2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51085

Work in Progress:Longitudinal Insights into the Reliability of the Tactile Mental Cutting Test: A Two-Year Study with Native American Elementary Students

Sujata Basnet, Utah State University

Sujata Basnet is a PhD student in Department of Engineering education at Utah State University. She received her master degree in Water Engineering from Asian Institute of Technology(AIT),Thailand. Her research interest lie in exploring spatial ability development and assessment, particularly through tools like the Tactile Mental Cutting Test (TMCT), and understanding its intersections with gender, socio-economic status, and underrepresented groups such as Native American children in STEM education.

Dr. Wade H Goodridge, Utah State University

Wade Goodridge is a tenured Associate Professor in the Department of Engineering Education at Utah State University. His research often has a focus on what drives us to learn, how we learn and how well we do it, and what can improve that learning. This work is often directed with a spatial thinking emphasis. Dr. Goodridge has investigated different learning modalities (fully online, hybrid, blended courses), cognitive and metacognitive learning processes (self-regulation, engineering problems solving, neural efficiency in engineering problem solving) learning motivators (engineering self-efficacy), and effective and personalized engineering education interventions (tactile engineering curriculum, physical teaching models, XR environments and AI informed systems of delivering and monitoring engineering learning). Dr. Goodridge has created a valid and reliable instrument to assess spatial ability in blind and low vision (BLV) learners thus opening new areas of research centered on tactile spatial ability. He has created tactile and visual learning tools and has offered professional development experiences on how to improve the education of tomorrows engineers.

Mr. Daniel Kane, Utah State University

Daniel Kane is a third-year Ph.D. student in the department of engineering education at Utah State University. His research interests include spatial ability, accessibility for students with disabilities, artificial intelligence in education, and enhancing electric vehicle charging system infrastructure. Daniel has contributed significantly to the development of the Tactile Mental Cutting Test (TMCT) which is a significant advancement in assessing spatial ability for blind and low-vision populations. His research has helped inform teaching methods and develop strategies for improving STEM education accessibility. Currently, he is studying how AI tools are utilized by students across USU's colleges to optimize their educational value. Daniel has also served as president of the ASEE student chapter at USU where he initiated outreach activities at local K-12 schools and promoted student engagement in research.

Work in Progress: Longitudinal Insights into the Reliability of the Tactile Mental Cutting Test: A Two-Year Study with Native American Elementary Students

Introduction

Spatial ability is a foundational cognitive resource that enables individuals to perceive, mentally manipulate, and reason about objects and spatial relationships in two- and three-dimensional contexts, and it is best understood as a set of related processes such as spatial visualization, spatial perception, and mental rotation that vary in complexity and cognitive demand [1], [2]. These skills support everyday tasks like interpreting diagrams and assembling objects and are especially important in STEM learning, where students must coordinate multiple representations and reason about 3D structures [3], [4]. In engineering education, spatial ability has been consistently associated with academic performance and persistence, and longitudinal and synthesis studies further show it predicts achievement and continued participation in STEM fields, particularly those requiring advanced visualization [5], [6], [7]. However, substantial variability exists across learners due to experience, instruction, and contextual factors, underscoring the need for reliable assessment and equitable opportunities to develop spatial skills early[2], [8].

Importantly, spatial ability is malleable, and meta-analytic evidence shows that spatial training can produce meaningful gains across tasks and age groups, with early, hands-on, and structured interventions improving children's spatial performance and sometimes supporting related outcomes such as mathematical reasoning [1][9], [10], [11]. Experimental work further suggests that these benefits can generalize beyond a single task, strengthening broader spatial cognition [12], which highlights the value of early identification and assessment to guide instructional support [13]. However, equitable measurement in early elementary settings remains challenging because many widely used spatial tests rely on visual, paper-based formats that may be developmentally inappropriate or inaccessible for some learners, were often designed for older populations, and assume familiarity with abstract visual conventions [14][15]. Additionally, learners may benefit differently from representational formats (e.g., 3D models versus 2D visuals), indicating that a single modality may not capture spatial reasoning equally well, especially for underrepresented populations, where accessibility and equity are central concerns [16].

