

Reliability and Validity of a Spatial Ability Assessment Tool for 5- to 6-Year-Old Children: An Empirical Study of 112 Children

Xu Gehang, Lai Yihan, Mu Rongqi, Chen Xinhang, Chen Yuwei

Zhejiang Normal University, Jinhua, Zhejiang 321004, China.

DOI: 10.64649/yh.eg.issn3078-4794.20260415

Abstract: Objective: The present study evaluated the quality of individual items and obtained preliminary evidence on the reliability and validity of a spatial ability assessment tool for children aged 5–6 years. Methods: Liu's (2021) assessment tool served as the basis for the assessment. The participants were 112 children from four classes in one kindergarten. The 15 dichotomously scored items were grouped into spatial perception, mental rotation, and spatial visualization. Item characteristics, internal consistency, preliminary structural evidence, and associations between age and spatial ability were examined. Results: The items varied considerably in difficulty and discriminatory performance. Item 5 was answered correctly by 92.9% of the children and showed a relatively weak corrected item–total correlation, suggesting that its difficulty may need adjustment. Items 6 and 7 were difficult, but both retained some discriminatory capacity, so their low accuracy alone does not warrant deletion. KR–20 coefficients were 0.713 for spatial perception, 0.293 for mental rotation, 0.778 for spatial visualization, and 0.657 for the full assessment. For children aged 61–75 months, age was not significantly related to any dimension or to the total score. Conclusion: The assessment provided preliminary evidence of acceptable item performance and reliability for several components of spatial ability in 5- to 6-year-old children. However, the mental rotation dimension showed limited internal consistency, and Item 5 may require further revision. Further studies with larger and more diverse samples are needed to examine the stability of the proposed structure and to strengthen the validity evidence for the assessment.

Keywords: 5- to 6-year-old children; spatial ability; assessment tool; item analysis; reliability; validity

0. Introduction

Spatial ability refers to several related ways of processing spatial information: noticing relations between objects, keeping a mental representation in mind, and mentally comparing or transforming shapes. These processes overlap, but they are not interchangeable. A child who performs well on one type of spatial task may not perform equally well on another. Spatial thinking has also been associated with learning in mathematics and science, which makes it relevant to the study of early cognitive development^{[2][3]}.

These abilities are still developing during the preschool years. Children gradually become better at identifying spatial relations, imagining how a shape would look after being turned, and working with arrangements that are not directly visible. Mental rotation can be especially demanding because the child has to transform a representation mentally rather than rely only on what is in front of them. Studies have documented changes in mental rotation during the preschool period, including between three and five years of age^{[4][5]}. Research on spatial training also suggests that experience and structured activities may support the development of spatial skills^{[3][6][7]}.

Studying these abilities in young children depends in part on the quality of the assessment. The tasks need to reflect the skills they are intended to measure and to show differences in children's performance. Liu^[1] developed a spatial ability assessment for Chinese preschool children aged 5–6 and organized it around three dimensions: spatial perception, mental rotation, and spatial

visualization. The present study follows Liu's classification as the basis for grouping the items; it does not assume that this is the only possible way to classify spatial ability.

An assessment's performance may vary from one group of children to another. An item that is useful in one sample may be too easy or too difficult in another, and this can affect how well it distinguishes between children. The issue matters for items scored as correct or incorrect: when most children receive the same score on an item, the item provides less variation for examining discrimination and internal consistency. Item-level findings therefore need to be considered together with reliability estimates and evidence relevant to the intended structure of the assessment [8].

This study examined the item and measurement characteristics of Liu's [1] assessment using data from children aged 5–6 years. We examined how the 15 items performed, estimated internal consistency for each of the three dimensions and for the full assessment, and considered the available evidence concerning the proposed item grouping. We also tested whether age was associated with dimension scores or the total score among children with usable age information. The purpose was to identify features that may be retained and those that need further review, rather than to treat the current version as a fully validated instrument.

