Psychometric Validation And Structural Analysis Of A Gamified Mobile Cognitive Assessment For Adolescents: The Skilme Craft-The-Block Task

Dr. Ameer Hassan[1], Shyam Kumar[2], Dr. Roshan Anie Alex[2], Amal Tomy[2]

[1]Assistant Professor and Head of the Department, Amal College of Advanced Studies (Autonomous), Nilambur, Kerala
[2]Research & Development Department, Master n Masters, Kozhikode, Kerala, India
*Corresponding author: *Shyam Kumar, Research & Development Department, Master n Masters, Kozhikode, Kerala, India, Email: shyamkumar17a@gmail.com

Abstract

The digitisation of cognitive assessment has catalysed the development of gamified tools designed to improve engagement, scalability, and ecological validity. However, concerns persist regarding psychometric robustness and theoretical grounding, particularly in non-Western, mobile-first contexts. The present study evaluated the psychometric properties of the skilme Craft-the-Block Task, a gamified mobile assessment designed to measure Fluid Reasoning (Gf) and Psychomotor Speed (Gps) within the Cattell–Horn–Carroll (CHC) framework. A stratified random sample of 260 adolescents (aged 13–19 years; 50% male, 50% female; 56.1% rural, 43.9% urban) from educational institutions in Kerala, India completed the assessment under standardised conditions. Data analyses were conducted using R (Version 4.5.0; R Core Team, 2025) within the RStudio integrated development environment. Reliability analysis yielded strong internal consistency (Cronbach's $\alpha = .812$, 95% CI [.78, .85]). Exploratory Factor Analysis (EFA) with Varimax rotation revealed a theoretically coherent two-factor structure accounting for 62.24% of total variance, differentiating Nonverbal Concept Formation (Gf/Gv; Levels 3–10) from foundational Manual Dexterity (Gps; Levels 1–2). Corrected item-total correlations ranged from .443 to .763. No statistically significant bias was observed across gender ($p = .197$) or locality ($p = .788$). Academic grade level demonstrated a significant effect on performance (Kruskal–Wallis $H = 22.14$, $p = .001$), evidencing developmental sensitivity. These findings provide preliminary evidence supporting the reliability and structural validity of the skilme Craft-the-Block Task as a scalable, equitable digital cognitive assessment instrument.

Keywords: Block design, Gamified cognitive assessment, CHC theory, Fluid reasoning, Psychometrics, Mobile assessment, Adolescents, India, Digital divide, Psychomotor speed.

Introduction

The measurement of human intelligence is undergoing a paradigm shift, moving from static, examiner-administered instruments towards dynamic, digital, and gamified assessment models. For nearly a century, the assessment of visual-spatial intelligence has been dominated by physical manipulative tasks. The most prominent among these is the Block Design Test, originally developed by Kohs (1923) and integrated as a core subtest in the Wechsler Intelligence Scales (Wechsler, 1939, 2014). In clinical and educational psychology, block design performance is considered the gold standard for measuring nonverbal fluid intelligence and is a robust predictor of academic achievement, particularly in STEM domains (Wai et al., 2009).

Traditional block design administration faces significant constraints in contemporary educational contexts. Physical apparatus requires costly proprietary kits and mandates one-on-one administration by trained psychologists, creating scalability bottlenecks especially pronounced in developing nations. Additionally, formal testing environments induce test anxiety, a phenomenon that introduces construct-irrelevant variance, causing performance to reflect situational arousal rather than cognitive capacity (Cassady & Johnson, 2002; Putwain & Remedios, 2014). To address these limitations, researchers have turned to gamification; the application of game-design elements in non-game contexts (Hamari et al., 2014). Gamified assessments embed cognitive tasks within interactive environments, inducing states of flow that sustain engagement and attenuate evaluative anxiety (Nacke & Deterding, 2017).

