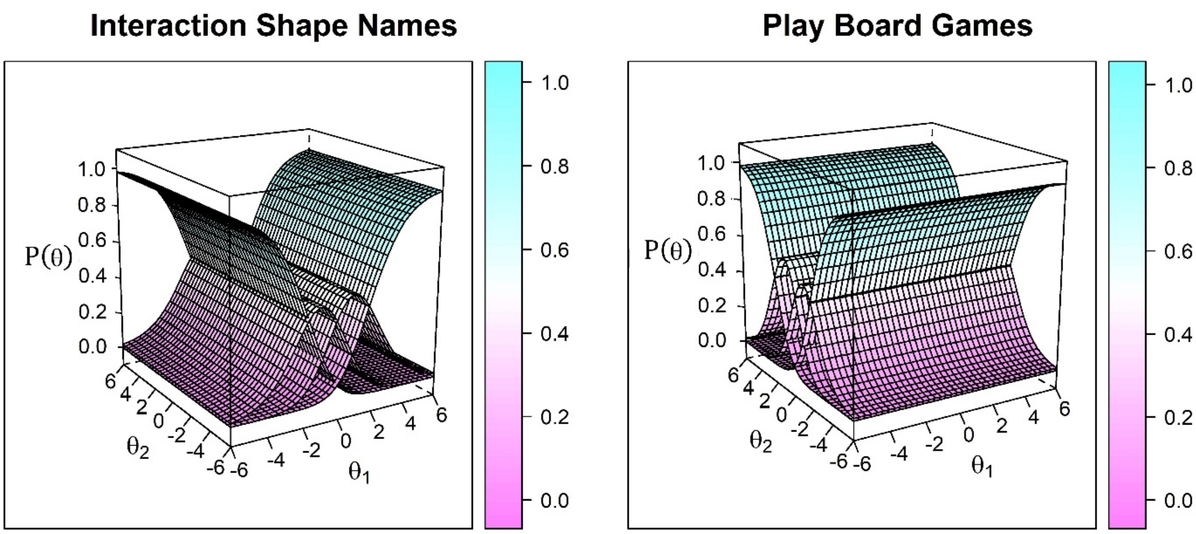
After collecting data from a pilot of the SCOLARS survey, we analyzed the results to examine inferences relating to the response processes and the internal structure of the survey. Evidence of response processes provides a better understanding of how parents interacted with the items. We found that they

Figure 3. Test information surface



[1] All item curve figures are available upon request.

Figure 4. Item characteristic surfaces for selected items



did not typically use the entirety of the five-point frequency scale, and adjustments were needed before further analysis. Internal structure validity evidence lets us understand the common factors measured by the items on the instrument. CFA and IRT analyses indicated that we could reasonably keep 17 of the 22 items that remained after the initial CTT analyses. We also removed an item for the sake of improving the balance between the spatial visualization and orientation subcomponents. The 16 remaining items maintained acceptable fit, as indicated by CFA model fit and GPCM item-fit statistics, and high internal consistency, as indicated by Cronbach's alpha. (See Appendix C for the item stems, item text, and response options for the final set of survey items.) These analyses helped us validate the scoring inferences for the survey we developed, but further research is still needed to investigate whether the survey is useful to teachers, parents, and students.

Discussion and Limitations

Students' spatial reasoning skills are not only shaped within the classroom but through their experiences at home. The more information we can gather about these experiences, the more effective we can make learning in the classroom. The purpose of the current study was to design an instrument to measure at-home spatial reasoning and to explore initial validity evidence to support the use of the survey to measure students' opportunities to learn spatial reasoning skills at home. We used Kane's (2013) validation framework to identify and to collect relevant validity evidence related to the scoring inferences we made about the survey. We found strong content-related evidence to support the inference that the survey measures the full range of at-home spatial reasoning skills. This was evidenced by the extensive literature review and rigorous expert review of the survey. While the experts had suggested expanding the frequency-related response options to a five-point Likert scale, we found that respondents did not typically use the entire range of options. This led us to collapse the "never" and 1-2 times/year options into an "almost never" option.