1. Methods

1.1 Participants

Data were collected from four classes in one kindergarten in Zhejiang Province, China. The initial group consisted of 120 children aged 5–6 years. Eight records were excluded before analysis because responses were incomplete or did not provide sufficient information for scoring, leaving 112 children in the final sample. The valid-response rate was 93.33%.

Birth dates were available for 105 children in the final sample. Their ages ranged from 61 to 75 months ($M = 66.60$, $SD = 3.29$). The other seven children were omitted from analyses involving age because their birth-date information was missing or could not be established with sufficient confidence.

Participation was voluntary. Each child completed the assessment individually in a quiet area of the kindergarten. The same general administration conditions were used throughout data collection.

1.2 Assessment Tool

The item set was based on the spatial ability assessment developed by Liu [1] in the dissertation *Development of an Assessment Tool for Spatial Ability in 5- to 6-Year-Old Preschool Children and Its Developmental Status*. Liu's tool distinguishes three dimensions of preschool spatial ability: spatial perception, mental rotation, and spatial visualization.

For this study, the assessment materials were organized into 15 items according to those dimensions. Items 1–5 were assigned to spatial perception, Items 6–9 to mental rotation, and Items 10–15 to spatial visualization. Each item was scored as correct (1) or incorrect (0), giving a possible total score of 0–15. The possible score ranges were 0–5 for spatial perception, 0–4 for mental rotation, and 0–6 for spatial visualization.

Item	dimension
1–5	spatial perception
6–9	mental rotation
10–15	spatial visualization

Responses were coded 1 for a correct answer and 0 for an incorrect answer. Dimension scores

were calculated by summing the relevant items, and the total score was the sum of all 15 items.

1.3 Procedure

The examiner assessed each child individually in a quiet area of the kindergarten, gave the task instructions, and recorded the child's responses. The recorded responses were then converted to binary scores for analysis.

Before analysis, records with incomplete answers or response patterns that could not be interpreted sufficiently for scoring were removed. The remaining data were examined item by item first; internal consistency and the evidence available concerning the score structure were then considered.

1.4 Data Analysis

Analyses were conducted in IBM SPSS Statistics. The analyses included descriptive summaries, item-level statistics, estimates of internal consistency, correlations among dimension scores, and associations between age and spatial ability scores.

For each item, we examined the proportion of correct responses, the discrimination index, and the corrected item-total correlation. The discrimination index compared the upper and lower 27% of participants. These indicators were considered together rather than using the correct-response rate alone to judge an item.

KR-20 was used as the primary internal-consistency coefficient because all items were scored dichotomously. Cronbach's alpha was also reported for reference; with binary-scored items, it produces the same coefficient as KR-20 when calculated from the same item scores.

KMO and Bartlett's test of sphericity were reported as preliminary checks on the item correlation matrix. These statistics were not treated as confirmation of a three-factor solution. Pearson correlations were used to describe associations among the three dimension scores and, for the 105 children with usable age information, between age and each dimension score and the total score. Statistical significance was evaluated at $\alpha = .05$.

2. Results

2.1 Descriptive Statistics

Descriptive statistics for the three dimensions and the total score are presented in Table 1.

Table 1. Descriptive Statistics of Spatial Ability Scores (N = 112)

Dimension	M	SD	Minimum	Maximum	Median
Spatial perception	2.53	1.55	0	5	2
Mental rotation	1.25	1.02	0	4	1
Spatial visualization	4.03	1.92	0	6	4
Total score	7.80	2.85	1	13	8

The mean scores were 2.53 for spatial perception, 1.25 for mental rotation, and 4.03 for spatial visualization. The mean total score was 7.80 (SD = 2.85). Relative to their respective maximum scores, children performed least well on mental rotation and best on spatial visualization.

2.2 Item Analysis

Table 2 summarizes the performance of each item, including the proportion of correct responses, discrimination index, and corrected item-total correlation.