Despite the proliferation of gamified cognitive tools, a critical gap persists. Few instruments have been rigorously validated against established psychometric frameworks such as the Cattell–Horn–Carroll (CHC) model (Schneider & McGrew, 2012), and validation studies are overwhelmingly conducted in Western populations, neglecting mobile-dominant developing regions. The present study addresses this gap by providing a comprehensive psychometric evaluation of the skilme Craft-the-Block Task among Indian adolescents.

Literature Review

The CHC Theoretical Framework

The CHC model constitutes the dominant evidence-based taxonomy of cognitive abilities, synthesising Carroll's (1993) three-stratum model with the Gf-Gc theory of Cattell (1963) and Horn (1968; as cited in Schneider & McGrew, 2012). The model posits a hierarchical structure with general intelligence (g) at Stratum III, broad abilities at Stratum II, and narrow abilities at Stratum I. For the present study, two Stratum II broad abilities are directly implicated. Fluid Reasoning (Gf) involves solving novel problems through abstract reasoning, specifically engaging Inductive Reasoning (I) and Visualisation (Vz) in pattern-based reconstruction tasks. Psychomotor Speed (Gps), specifically Finger Dexterity (P2), is operationalised through the touchscreen interaction mechanics of the digital interface. A fundamental psychometric challenge in digital test design is ensuring that interface-level Gps demands do not conflate with the cognitive Gf/Gv constructs of primary interest.

Gamification and Measurement Integrity

Gamification leverages principles from Self-Determination Theory (Ryan & Deci, 2000), which posits that intrinsic motivation requires fulfilment of competence, autonomy, and relatedness. A systematic review by Lumsden et al. (2016) demonstrated that carefully designed gamified cognitive tasks can maintain scientific validity while significantly improving participant engagement. Meta-analytic evidence indicates that gamification produces significant small-to-moderate effects on cognitive learning outcomes ($g = .49$, 95% CI [0.30, 0.69]; Sailer & Homner, 2020). However, Lumsden et al. (2016) also noted that design quality is a critical determinant of psychometric robustness, with mixed-domain measurement representing a particular risk. Concerning the validity of theory-driven game-based assessments specifically, Landers et al. (2022) demonstrated that such tools can achieve both positive participant reactions and high-quality psychometric measurement of general cognitive ability when designed with explicit theoretical grounding.

Test anxiety constitutes a further dimension of measurement integrity in adolescent populations. Zeidner (1998) documented that test anxiety interferes with working memory, depleting cognitive resources available for problem-solving. Gamified reframing of assessments as challenges rather than tests has been proposed as a mechanism for attenuating evaluative threat, with Putwain and Remedios (2014) documenting that 16.4% of English secondary school students reported high test anxiety, with significantly higher prevalence among female students (22.5%) than male students (10.3%), underscoring the practical importance of anxiety-reducing assessment design.

Digital Assessment Equity and the Digital Divide

A persistent concern in the digitalisation of educational assessment is the digital divide; the gap between those with and without ready access to digital technology (Warschauer, 2004). In the Indian context, however, mobile smartphone penetration has effectively leapfrogged desktop computing adoption. The Nacke and Deterding (2017) editorial on the maturing of gamification research highlighted mobile platforms as emerging primary delivery vehicles for gamified assessments, particularly in regions with high smartphone-to-desktop ratios. Validating whether a mobile-based assessment is free from locality bias is a critical requirement for any new instrument designed for use in diverse educational contexts.

Psychometric Standards for Digital Cognitive Tools

Contemporary validation standards require demonstration of internal consistency reliability, theoretically coherent factor structure, adequate item discrimination, measurement invariance across demographic groups, and developmental sensitivity (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Despite the rapid proliferation of gamified cognitive applications, few instruments have undergone validation meeting these criteria. Lumsden et al. (2016) identified in their systematic review that only a minority of reviewed gamified cognitive tools had published reliability data and fewer had reported factorial validity. The present study directly addresses this gap.