Challenges emerged throughout both the development and analysis phases. Some items that we originally mapped back to spatial visualization loaded with the items from the spatial orientation dimension and needed to be removed. This may be due to the items themselves or due to the background of the participants. These findings may indicate that children are not exposed to these skills as originally hypothesized. Further research is needed to understand the relationships among the skills measured by SCOLARS. Despite these challenges, we found promising evidence to support the scoring inference that SCOLARS data fit the scoring model after some items were removed. More work is needed, however, to further understand the derived scales.

Further analyses, such as differential item functioning, could highlight differences in item difficulty across demographic groups. These analyses can provide evidence of fairness across participants.

Limitations

Some limitations exist in the current study that may have impacted our analyses. First, our sample size was smaller than are those typically included in studies that use item response theory. Future work should include an independent sample to confirm the factor structure. A second limitation is the opt-out feature of a few items within the survey. This allowed participants to skip over certain questions if they stated previously that they did not have access to the requisite electronic devices or programs. In order to retain all 201 participants, we recoded missing values as indicating that students never engaged in those activities, which may not have given us an accurate account of their behavior. Future iterations of the survey should consider changes that allow parents to select the "almost never" response option rather than skipping survey questions.

Intended Use

Our hope is that the SCOLARS instrument can be used by researchers and teachers to better understand early elementary students' opportunities to learn spatial reasoning skills at home. Teachers may be able to use this knowledge to adjust classroom practices by identifying which aspects of spatial reasoning students are less familiar with and tailoring classroom activities to address any missing skills. Researchers could use the instrument to gauge at-home reasoning before an intervention. This tool could also be used to help researchers understand between-group disparities in students' access to spatial reasoning activities at home. While the analyses presented in the studies described above make us confident in the design and the scoring of our survey, further research still needs to be done to determine whether our intended use cases are viable.

Conclusion

Because of the importance and predictive nature of spatial reasoning for future mathematics and STEM success, data from surveys such as SCOLARS are needed to better understand the spatial reasoning skills students develop at home. The purpose of the current study was to develop a survey to measure students' exposure to and development of spatial reasoning skills at home and to collect initial validity evidence to support our scoring inferences. A clearer understanding of students' opportunities to learn spatial reasoning skills at home may provide both researchers and teachers with information that can be used to design instruction and interventions to support students' continued development of these skills. While we did find clear support for our scoring inferences, more work is needed to explore the generalization and extrapolation inferences that must hold for the survey to be used as intended.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Battista, M. T. (1990). Spatial visualization and gender differences in high school geometry. *Journal for Research in Mathematics*, *21*(1), 47-60.

Bishop, A. J. (1980). Spatial abilities and mathematics education: A review. *Educational Studies in Mathematics*, *11*(3), 257-269.

Blevins-Knabe, B. (2016). Early mathematical development: How the home environment matters. In B. Blevins-Knabe & A.M.B. Austin (eds.), *Early childhood mathematics skills development in the home environment* (pp7-28). Springer International: Switzerland.

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number use at home by children and their parents and its relationship to early mathematical performance. *Early Development and Parenting, 5*(1), 35-45.

Cannon, J., and Ginsburg, H.P. (2008). "Doing the math": Maternal beliefs about early mathematics versus language learning. *Early Education and Development, 19*(2), 238-260.

Chalmers, R. P. (2012). **mirt**: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*, 1–29.

Cheng, Y-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2-11.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25-32.

Ferrara, K., Hirsch-Pasek, K., Newcombe, N., Michnick Golinkoff, R., & Shallcross Lam, W. (2011). Block talk: Spatial language during block play. *Mind, Brain, and Education, 5*(3), 143-151.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education*, *33*(1), 18-28.

Hart, S. A., Ganley, C. M., & Purpura, D. J. (2016). Understanding the home math environment and its role in predicting parent report of children's math skills. *PLoS ONE, 11*(12), e0168227.

Ho, A., Lee, J., Wood, E., Kassies, S., & Heinbuck, C. (2017). Tap, swipe, and build: Parental spatial input during iPad® and toy play. *Infant and Child Development, 27*(1), e2061.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Huntsinger, C.S., Jose, P. E., & Luo, Z. (2016). Parental facilitation of early mathematics and reading skills and knowledge through encouragement of home-based activities. *Early Childhood Research Quarterly, 37*, 1-15.