Table 2. Item Difficulty and Discrimination Indices (N = 112)

Item	Proportion correct	Discrimination index	Corrected item – total correlation
Q1	.357	.300	.117
Q2	.366	.467	.257
Q3	.438	.500	.336
Q4	.438	.367	.208
Q5	.929	.100	.099
Q6	.205	.233	.138
Q7	.188	.367	.236
Q8	.384	.500	.176
Q9	.473	.400	.168
Q10	.741	.533	.368
Q11	.598	.767	.476
Q12	.598	.767	.453
Q13	.589	.767	.456
Q14	.750	.400	.240
Q15	.750	.500	.296

Item difficulty was uneven across the assessment. Item 5 was answered correctly by .929 of the children, making it the easiest item in this sample. Items 7 and 6 were the most difficult, with correct-response proportions of .188 and .205, respectively.

Discrimination was acceptable to strong for most items. Items 3, 8, and 10–13 were comparatively more discriminating, with Items 11–13 standing out: each had a discrimination index of .767, and their corrected item–total correlations were above .45.

Item 5 was the clearest exception. Although 92.9% of the children answered it correctly, its discrimination index was .100 and its corrected item–total correlation was only .099. The item therefore offered little information for separating children with different levels of spatial ability, a pattern consistent with a possible ceiling effect.

The low correct-response rates for Items 6 and 7 need to be considered alongside their discrimination indices. The values were .233 and .367, respectively, indicating that both items still differentiated among children to some extent. Difficulty alone, therefore, is not sufficient reason to regard either item as defective.

2.3 Internal Consistency

The internal consistency coefficients are shown in Table 3.

Table 3. Internal Consistency of the Assessment Tool

Dimension	Number of items	KR–20	Cronbach's α
Spatial perception	5	.713	.713
Mental rotation	4	.293	.293
Spatial visualization	6	.778	.778
Overall assessment tool	15	.657	.657

For the complete 15-item assessment, KR–20 was .657. The coefficients were higher for spatial perception (.713) and spatial visualization (.778), whereas mental rotation yielded a substantially lower value of .293.

The mental rotation result warrants some caution. This dimension contains only four items, and the items differ noticeably in difficulty. Both features can affect the stability of an internal-consistency estimate, so the coefficient should be interpreted in the context of the current item set

rather than taken as evidence that mental rotation is an unsuitable component of spatial ability.

2.4 Reliability after Item Deletion

To determine whether any single item had a marked effect on the reliability of the full assessment, KR-20 was recalculated after removing each item in turn. The resulting coefficients are shown in Table 4.

Table 4. KR-20 Coefficients after Deleting Each Item

Deleted item	KR-20 if item deleted
Q1	.664
Q2	.644
Q3	.632
Q4	.651
Q5	.659
Q6	.658
Q7	.646
Q8	.656
Q9	.657
Q10	.628
Q11	.609
Q12	.613
Q13	.612
Q14	.646
Q15	.638
All items	.657

Removing Item 5 raised the overall KR-20 only from .657 to .659. In practical terms, the change is negligible. Although Item 5 has weak item-level statistics, its deletion would not materially improve the reliability of the complete assessment.

The opposite pattern appeared for Items 10–13. Removing any of these items lowered the overall coefficient, with the reductions particularly evident for Items 11–13. Their contribution to the current measurement therefore appears meaningful.

Taken together, the deletion results do not provide a strong statistical case for dropping Item 5 at this stage. Revising the task itself, especially its level of difficulty, is a more defensible next step than removing it solely because of its item statistics.

2.5 Preliminary Evidence for Construct Validity

For the 15 items, the KMO value was .640, and Bartlett's test of sphericity was significant, $\chi^2(105) = 441.97$, $p < .001$. The correlation matrix was therefore sufficiently structured to permit an exploratory examination, although these statistics alone do not establish the proposed factor structure.

The associations among the three predefined dimensions were modest. Spatial perception correlated .041 with mental rotation and $-.050$ with spatial visualization, while the correlation between mental rotation and spatial visualization was approximately .296.

These results need to be read in light of the characteristics of the data. The sample contained 112 children, and all 15 indicators were dichotomous. A conventional Pearson correlation matrix may not represent relationships among binary indicators as well as methods based on tetrachoric correlations, particularly when item difficulties are uneven.