Research Gap, Aims, and Hypotheses

Research Gap

Despite the rapid expansion of gamified cognitive assessment platforms in mobile-dominant economies, four critical gaps persist in the literature: (a) insufficient psychometric validation against established intelligence frameworks, particularly the CHC model, in non-Western samples; (b) limited empirical evidence on the structural separability of cognitive (Gf/Gv) and motor (Gps) components in digital block design adaptations; (c) near-absence of mobile-first validation studies targeting adolescents in South Asia; and (d) scarcity of demographic fairness analyses examining rural-urban and gender-based measurement invariance in Indian digital assessment contexts.

Aims and Objectives

The primary aim was to evaluate the psychometric properties and structural validity of the skilme Craft-the-Block Task as a gamified mobile assessment of fluid reasoning and psychomotor speed among Indian adolescents, in alignment with the CHC theoretical framework. Specific objectives were: (1) to assess internal consistency reliability; (2) to examine the factor structure using EFA; (3) to evaluate item discrimination indices across difficulty levels; (4) to investigate demographic fairness across gender and locality; (5) to assess developmental sensitivity across academic grade levels; and (6) to explore the relationship between processing speed and accuracy.

Hypotheses

1. H1: The skilme Craft-the-Block Task will demonstrate adequate internal consistency reliability (Cronbach's $\alpha \geq .70$) across its 10 levels.
2. H2: The task will exhibit a distinct two-factor structure separating Psychomotor Speed/Dexterity (Gps) from Fluid and Visual-Spatial Reasoning (Gf/Gv), consistent with the CHC framework.
3. H3: Items will demonstrate acceptable discrimination, with corrected item-total correlations ≥ .30 varying systematically across difficulty levels.
4. H4: Gender will not significantly influence total scores, indicating demographic fairness.
5. H5: Locality (rural vs. urban) will not significantly influence total scores; however, academic grade level and task completion time will demonstrate significant associations with performance.

Method

Research Design

A quantitative, cross-sectional psychometric validation design was employed, consistent with contemporary standards for the evaluation of digital cognitive instruments (AERA, APA, & NCME, 2014).

Participants

A total of 260 participants ($N = 260$) aged 13–19 years ($M = 16.8$, $SD = 2.1$) were recruited through stratified random sampling across secondary and higher-secondary educational institutions in Kerala, India. Stratification variables included gender and locality. The final sample comprised 130 male (50%) and 130 female (50%) participants; 146 rural (56.1%) and 114 urban (43.9%) participants. Academic levels spanned Grade 8 through undergraduate level.

Inclusion criteria required: (a) age 13–19 years; (b) enrolment in a recognised educational institution; and (c) basic familiarity with touchscreen smartphones, operationally defined as self-reported regular use for a minimum of six months. Exclusion criteria comprised visual impairments uncorrected by standard lenses, motor disabilities precluding independent touchscreen interaction, or incomplete task completion (< 8 of 10 levels). Ethical approval was obtained from the Institutional Ethics Committee, and written informed consent was secured from all participants and, for minors, from guardians.

Instrument

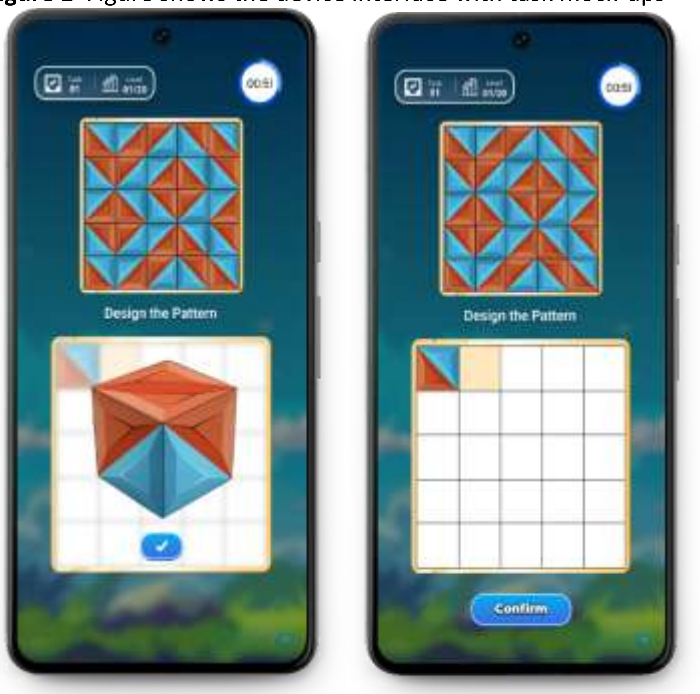
Overview.