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science, 26*(3), 302-310.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Ketterlin-Geller, L.R., Perry, L., & Adams, E. (2019). Integrating validation arguments with the assessment triangle: A framework for operationalizing and instantiating validation. *Applied Measurement in Education, 32*(1), 60-76. 10.1080/08957347.2018.1544136

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2011). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*(2), 530-542.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437.

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math. In J. B. Benson (Ed.), *Advances in child development and behavior* (Vol. 42) (pp. 197-243). New York, NY: Elsevier.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.) Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. C. T. Cross, T. A. Woods, & H. Schweingruber (Eds.). Committee on Early Childhood Mathematics, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Newcombe, N. S. (2010). Early education for spatial intelligence: Why, what, and how. *Mind, Brain, and Education*, *4*(3), 102-111.

Perry, L., Pinilla, R., Geller, J., Hatfield, C., & Ketterlin-Geller, L. R. (2020). *Spatial reasoning: Learning progressions development* (Technical Report No. 20-06) [Technical Report]. Southern Methodist University, Research in Mathematics Education.

Pruden, S. M., & Levine, S., C. (2017). Parents' spatial language mediates a sex difference in preschoolers' spatial-language use. *Psychological Science, 28*(11), 1583-1596.

Revelle, W. (2018). **psych**: Procedures for psychological, psychometric, and personality research. *R Package Version*, *1*(10).

Rosseel, Y. (2012). **lavaan**: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rourke, B. P. (1993). Arithmetic disabilities, specific and otherwise: A neuropsychological perspective. *Journal of Learning disabilities*, *26*(4), 214-226.

Sarama, J., & Clements, D. (2009a). Geometry and spatial thinking. In *Early childhood mathematics education research: Learning trajectories for young children* (pp. 247–269).

Sarama, J., & Clements, D. (2009b). Shape. In *Early childhood mathematics education research: Learning trajectories for young children* (pp. 199–246).

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (Vol. 4). Pearson Boston, MA.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? In B. Ross (Ed.), *Psychology of learning and motivation* (pp. 147-181). Oxford: Academic Press.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development, 85*(3), 1062-1076.

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.

Zippert, E. L., & Rittle-Johnson, B. (2020). The home math environment: More than numeracy. *Early Childhood Research Quarterly*, *50*, 4-15.

Citation:
Wellberg, S., Sparks, A., & Ketterlin Geller, L. (2023). The development and validation of a survey measuring opportunity to learn spatial reasoning skills at home." *Practical Assessment, Research, and Evaluation* 28(14). doi: https://doi.org/10.7275/pare.1885

Corresponding Author:
Sarah Wellberg
University of Virginia
Email: sarah.wellberg [at] virginia.edu

Appendix A: Study 1

Table A-1. Survey item stems, response options, and text before and after expert review

Label	Reviewed text	Updated text
Building/drawing items		
Section stem	Has your child done any of the following activities?	Does your child do any of the following activities on their own?
Response options	Not yet; Yes, with much assistance; Yes, with some assistance; Yes, independently	Not Yet; Yes, with a lot of help; Yes, with a little bit of help; Yes, without help
Build Papercraft	Papercraft (e.g., origami, paper snowflakes, paper airplanes, etc.)	Papercraft (e.g., paper airplanes, origami, paper snowflakes, papel picado, etc.)
Draw Maps	Drawing maps (e.g., treasure hunt, giving directions, while telling a story, etc.)	Draw maps (e.g., treasure map, giving directions, while telling a story, etc.)
Draw Plans	Drawing plans for buildings or spaces (e.g., treasure hunt, giving directions, while telling a story, etc.)	Drawn plans for buildings or spaces (e.g., the layout of your home, a fort, castle, classroom)
Draw Bird's-eye	Drawing a picture from a bird's-eye view (e.g., a picture of their school or neighborhood as seen from above)	Draw a picture from a bird's-eye view (e.g., a picture of their school or soccer field as seen from above)
Identification items		
Section stem	Have you noticed your child doing any of the following?	If prompted, does your child do any of the following?
Response options	Not yet; Yes, with much assistance; Yes, with some assistance; Yes, independently	Not yet; Yes, with a lot of help; Yes, with a little bit of help; Yes, without help
ID 2D Shapes	Recognize that shapes have the same name even when they are facing different ways or are different sizes (e.g., a triangle is still a triangle even if it is pointing down or to the side)	No change
ID 3D Shapes	Identifying that two or more objects are the same shape even if they have different sizes or orientations (e.g., an ice cream cone and a road pylon are both cones)	Identify that two or more objects are the same shape even if they have different sizes or orientations (e.g., an ice cream cone and a traffic cone are both cones)
ID Photos	Recognizing a photo of an object or a location taken from a different perspective (e.g., their neighborhood as seen from above)	Recognize a photo of an object or a location taken from a different point of view (e.g., their neighborhood or soccer field as seen from above)