Tactile and multimodal assessments can support more inclusive measurement of spatial ability by reducing barriers associated with purely visual tasks and providing learners with additional ways to demonstrate spatial reasoning [17]. Developed in this context, the Tactile Mental Cutting Test (TMCT) adapts the traditional Mental Cutting Test (MCT), which assesses reasoning about cross-sections of 3D objects [18], by converting key visual elements into tactile models and cutting-plane indicators while preserving the task's core cognitive demands [19]. The TMCT was initially developed to support accessibility for blind and low-vision populations and has been examined in prior work for its reliability and feasibility in tactile assessment contexts [19], [20]. More recently, the TMCT has been explored as a potential spatial assessment tool for broader populations, including early elementary learners. In our prior work-in-Progress study, we reported initial internal consistency reliability estimates for the TMCT when administered to

Native American elementary students, finding moderate reliability for both subtests and identifying the need for additional data across administrations and cohorts [21].

The prior work [21]was intended as an early step toward understanding whether the TMCT could serve as a reliable assessment tool in a population underrepresented in engineering education research and often underserved by traditional assessment approaches. Native American children remain underrepresented in psychometric research on early spatial assessment, and instruments developed or validated primarily in other populations may not exhibit the same measurement behavior when administered in different cultural and educational contexts. Establishing reliability and other validity-relevant evidence in the intended population is a necessary precursor to the responsible interpretation and use of test scores [22]. Moreover, evidence from one population does not automatically generalize to another, particularly when measurement invariance has not been examined [23]. Therefore, this study focuses on documenting internal consistency and score distribution characteristics of the TMCT in a sample of early elementary Native American students, as an initial step toward strengthening equitable measurement foundations rather than making between-group comparisons.

The present study extends that line of inquiry by examining the stability of TMCT internal consistency across two academic years. This longitudinal extension contributes to engineering education in two primary ways. By evaluating reliability across multiple administrations rather than relying on a single cohort snapshot, this work strengthens the evidence base for the TMCT. It contributes to broader efforts aimed at equitable assessment in STEM education. This study, which focuses on Native American elementary students, aligns with calls for culturally responsive and accessible measurement tools. It supports the potential of the TMCT as an early assessment instrument that can inform instructional design and spatial intervention efforts [1], [9], [13].

Because this study focuses on score reliability in early elementary students, it is important to note that internal consistency indices (e.g., Cronbach's alpha) are influenced by score variability and item difficulty [24], [25] and may be attenuated in young populations due to restricted score ranges or floor effects. In such contexts, lower alpha values do not necessarily indicate poor instrument functioning; rather, they may reflect developmental variability and limited differentiation among participants. Accordingly, reliability should be interpreted as evidence of score consistency within the specific sample and administration, and results should be considered alongside the cohort's age composition.

Methods

Instrument: Tactile Mental Cutting Test

The TMCT is adapted from the traditional Mental Cutting Test (MCT), originally developed in 1939 as a component of the College Entrance Examination Board (CEEB) Special Aptitude Test in Spatial Relations. The Mental Cutting Test (MCT) measures spatial visualization skills by asking participants to interpret two-dimensional isometric representations of three-dimensional objects intersected by cutting planes[18]. The TMCT converts key visual elements into tactile representations using three-dimensional models and tactile cutting-plane indicators [19]. The

TMCT is divided into two subtests, A and B, each with 12 questions and intended to function as parallel forms (i.e., comparable difficulty). However, form equivalence has not yet been established in early elementary cohorts. TMCT was administered in a controlled, organized setting to ensure accessibility, comfort, and equity for all participants, especially those with visual impairments. Participants took the Tactile Mental Cutting Test (TMCT) using 3D objects placed on a rotating turntable, so they could feel and examine each object easily without having to change position. They chose their answers on a tablet, selecting from five possible cross-section shapes for each item. A proctor read the instructions aloud and helped when needed. Unlike the original Mental Cutting Test, there was no time limit, which was intended to reduce pressure and support accurate responses. Testing was conducted in a quiet setting with minimal distractions, allowing students to concentrate.

Participants and Study Design

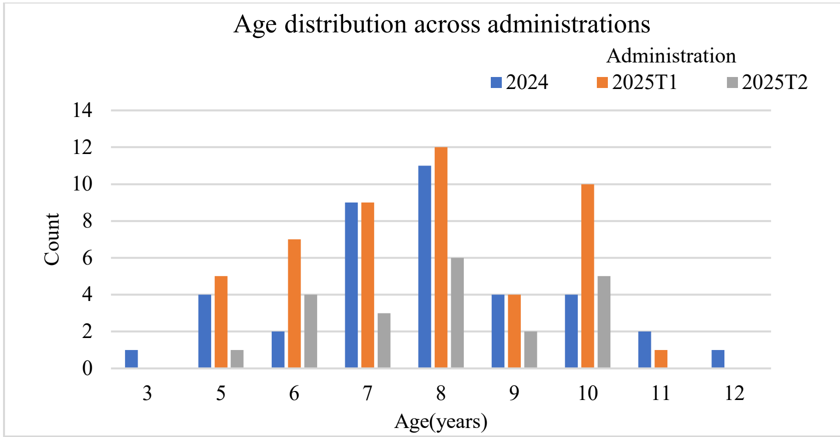
Participants were Native American elementary students recruited from several neighboring rural communities to evaluate TMCT score reliability in this setting; the study was not designed to make between-group comparisons or draw inferences about group differences in spatial ability. Data were collected across two academic years. In 2024, 38 Native American students aged 3 to 12 years completed the TMCT as part of an initial reliability investigation [21].In 2025, the TMCT was administered at two time points (T1 and T2), with some students participating in both administrations and others in only one.