The skilme Craft-the-Block Task is a gamified, language-independent digital module within the skilme cognitive assessment application. It consists of 10 levels of systematically increasing difficulty, progressing from simple 2×2 grid patterns to complex abstract configurations requiring mental rotation and multistep synthesis. The task is hypothesised, within the CHC framework, to measure Fluid Reasoning (Gf; Inductive Reasoning and Visualisation), Visual Processing (Gv), and Psychomotor Speed (Gps; specifically Finger Dexterity).

Task mechanics.

The task employs a “Select and Rotate” mechanic. The participant views a target pattern in an upper display panel, selects individual cells in an empty response grid, and rotates a virtual dice to match the correct colour face or diagonal orientation. This mechanic was chosen over drag-and-drop to minimise slip errors and standardise motor demands.

Figure 1 Figure shows the device interface with task mock-ups



Procedure

All sessions were conducted in distraction-free, supervised classroom environments. A uniform set of identical Android smartphones was provided by the research team to eliminate hardware variance, a recognised source of systematic error in digital assessment research. Devices were fully charged and connected to a dedicated local Wi-Fi network at each venue. Trained administrators briefed participants using a challenge framing to minimise evaluative anxiety. A mandatory tutorial phase preceded the scored assessment. Post-assessment qualitative feedback was collected via Google Forms.

Data Analysis

Data analyses were conducted using R (Version 4.5.0; R Core Team, 2025) within the RStudio integrated development environment. Normality was assessed via the Shapiro–Wilk test, which indicated significant non-normality ($W = .94$, $p < .001$), necessitating non-parametric procedures. Reliability was estimated using Cronbach’s α with 95% bootstrap confidence intervals. Item discrimination was evaluated through corrected item-total correlations, with values ≥ .30 considered acceptable and ≥ .50 considered good (Nunnally & Bernstein, 1994).

Structural validity was examined via EFA using Principal Component Analysis (PCA) with Varimax rotation. Data factorability was assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett’s Test of Sphericity. Factor retention followed Kaiser’s criterion (eigenvalues > 1.0), supported by Scree Plot inspection. Group comparisons across gender and locality employed the Mann–Whitney U test with effect size r; comparisons across academic levels used the Kruskal–Wallis H test with η^2. Speed–accuracy relationships were examined using Spearman rank-order correlations (ρ), consistent with the non-normal data distribution.

Results

Descriptive Statistics and Item Difficulty

The mean total score was 17.58 (SD = 6.09, Mdn = 16.00) out of a maximum of 40, indicating appropriate difficulty without floor or ceiling effects at the composite level. Mean total task completion time was 13.91 minutes (SD = 1.34), with low variability suggesting consistent engagement across participants. Table 1 presents descriptive statistics for all performance metrics. Notable features include a pronounced negative skew for Level 1 (−1.45), confirming an expected ceiling effect on the introductory dexterity level, and positive skew values for Levels 6 and 9 (0.89 and 0.73, respectively), indicating effective difficulty calibration for advanced reasoning levels.