Location Grid		**Added item:** Find his/her location on a map with a grid (e.g., directory at the mall, location on a hiking trail, at a theme park, etc.)
ID Cross-sections	Notice the shape of an object's flat face after it has been cut into parts or sliced in half (e.g., a stick of butter is rectangular, but when you cut a slice, it looks like a square)	Notice the shape of an object's flat face after it has been cut into parts or sliced in half (e.g., when you cut a lemon or an orange, the inside looks like a circle)
ID Real	Identifying real-world objects as shapes (e.g., the moon looks like a circle, the door is a rectangle, this ball is a sphere, etc.)	Associate or draw real-world objects as shapes (e.g., the moon looks like a circle, the door is a rectangle, this ball is a sphere, draw a circle to represent the sun, etc.)
Digital experiences items		
Section stem	About how often does your child use a computer/video game, app, or interactive website to do the following activities?	About how often does your child use a computer, video game, phone or tablet application to do the following activities?
Response options	Almost Never; Occasionally; Frequently; Almost Daily	Never; 1-2 times/year; 1-2 times/month; 1-2 times/week; Almost daily
Digital Building	Build things (e.g., play Minecraft, use LEGO building websites or apps, etc.)	Build things (e.g., play Minecraft, use building websites or apps, etc.)
Digital Organize	Organize or arrange shapes (on their own or in combination) to match or fit a space (e.g., play Tetris or Tangrams)	No change
Digital Avatar	Move a digital avatar through space (e.g., in Minecraft, Pokémon Go, etc.)	Move a digital avatar through space (e.g., Harry Potter: Wizards Unite, Pokémon Go, etc.)
Digital Navigate	Navigate through virtual spaces using a map (e.g., using a map in a video game to figure out where to go)	No change
Play items		
Section stem	About how often does your child play with the following items/toys?	No change
Response options	Almost Never; Occasionally; Frequently; Almost Daily	Never; 1-2 times/year; 1-2 times/month; 1-2 times/week; Almost daily
Play Puzzles	Puzzles	Jigsaw Puzzles
Play Blocks	Blocks	No change
Play Board Games	Board games in which they move a player through a route with other players (e.g., Chutes and Ladders, Candy Land, etc.)	No change

Language items		
Section stem	How often have you noticed your child doing the following?	No change
Response options	Not yet; Sometimes; Frequently	Never; 1-2 times/year; 1-2 times/month; 1-2 times/week; Almost daily
Features of Figures	Talking about or counting the number of edges, corners, or faces there are on a three-dimensional object	Describe the features of a figure (e.g., the side of a cube is a square)
Using Landmarks	Using landmarks to describe locations (e.g., the park by the school)	Using landmarks or specific places to describe locations (e.g., the park is by the school, the restaurant near the lake, etc.)
Relative Position	Describing an object's position relative to other objects (i.e., "the pencil *behind* the book," or "My friend sits *between* me and the teacher.")	No change
Hand Motions	Using hand motions (e.g., pointing, pantomiming, etc.) while they are describing an object's position (e.g., saying under the bridge while motioning going under something)	Using hand motions (e.g., pointing, gesturing, etc.) while they are describing an object's position (e.g., saying under the bridge while motioning going under something)
Interactions with your child items		
Section stem	About how often do you do the following with your child?	About how often do you (or someone in your household) do the following with your child?
Response options	Almost Never; Occasionally; Frequently; Almost Daily	Never; 1-2 times/year; 1-2 times/month; 1-2 times/week; Almost daily
Interaction Games	Play games in which objects can be seen by one person and not by the other (e.g., card games, Battleship, hide and seek, etc.)	Removed
Interaction Building	Build things by following a set of instructions (e.g., LEGO sets or DIY furniture)	Build things with your child by following a set of written, illustrated, or oral instructions (e.g., LEGO sets, science kits, etc.)
Interaction Hand Motions	Use hand motions or other movements when describing an object's position to your child (e.g., on top of, behind, on the right of, etc.)	Use hand motions or other movements when describing an object's position (e.g., on top of, behind, on the right of, etc.)
Positional Language	Ask your child to place or retrieve an object using positional language (e.g., put the book *on top of* the table, pick up the toy *in front of* the chair)	No change