Participant age characteristics and sample sizes for each administration are summarized in Table 1, and age distribution across administrations is shown in Figure 1. Across all administrations, participant ages ranged from 3 to 12 years. In 2024, students completing Subtest A (n = 23) had a mean age of 8.13 years (SD = 1.74; range = 5-12), while those completing Subtest B (n = 15) had a mean age of 7.27 years (SD = 2.02; range = 3-11). In 2025 Time 1, Subtest A participants (n = 14) had a mean age of 8.07 years (SD = 1.44; range = 6-10), and Subtest B participants (n = 34) had a mean age of 7.65 years (SD = 1.77; range = 5-11). In 2025 Time 2, Subtest A participants (n = 8) had a mean age of 8.00 years (SD = 2.00; range = 5-10), and Subtest B participants (n = 13) had a mean age of 7.85 years (SD = 1.34; range = 6-10). Most participants were approximately 5 to 10 years old, with a small number of younger and older participants.

Table 1: Participants' Age Characteristics by Subtest Administration

	Subtest	Sample Size(n)	Mean Age(years)	Standard Deviation	Age Range
2024	A	23	8.13	1.74	5-12
	B	15	7.27	2.02	3-11
2025T1	A	14	8.07	1.44	6-10
	B	34	7.65	1.77	5-11
2025T2	A	8	8.00	2.00	5-10
	B	13	7.85	1.34	6-10

Figure 1: Age distribution of participants across administrations (2024, 2025T1,2025T2)



Note: Bars represent the count of participants at each age(years).

Data Analysis

For each administration (2024, 2025 Time 1, and 2025 Time 2), descriptive statistics were computed separately for TMCT Subtests A and B, including the mean percent correct to summarize average performance and the standard deviation to characterize score variability across students; these summaries also supported interpretation of performance dispersion. To evaluate potential floor effects, we inspected total-score distributions for each subtest and administration to determine whether scores clustered near the minimum possible score [25]. Internal consistency reliability was then evaluated using Cronbach's alpha (α) for each subtest within each administration to examine whether items functioned coherently as a set and whether reliability remained stable across time points and classroom contexts. This analytic approach is consistent with prior TMCT reliability work and mirrors the methods used in the earlier Work-in-Progress study with this population [20], [21].

Results

Descriptive Statistics

Across administrations, TMCT performance showed a moderate mean score and substantial variability. In 2024, Subtest A had a mean score of 64.78% (SD=25.91%, n=23), while Subtest B had a mean of 44% (SD=23.84%, n=15) [21].In 2025, Time(T1) mean scores were 59.5%(SD=29.75%,n=14)for Subtest A and 48%(SD=23.83%,n=34) for Subtest B. Similarly for 2025 Time(T2) Subtest A had a mean score of 50% (SD=34.75%, n=8) while Subtest B had a mean of 57.66% (SD=25.08%, n=13). A summary table is provided below:

Table 2: Descriptive Statistics for Subtest A and B

	Subtest	Mean	Standard Deviation	Sample Size(n)
2024	A	64.78%	25.91%	23
	B	44.00%	23.84%	15
2025T1	A	59.50%	29.75%	14
	B	48.00%	23.83%	34
2025T2	A	50.00%	34.75%	8
	B	57.66%	25.08%	13

Reliability Analysis

Cronbach's alpha estimates for each administration are summarized in Table 3. For Form A, alpha increased from 0.66 (2024) to 0.87 (2025 Time 1) and 0.90 (2025 Time 2). For Form B, alpha increased from 0.60 (2024) to 0.71 (2025 Time 1) and 0.75 (2025 Time 2).