For that reason, the exploratory findings were not used to impose a fixed factor solution. Instead,

the observed correlations were considered together with the original theoretical framework and the item-level results. The data offer some preliminary support for viewing the assessment in three dimensions—spatial perception, mental rotation, and spatial visualization. However, the present findings are insufficient to determine whether the proposed grouping of items is appropriate.

2.6 Relationship between Age and Spatial Ability

Among the 105 children with valid age information, ages ranged from 61 to 75 months. Pearson correlations between age and spatial ability scores are presented in Table 5.

Table 5. Correlations between Age and Spatial Ability Scores (N = 105)

Variable	<i>r</i>	<i>p</i>
Spatial perception	.075	.448
Mental rotation	-.032	.747
Spatial visualization	.087	.377
Total score	.089	.366

None of the four correlations was statistically significant (all $p > .05$). Within the 61–75-month range represented in this sample, age was not significantly associated with spatial perception, mental rotation, spatial visualization, or the total spatial ability score.

This result should not be read as evidence that spatial ability stops developing during the preschool years. The participants occupied a fairly narrow age interval, from 61 to 75 months, and the cross-sectional design also limits what can be inferred about developmental change over time.

3. Discussion

3.1 Overall Quality of the Assessment Tool

The 15-item assessment showed a mixed psychometric profile. Several parts of the tool performed reasonably well, but the evidence is not strong enough to regard the present version as a fully established measure.

At the item level, Items 10–13 performed particularly well. Their discrimination indices and corrected item-total correlations were comparatively strong, and the deletion analysis also showed that removing these items would reduce the reliability of the full assessment. For the complete tool, KR-20 was .657; the corresponding coefficients for spatial perception and spatial visualization were .713 and .778.

The main weakness was concentrated in the mental rotation dimension, whose KR-20 was .293. The result points to a measurement issue that is more specific than a general failure of the assessment. Four items provide limited coverage of the construct, and the current tasks also span a broad range of difficulty. Both factors deserve attention when the dimension is revised.

3.2 Item 5: High Accuracy but Limited Discrimination

Item 5 stands out because almost every child answered it correctly: the proportion correct was .929. Its discrimination index (.100) and corrected item-total correlation (.099) were correspondingly low. In this sample, the task appears to have had little room to distinguish children who differed in overall spatial ability.

The most plausible interpretation is that the task was too easy for this particular group, producing a ceiling effect. That interpretation is preferable to concluding immediately that the item fails to represent spatial perception, because the statistical pattern concerns the difficulty of the task as administered rather than the relevance of the underlying construct.

The deletion analysis reinforces this point. Removing Item 5 changed KR-20 from .657 to .659,

a difference too small to justify discarding the item on reliability grounds alone. The item is better treated as a candidate for revision than as an item that should simply disappear from the assessment.

One possible revision would be to retain the same spatial-perception target while increasing the complexity of the spatial relationship that must be identified, or by reducing cues that allow the answer to be obtained directly from perception. Whether such changes improve discrimination would need to be tested with a new and more heterogeneous sample.

3.3 Mental Rotation: High Difficulty Does Not Necessarily Indicate Poor Item Quality

Mental rotation had the lowest mean score of the three dimensions. Items 6 and 7 were especially difficult, with correct-response rates of .205 and .188. The low rates are notable, but they do not tell the whole story.

Both items retained some ability to distinguish among participants: the discrimination indices were .233 for Item 6 and .367 for Item 7. Thus, a large proportion of incorrect responses does not automatically mean that an item is functioning poorly.

Difficulty and discrimination answer different questions. An item can be hard because the spatial transformation it requires is still challenging for children at this developmental stage, while at the same time providing useful information about individual differences. The results for Items 6 and 7 fit that possibility better than a simple interpretation of “too difficult”.