Interaction Shape Names		**Added item:** Ask your child why a shape has a certain name (e.g., "How did you know that this is a triangle?")
Interaction Shape IDs	Ask your child to identify an object of a certain shape (e.g., play "I Spy" with shapes: You say "I spy with my little eye, an octagon", and the child finds the stop sign)	Ask your child to identify an object of a certain shape (e.g., you ask "what shape is that window?" and it is a rectangle or "can you find some circles in this room?")

Table A-2. Survey blueprint after expert review

	Visualization			Orientation		
Item	A1	A2	A3	B1	B2	B3
Build Papercraft		X				
Draw Maps					X	
Draw Plans					X	
Draw Bird's-Eye						X
ID 2D Shapes	X					
ID 3D Shapes	X					
ID Photos						X
Location Grid					X	
ID Cross-Sections			X			X
ID Real	X					
Digital Building		X	X			
Digital Organize		X	X			
Digital Avatar						X
Digital Navigate					X	X
Play Puzzles		X	X			
Play Blocks		X	X			
Play Board Games				X	X	X
Features of Figures	X					
Using Landmarks				X		
Relative Position				X		
Hand Motions				X		
Interaction Building		X	X	X		
Interaction Hand Motions				X		
Positional Language				X		
Interaction Shape Names	X					
Interaction Shape IDs	X					
Total	6	6	6	7	5	5

Appendix B: Study 2

Eligibility survey

The items below were sent to people who indicated interest in completing the SCOLARS pilot survey to assess their eligibility. The items that do not have Yes/No response options indicated provided respondents with a text box into which they typed their answers.

1. Would you like to participate in this research study? By clicking "Yes" below, you agree to participate in this research study.
   - Yes
   - No

2. Do you have a child who is entering Kindergarten through Second grade in the 2020-2021 school year?
   - Yes
   - No

3. In which country do you currently reside?
4. In which state do you currently reside?
5. Please enter your first and last name.
6. Please enter a valid email address below.
7. Confirm email address below.

Tables

Table B-1. Response Frequencies for Independence Items

	Not yet		Lots of help		Little help		No help		Total	
	N	%	N	%	N	%	N	%	N	%
Build/Draw										
Papercraft	4	2.0%	77	38.3%	78	38.8%	42	20.9%	201	100.0%
Maps	15	7.5%	51	25.4%	81	40.3%	54	26.9%	201	100.0%
Plans	16	8.0%	84	41.8%	73	36.3%	28	13.9%	201	100.0%
Bird's-Eye	28	13.9%	55	27.4%	93	46.3%	25	12.4%	201	100.0%
Identification										
2D Shapes	2	1.0%	78	38.8%	76	37.8%	45	22.4%	201	100.0%
3D Shapes	8	4.0%	60	29.9%	95	47.3%	38	18.9%	201	100.0%
Photos	13	6.5%	70	34.8%	78	38.8%	40	19.9%	201	100.0%
Location Grid	26	12.9%	84	41.8%	72	35.8%	19	9.5%	201	100.0%
Cross-sections	11	5.5%	66	32.8%	86	42.8%	38	18.9%	201	100.0%
Real	13	6.5%	58	28.9%	95	47.3%	35	17.4%	201	100.0%