Table 3: Internal Consistency of TMCT for Subtest A and B

	Subtest	Cronbach's Alpha	Sample Size(n)
2024	A	0.66	23
	B	0.60	15
2025T1	A	0.87	14
	B	0.71	34
2025T2	A	0.90	8
	B	0.75	13

Discussion

Across administrations, TMCT scores showed moderate mean performance and substantial between-student variability (Table 2). Although a strong floor effect was not observed at the total-form level, some administrations and subtests showed low-end clustering, suggesting a restricted score range for portions of the sample. Because the administrations involved different groups with varying sample sizes and age compositions (Table 1; Figure 1), differences in mean scores across time points should not be interpreted as growth or decline. Instead, the descriptive patterns motivate careful evaluation of measurement behavior (e.g., internal consistency) within each administration.

Internal consistency estimates (Table 3) were generally higher in the 2025 administrations than in 2024 for both subtests. Importantly, Cronbach's alpha reflects within-sample score coherence and is sensitive to score variance and item difficulty [24], [25]; therefore, restricted range or low-end clustering in young samples can attenuate alpha even when an instrument is functioning as intended. Given the early elementary age range (Figure 1) and modest sample sizes in some administrations, reliability estimates should be interpreted cautiously as administration-specific evidence rather than definitive indicators of overall test quality. Moreover, internal consistency provides one component of reliability evidence and, by itself, does not establish broader validity. Future work should evaluate additional validity evidence (e.g.,item-level analyses) and include larger samples to examine form equivalence and strengthen reliability in Native American elementary settings.

Conclusion

This multi-year, multi-cohort study extends prior Work-in-Progress research [21] by demonstrating that the Tactile Mental Cutting Test exhibits stable and, in later administrations, strong internal consistency among Native American elementary student cohorts. Across administrations, the TMCT demonstrated acceptable to strong internal consistency, supporting its continued development as a tactile spatial assessment for early elementary contexts. At the same time, variability in score distributions across cohorts suggests that administration context and task familiarity may influence performance and observed reliability, supporting its continued development and use as an equitable spatial ability assessment in early STEM education.

Limitation and Future Work

Modest sample sizes limit this study at some administrations and by its focus on internal consistency rather than broader validity evidence. This study provides evidence of internal consistency reliability for TMCT scores in this context, but it does not establish broader validity evidence for the intended use and population. Future work should include item-level analyses (e.g., item-total correlations) and should build validity evidence (e.g., internal structure and relations to other variables), including construct- and criterion-related validity. Future studies with participant-level linkage across administrations should also evaluate test-retest reliability and developmental trajectories

References

[1] D. H. Uttal *et al.*, "The malleability of spatial skills: A meta-analysis of training studies.," *Psychol. Bull.*, vol. 139, no. 2, pp. 352–402, Mar. 2013, doi: 10.1037/a0028446.

[2] M. C. Linn and A. C. Petersen, "Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis," *1985*, vol. 56, no. 6, pp. 1479–1498, 1985, doi: https://doi.org/10.2307/1130467.

[3] C. A. Cohen and M. Hegarty, "Inferring cross sections of 3D objects: A new spatial thinking test," *Learn. Individ. Differ.*, vol. 22, no. 6, pp. 868–874, Dec. 2012, doi: 10.1016/j.lindif.2012.05.007.

[4] N. Newcombe and T. Shipley, "Studying Visual and Spatial Reasoning for Design Creativity. Studying Visual and Spatial Reasoning for Design Creativity, November," 2015.

[5] J. Wai, D. Lubinski, and C. P. Benbow, "Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance.," *J. Educ. Psychol.*, vol. 101, no. 4, pp. 817–835, Nov. 2009, doi: 10.1037/a0016127.

[6] D. L. Shea, D. Lubinski, and C. P. Benbow, "Importance of Assessing Spatial Ability in Intellectually Talented Young Adolescents: A 20-Year Longitudinal Study," *2001*, vol. 93, no. 3, pp. 604–614, 2001, doi: http://dx.doi.org/10.1037/0022-0663.93.3.604.

[7] D. Lubinski, "Spatial ability and STEM: A sleeping giant for talent identification and development," *2010*, vol. 49, no. 4, pp. 344–351, 2010, doi: http://dx.doi.org/10.1016/j.paid.2010.03.022.

[8] D. Voyer, S. Voyer, and M. P. Bryden, "Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables," *1995*, vol. 117, no. 2, pp. 250–270, 1995, doi: http://dx.doi.org/10.1037/0033-2909.117.2.250.