This interpretation is consistent with previous developmental evidence. Frick et al.^[4] reported substantial changes in mental rotation between 3 and 5 years, and their touchscreen study also identified developmental changes between approximately 3.5 and 5.5 years^[5]. Work with Chinese 5- to 6-year-olds has further indicated that mental-rotation training can improve spatial and mathematical performance^[6]. Recent research has likewise linked block play with mental-rotation development in children aged 5–6 years^[7].

The performance of Items 6 and 7 may therefore reflect, at least in part, the fact that mental rotation is still developing at this age. That possibility does not rule out item-design problems, however. Future revisions should preserve a meaningful spread of difficulty so that the assessment can distinguish children who are at different points along the emerging mental-rotation continuum.

3.4 Low Internal Consistency of the Mental Rotation Dimension

Across Items 6–9, discrimination varied, while the KR-20 for the dimension was only .293. This is the weakest reliability coefficient in the assessment and is the clearest psychometric issue identified in the current data.

There are at least two plausible reasons. The dimension contains only four items, which limits the amount of information available for estimating internal consistency. In addition, the tasks do not appear to place exactly the same demands on spatial transformation, and their difficulty levels differ. Such variation can lower the degree to which responses move together even when the items are related to the same broader ability.

The coefficient should consequently be interpreted as a limitation of the present operationalization rather than as evidence against mental rotation as a dimension of spatial ability. At present, four items do not appear to provide stable enough coverage of the construct.

A revised version would benefit from more mental-rotation items and a clearer sequence from simpler to more demanding transformations. Increasing both the number and range of tasks would give the dimension broader coverage and provide a stronger basis for estimating its reliability.

3.5 Correlations Among the Three Dimensions of Spatial Ability

The correlations between the three dimension scores were small overall. Spatial perception was almost unrelated to mental rotation ($r = .041$) and spatial visualization ($r = -.050$) in this sample. The association between mental rotation and spatial visualization was higher, at approximately $r = .296$, but remained modest. These results describe how the scores varied together in this group; by themselves, they do not show that the dimensions are independent or establish a particular factor structure.

The interpretation also needs to take the data into account. The study included 112 children, and each of the 15 items was scored as correct or incorrect. With a sample of this size, the dimension correlations should be treated as preliminary. In addition, analyses based on Pearson correlations may not capture relationships among binary indicators as well as methods using tetrachoric correlations.

The KMO value was .640, and Bartlett's test of sphericity was significant, $\chi^2(105) = 441.97$, $p < .001$. These results suggest that the item correlation matrix may warrant further examination, but they do not verify the correlations among the three dimensions. A larger study could test the proposed grouping using methods designed for categorical data, followed by confirmatory analysis in an independent sample.

3.6 Age and Spatial Ability

Age was not significantly related to any of the three dimension scores or to the total score in the present sample. Given the developmental nature of the construct, this null finding deserves a restrained interpretation rather than a broad developmental conclusion.

The children were between 61 and 75 months old, so the amount of age variation was relatively small. When participants are concentrated within such a narrow developmental interval, the correlation between age and performance can be difficult to detect. The cross-sectional design provides another reason for caution, since it cannot directly capture individual developmental trajectories.

Accordingly, the absence of a significant age association here does not mean that spatial ability is unrelated to development. It indicates only that, within this sample and age range, age did not account for a statistically detectable portion of the individual differences in spatial ability scores.

A broader age span and longitudinal follow-up would make it possible to examine developmental change more directly.

4. Conclusions and Limitations

4.1 Conclusions

The present analysis provides an initial psychometric picture of the 15-item spatial ability assessment for 5- to 6-year-old children. The evidence is encouraging in some respects, but it also identifies clear areas for refinement.

The three-dimensional framework—spatial perception, mental rotation, and spatial visualization—was retained throughout the analysis, and most items showed at least some discriminatory value. The results therefore provide a preliminary empirical basis for continuing to examine the assessment within this framework.

Item 5 is the main item-level concern. Its correct-response rate was .929, while both the discrimination index (.100) and corrected item-total correlation (.099) were low. Yet removing it raised the overall KR-20 only from .657 to .659. On the present evidence, revising the item is preferable to deleting it outright.