Table B-2. Original Responses for Frequency Items

	Never		1-2/Year		1-2/Month		1-2/Week		Almost Daily		Missing		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Digital														
Build	1	0.5%	16	8.0%	49	24.4%	62	30.8%	13	6.5%	60	29.9%	201	100.0%
Organize	1	0.5%	27	13.4%	65	32.3%	50	24.9%	24	11.9%	34	16.9%	201	100.0%
Avatar	1	0.5%	11	5.5%	44	21.9%	42	20.9%	23	11.4%	80	39.8%	201	100.0%
Navigate	1	0.5%	20	10.0%	37	18.4%	53	26.4%	13	6.5%	77	38.3%	201	100.0%
Play														
Puzzles	1	0.5%	25	12.4%	70	34.8%	71	35.3%	34	16.9%	0	0.0%	201	100.0%
Blocks	2	1.0%	31	15.4%	41	20.4%	79	39.3%	48	23.9%	0	0.0%	201	100.0%
Board Games	5	2.5%	19	9.5%	73	36.3%	84	41.8%	20	10.0%	0	0.0%	201	100.0%
Language														
Features of Figures	4	2.0%	22	10.9%	82	40.8%	70	34.8%	23	11.4%	0	0.0%	201	100.0%
Using Landmarks	5	2.5%	40	19.9%	65	32.3%	58	28.9%	33	16.4%	0	0.0%	201	100.0%
Relative Position	4	2.0%	17	8.5%	74	36.8%	76	37.8%	30	14.9%	0	0.0%	201	100.0%
Hand Motions	1	0.5%	29	14.4%	55	27.4%	82	40.8%	34	16.9%	0	0.0%	201	100.0%
Interactions														
Building	1	0.5%	22	10.9%	64	31.8%	79	39.3%	35	17.4%	0	0.0%	201	100.0%
Hand Motions	3	1.5%	34	16.9%	60	29.9%	72	35.8%	32	15.9%	0	0.0%	201	100.0%
Positional Language	7	3.5%	23	11.4%	58	28.9%	74	36.8%	39	19.4%	0	0.0%	201	100.0%
Shape Names	7	3.5%	19	9.5%	71	35.3%	77	38.3%	27	13.4%	0	0.0%	201	100.0%
Shape ID	8	4.0%	28	13.9%	60	29.9%	73	36.3%	32	15.9%	0	0.0%	201	100.0%

Table B-3. Discrimination values

Item	Original	Round 1	Round 2	Round 3	Round 4
Draw Brid's-Eye	0.212	-	-	-	-
Location Grid	0.249	0.224	-	-	-
Digital Navigate	0.275	0.271	0.269	0.265	-
Draw Plans	0.281	0.262	0.249	-	-
Draw Maps	0.333	0.332	0.332	0.331	0.334
Digital Building	0.337	0.335	0.340	0.342	0.308
ID 3D Shape	0.367	0.368	0.366	0.364	0.369
Digital Avatar	0.387	0.386	0.391	0.384	0.359
ID Photos	0.402	0.399	0.392	0.369	0.363
ID Real	0.440	0.442	0.445	0.433	0.436
Positional Language	0.467	0.472	0.470	0.469	0.467
Digital Organize	0.475	0.477	0.483	0.501	0.516
Build Papercraft	0.476	0.481	0.483	0.475	0.476
Interaction Shape Names	0.494	0.501	0.492	0.496	0.504
Play Puzzles	0.495	0.502	0.502	0.514	0.529
Features of Figures	0.495	0.498	0.493	0.492	0.491
Play Board Games	0.502	0.503	0.505	0.508	0.513
ID Cross-sections	0.508	0.505	0.508	0.498	0.495
Play Blocks	0.518	0.530	0.535	0.546	0.551
Interaction Shape IDs	0.525	0.529	0.529	0.538	0.545
Hand Motions	0.530	0.530	0.536	0.539	0.545
Using Landmarks	0.547	0.558	0.561	0.562	0.562
Relative Position	0.564	0.564	0.556	0.551	0.560
Interaction Hand Motions	0.578	0.581	0.588	0.598	0.598
ID 2D Shape	0.586	0.587	0.593	0.583	0.585
Interaction Building	0.589	0.586	0.582	0.589	0.592