[9] W. Yang, H. Liu, N. Chen, P. Xu, and X. Lin, "Is Early Spatial Skills Training Effective? A Meta-Analysis," *Front. Psychol.*, vol. 11, p. 1938, Aug. 2020, doi: 10.3389/fpsyg.2020.01938.

[10] Y.-L. Cheng and K. S. Mix, "Spatial Training Improves Children's Mathematics Ability," *J. Cogn. Dev.*, vol. 15, no. 1, pp. 2–11, Jan. 2014, doi: 10.1080/15248372.2012.725186.

[11] Z. Hawes, J. Moss, B. Caswell, and D. Poliszczuk, "Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study," *Trends Neurosci. Educ.*, vol. 4, no. 3, pp. 60–68, Sep. 2015, doi: 10.1016/j.tine.2015.05.001.

[12] R. Wright, W. L. Thompson, G. Ganis, N. S. Newcombe, and S. M. Kosslyn, "Training generalized spatial skills," *Psychon. Bull. Rev.*, vol. 15, no. 4, pp. 763–771, Aug. 2008, doi: 10.3758/PBR.15.4.763.

[13] M. Stieff and D. Uttal, "How Much Can Spatial Training Improve STEM Achievement?," *Educ. Psychol. Rev.*, vol. 27, no. 4, pp. 607–615, Dec. 2015, doi: 10.1007/s10648-015-9304-8.

[14] J. W. French, R. B. Ekstrom, and L. A. Price, "MANUAL FOR KIT OF REFERENCE TESTS FOR COGNITIVE FACTORS (REVISED 1963):," Defense Technical Information Center, Fort Belvoir, VA, Jun. 1963. doi: 10.21236/AD0410915.

[15] G. M. Bodner and R. B. Guay, "The Purdue Visualization of Rotations Test," *Chem. Educ.*, vol. 2, no. 4, pp. 1–17, Oct. 1997, doi: 10.1007/s00897970138a.

[16] T. Huk, "Who benefits from learning with 3D models? the case of spatial ability," *J. Comput. Assist. Learn.*, vol. 22, no. 6, pp. 392–404, Dec. 2006, doi: 10.1111/j.1365-2729.2006.00180.x.

[17] M. B. Casey, E. Pezaris, and R. L. Nuttall, "Spatial ability as a predictor of math achievement: The importance of sex and handedness patterns," *Neuropsychologia*, vol. 30, no. 1, pp. 35–45, Jan. 1992, doi: 10.1016/0028-3932(92)90012-B.

[18] D. Kane, N. Shaheen, and W. Goodridge, "An Analysis of Low-Scoring Blind and Low-Vision Individuals' Selected Answers on a Tactile Spatial Ability Instrument," in *2023 ASEE Annual Conference & Exposition Proceedings*, Baltimore , Maryland: ASEE Conferences, Jun. 2023, p. 42603. doi: 10.18260/1-2--42603.

[19] W. Goodridge, N. Shaheen, A. Hunt, and D. Kane, "Work in Progress: The Development of a Tactile Spatial Ability Instrument for Assessing Spatial Ability in Blind and Low-vision Populations," in *2021 ASEE Virtual Annual Conference Content Access Proceedings*, Virtual Conference: ASEE Conferences, Jul. 2021, p. 38203. doi: 10.18260/1-2--38203.

[20] C. Hamilton, E. Stratman, D. Kane, J. Blonquist, N. Shaheen, and W. Goodridge, "Parallel Form Reliability Analysis of a Tactile Mental Cutting Test for Assessing Spatial Ability in Blind and Low-vision Populations," in *2023 ASEE Annual Conference & Exposition Proceedings*, Baltimore , Maryland: ASEE Conferences, Jun. 2023, p. 43855. doi: 10.18260/1-2--43855.

[21] S. Basnet, W. Goodridge, and D. Kane, "Work in Progress: Exploring Reliability of the Tactile Mental Cutting Test in Assessing Spatial Abilities Among Native American Children," in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 57493. doi: 10.18260/1-2--57493.

[22] N. J. Dorans and L. L. Cook, *Fairness in Educational Assessment and Measurement*, 1st ed. New York: Routledge, 2016. doi: 10.4324/9781315774527.

[23] R. E. Millsap, *Statistical approaches to measurement invariance*. New York: Routledge, 2011.

[24] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, Sep. 1951, doi: 10.1007/BF02310555.

[25] D. L. Streiner, G. R. Norman, and J. Cairney, *Health Measurement Scales: A practical guide to their development and use*. Oxford University Press, 2014. doi: 10.1093/med/9780199685219.001.0001.