Items 6 and 7 were difficult, with correct-response rates below .21, but they still showed some discrimination. Their performance should consequently be considered in relation to the developmental demands of mental rotation rather than judged by accuracy alone.

The reliability problem was most evident for mental rotation, for which KR-20 was .293. Adding items and creating a clearer progression in transformation difficulty are reasonable priorities for the next version of the tool.

The analyses of relationships among the three assessed dimensions were inconclusive, and the available evidence was insufficient to determine whether the proposed grouping of items was supported by the data. In addition, no significant association between age and spatial ability was observed among the children aged 61–75 months.

4.2 Limitations

Several features of the study constrain how broadly the findings can be interpreted.

The participants came from a single kindergarten, and only 112 cases remained for the main analyses. The sample may not adequately represent children from other regions, socioeconomic backgrounds, or educational settings, so generalization should be made cautiously.

The data were collected at one point in time. As a result, the study can describe associations between age and performance but cannot establish developmental trajectories or causal effects of age on spatial ability.

All 15 items were scored dichotomously. KR-20 is appropriate for such data, but the structural analyses relied on conventional Pearson correlations, which may not capture latent relationships among categorical indicators as accurately as approaches based on tetrachoric correlations or other categorical-data methods. A larger sample would also be important for a more stable structural analysis.

The present study focused mainly on item functioning and internal structure. It did not directly examine criterion-related, convergent, or predictive validity. Subsequent work should compare the assessment with established spatial-ability measures and examine whether its scores relate to relevant cognitive or academic outcomes.

Taken as a whole, the findings argue for refinement of the current framework rather than its immediate replacement. The next stage of validation will require larger and more varied samples, stronger evidence from categorical-data factor analysis, and additional forms of validity evidence.

Fund Projec

This research is a result of the Zhejiang Provincial College Students' Innovation and Entrepreneurship Training Program (Project No.: S202610345009).

References

- [1] LIU Q X. *Development of an Assessment Tool for Spatial Ability in 5- to 6-Year-Old Preschool Children and Its Developmental Status* [D]. Changsha: Hunan Normal University, 2021. DOI: 10.27137/d.cnki.ghusu.2021.001308.
- [2] NEWCOMBE N S, SHIPLEY T F. Thinking about spatial thinking: New typology, new assessments[C]// GERO J S. *Studying Visual and Spatial Reasoning for Design Creativity*. Dordrecht: Springer, 2015: 179–192. DOI: 10.1007/978-94-017-9297-4_10.
- [3] UTTAL D H, MEADOW N G, TIPTON E, et al. The malleability of spatial skills: A meta-analysis of training studies[J]. *Psychological Bulletin*, 2013, 139(2): 352–402. DOI: 10.1037/a0028446.

- [4] FRICK A, HANSEN M A, NEWCOMBE N S. Development of mental rotation in 3- to 5-year-old children[J]. *Cognitive Development*, 2013, 28(4): 386-399. DOI: 10.1016/j.cogdev.2013.06.002.
- [5] FRICK A, FERRARA K, NEWCOMBE N S. Using a touch screen paradigm to assess the development of mental rotation between 3½ and 5½ years of age[J]. *Cognitive Processing*, 2013, 14(2): 117-127. DOI: 10.1007/s10339-012-0534-0.
- [6] KANG D, WEN X. Effects of mental rotation training on spatial ability and mathematical ability in 5- to 6-year-old children[J]. *Psychological Development and Education*, 2020, 36(1): 19-27. DOI: 10.16187/j.cnki.issn1001-4918.2020.01.03.
- [7] WEBER A, BOBROWICZ K, GREIFF S, LEUCHTER M. Mental rotation is supported by block play in boys and girls[J]. *Journal of Applied Developmental Psychology*, 2024, 91: 101630. DOI: 10.1016/j.appdev.2023.101630.
- [8] STREINER D L, NORMAN G R, CAIRNEY J. *Health Measurement Scales: A Practical Guide to Their Development and Use*[M]. 6th ed. Oxford: Oxford University Press, 2024.