Table B-4. Confirmatory factor analysis item loadings

Item	Original items	Modification	Final items
Spatial visualization			
Build Papercraft	0.533***	0.510***	0.507***
ID 2D Shape	0.728***	-	-
ID 3D Shape	0.451***	-	-
ID Real	0.549***	-	-
Digital Building	0.352***	0.365***	-
Digital Organize	0.654***	0.711***	0.716***
Play Puzzles	0.651***	0.717***	0.723***
Play Blocks	0.654***	0.691***	0.697***
Features of Figures	0.590***	0.638***	0.637***
Interaction Building	0.730***	0.775***	0.779***
Interaction Shape Names	0.589***	0.635***	0.647***
Interaction Shape IDs	0.645***	0.690***	0.686***
Spatial orientation			
Draw Maps	0.430***	0.402***	0.418***
ID Photos	0.464***	0.393***	0.393***
ID Cross-sections	0.609***	0.553***	0.547***
Digital Avatar	0.407***	0.441***	0.426***
Play Board Games	0.614***	0.642***	0.643***
Using Landmarks	0.682***	0.688***	-
Relative Position	0.651***	0.676***	-
Hand Motions	0.623***	0.642***	0.644***
Interaction Hand Motions	0.716***	0.736***	0.735***
Positional Language	0.550***	0.542***	0.538***

Table B-5. Relative fit statistics for graded response and partial credit IRT models

	AIC	SABIC	HQ	BIC	logLik	X2	Df
Generalized partial credit	8972.8	8983.2	9075.7	9227.1	-4409.4	0.0	
Graded response	8973.5	8983.9	9076.4	9227.8	-4409.7	-0.7	0

Table B-6. Chalmers' PV-Q1 item fit statistics for the generalized partial credit model

Item	Round 1	Round 2	Round 3	Balance
Spatial visualization				
Build Papercraft	0.449	0.247	0.052	0.219
Digital Building	0.080	0.217	0.145	-
Digital Organize	0.762	0.495	0.342	0.071
Play Puzzles	0.668	0.387	0.208	0.217
Play Blocks	0.656	0.421	0.662	0.683
Features of Figures	0.301	0.751	0.107	0.318
Interaction Building	0.049	0.195	0.415	0.583
Interaction Shape Names	0.431	0.115	0.427	0.239
Interaction Shape IDs	0.766	0.252	0.075	0.065
Spatial orientation				
Draw Maps	0.043	0.025	0.365	0.572
ID Photos	0.300	0.386	0.433	0.496
ID Cross-sections	0.988	0.789	0.311	0.428
Digital Avatar	0.704	0.068	0.292	0.342
Play Board Games	0.295	0.332	0.391	0.067
Using Landmarks	0.145	0.012	-	-
Relative Position	0.014	-	-	-
Hand Motions	0.457	0.181	0.093	0.159
Interaction Hand Motions	0.134	0.092	0.145	0.175
Positional Language	0.543	0.108	0.383	0.743

Table B-7. Generalized partial credit model parameters for the final set of items

Item	Location	Threshold 1	Threshold 2	Threshold 3
Spatial visualization				
Build Papercraft	0.75	-4.64	-0.20	1.24
Digital Organize	1.26	-0.49	0.41	1.47
Play Puzzles	1.47	-1.46	-0.08	1.20
Play Blocks	0.95	-0.94	-0.85	0.99
Features of Figures	1.10	-1.70	0.16	1.76
Interaction Building	1.90	-1.45	-0.25	1.14
Interaction Shape Names	0.92	-1.72	-0.12	1.76
Interaction Shape IDs	1.12	-1.11	-0.20	1.40
Spatial orientation				
Draw Maps	0.44	-3.31	-1.20	1.17
ID Photos	0.51	-3.83	-0.32	1.65
ID Cross-sections	0.83	-2.92	-0.48	1.45
Digital Avatar	0.43	1.54	0.29	1.96
Play Board Games	1.01	-1.77	-0.15	2.11
Hand Motions	1.09	-1.27	-0.45	1.41
Interaction Hand Motions	0.96	-1.11	-0.20	1.47
Positional Language	0.69	-1.52	-0.42	1.39

Table B-8. Item status and survey blueprint after analysis of pilot data

		Visualization			Orientation		
Item	Status	A1	A2	A3	B1	B2	B3
Build Papercraft	Retained		O				
Draw Maps	Retained					O	
Draw Plans	Removed (CTT)					X	
Draw Bird's-Eye	Removed (CTT)						X
ID 2D Shapes	Removed (CFA)	X					
ID 3D Shapes	Removed (CFA)	X					
ID Photos	Retained						O
Location Grid	Removed (CTT)					X	
ID Cross-Sections	Retained			O			O
ID Real	Removed (CFA)	X					
Digital Building	Removed (Balance)		X	X			
Digital Organize	Retained		O	O			
Digital Avatar	Retained						O
Digital Navigate	Removed (CTT)					X	X
Play Puzzles	Retained		O	O			
Play Blocks	Retained		O	O			
Play Board Games	Retained				O	O	O
Features of Figures	Retained	O					
Using Landmarks	Removed (IRT)				X		
Relative Position	Removed (IRT)				X		
Hand Motions	Retained				O		
Int. Building	Retained		O	O	O		
Int. Hand Motions	Retained				O		
Positional Language	Retained				O		
Int. Shape Names	Retained	O					
Int. Shape IDs	Retained	O					
Total		3	5	5	5	2	4

Note. O = Item retained, X = Item removed

Appendix C: The final survey

Item stem: Does your child do any of the following activities on their own?
Response options: Not Yet; Yes, with a lot of help; Yes, with a little bit of help; Yes, without help

1. Papercraft (e.g., paper airplanes, origami, paper snowflakes, papel picado, etc.)
2. Draw maps (e.g., treasure map, giving directions, while telling a story, etc.)

Item stem: If prompted, does your child do any of the following?
Response options: Not yet; Yes, with a lot of help; Yes, with a little bit of help; Yes, without help

3. Recognize a photo of an object or a location taken from a different point of view (e.g., their neighborhood or soccer field as seen from above)
4. Notice the shape of an object's flat face after it has been cut into parts or sliced in half (e.g., when you cut a lemon or an orange, the inside looks like a circle)

Item stem: About how often does your child use a computer, video game, phone, or tablet application to do the following activities?
Response options: Almost never; 1-2 times/month; 1-2 times/week; Almost daily

5. Organize or arrange shapes (on their own or in combination) to match or fit a space (e.g., play Tetris or Tangrams)
6. Move a digital avatar through space (e.g., Harry Potter: Wizards Unite, Pokémon Go, etc.)

Item stem: About how often does your child play with the following items/toys?
Response options: Almost never; 1-2 times/month; 1-2 times/week; Almost daily

7. Jigsaw puzzles
8. Blocks
9. Board games in which they move a player through a route with other players (e.g., Chutes and Ladders, Candy Land, etc.)

Item stem: How often have you noticed your child doing the following?
Response options: Almost never; 1-2 times/month; 1-2 times/week; Almost daily

10. Describe the features of a figure (e.g., the side of a cube is a square)
11. Using hand motions (e.g., pointing, gesturing, etc.) while they are describing an object's position (e.g., saying under the bridge while motioning going under something)

Item stem: About how often do you (or someone in your household) do the following with your child?
Response options: Almost never; 1-2 times/month; 1-2 times/week; Almost daily

12. Build things with your child by following a set of written, illustrated, or oral instructions (e.g., LEGO sets, science kits, etc.)
13. Use hand motions or other movements when describing an object's position (e.g., on top of, behind, on the right of, etc.)
14. Ask your child to place or retrieve an object using positional language (e.g., put the book on top of the table, pick up the toy in front of the chair)
15. Ask your child why a shape has a certain name (e.g., "How did you know that this is a triangle?")
16. Ask your child to identify an object of a certain shape (e.g., you ask "what shape is that window?" and it is a rectangle or "can you find some circles in this room?")