Table 1 Descriptive Statistics for Performance Metrics Across All Task Levels (N = 260)

Metric	M	SD	Mdn	Min	Max	Skewness
Total Score	17.58	6.09	16.00	0.00	30.00	−0.12
Total Time (min)	13.91	1.34	14.11	2.88	16.67	−0.08
Level 1 Score	3.17	1.03	3.00	0.00	4.00	−1.45
Level 2 Score	2.86	0.98	3.00	0.00	4.00	−0.87
Level 3 Score	1.57	1.21	2.00	0.00	4.00	0.41
Level 4 Score	2.22	1.15	2.00	0.00	4.00	0.10
Level 5 Score	1.25	1.29	1.00	0.00	4.00	0.62
Level 6 Score	1.05	1.24	0.00	0.00	4.00	0.89
Level 7 Score	1.48	1.19	1.00	0.00	4.00	0.53
Level 8 Score	1.62	1.22	2.00	0.00	4.00	0.31
Level 9 Score	1.03	1.21	1.00	0.00	4.00	0.73
Level 10 Score	2.33	1.27	2.00	0.00	4.00	0.15

Note. M = mean; SD = standard deviation; Mdn = median. Maximum possible score per level = 4; total maximum = 40.

Item Discrimination and Internal Consistency Reliability

Item discrimination was assessed through corrected item-total correlations. As shown in Table 2, correlations ranged from .443 (Level 1) to .763 (Level 10). Lower correlations for Levels 1 and 2 are consistent with the theoretical expectation that these introductory dexterity levels measure a construct partially distinct from the complex reasoning assessed by the total scale. All item-total correlations exceeded the .30 threshold for acceptable discrimination (Nunnally & Bernstein, 1994), supporting H3. The task demonstrated high internal consistency: Cronbach's α = .812 (95% CI [.78, .85]). Alpha-if-item-deleted analyses confirmed that no individual level removal would meaningfully improve scale reliability. These findings support H1.

Table 2 Item Difficulty and Discrimination Indices for the skilme Craft-the-Block Task (N = 260)

Level	M Score	Difficulty	Corrected r(it)	α if Deleted
Level 1	3.17	Easy	.443	.804
Level 2	2.86	Easy	.540	.798
Level 3	1.57	Moderate	.697	.781
Level 4	2.22	Moderate	.697	.783
Level 5	1.25	Difficult	.709	.780

Level	M Score	Difficulty	Corrected r(it)	α if Deleted
Level 6	1.05	Difficult	.730	.778
Level 7	1.48	Moderate	.722	.779
Level 8	1.62	Moderate	.727	.778
Level 9	1.03	Difficult	.727	.777
Level 10	2.33	Moderate	.763	.772

Note. Difficulty classification: Easy = $M \geq 2.50$; Moderate = 1.30–2.49; Difficult = ≤ 1.29. Corrected r(it) = corrected item-total correlation; α if Deleted = Cronbach's alpha if the level is excluded from scoring.

Structural Validity: Exploratory Factor Analysis

The KMO measure of sampling adequacy was .84 and Bartlett's Test of Sphericity was statistically significant ($\chi^2(45) = 512.34$, $p < .001$), confirming data suitability for factor analysis. PCA with Varimax rotation yielded a two-factor solution with eigenvalues > 1.0 (PC1 = 5.17; PC2 = 1.05), jointly explaining 62.24% of total variance (PC1 = 51.70%; PC2 = 10.54%). Table 3 presents the rotated component matrix. Levels 3–10 loaded strongly on Factor 1 (range: .76–.88), interpreted as Nonverbal Concept Formation (Gf/Gv). Levels 1 and 2 loaded predominantly on Factor 2 (.79 and .75, respectively), interpreted as foundational Manual Dexterity (Gps). These findings provide strong support for H2.

Table 3 Rotated Component Matrix from Exploratory Factor Analysis of the skilme Craft-the-Block Task (N = 260)

Level	Factor 1: Gf/Gv	Factor 2: Gps	h^2
Level 1	.12	.79	.64
Level 2	.50	.75	.81
Level 3	.76	−.22	.62
Level 4	.80	−.15	.66
Level 5	.82	−.23	.72
Level 6	.85	−.24	.77
Level 7	.83	−.27	.76
Level 8	.84	−.32	.80
Level 9	.85	−.24	.78
Level 10	.88	−.20	.81
Eigenvalue	5.17	1.05	;
% Variance	51.70%	10.54%	62.24%

Note. Extraction: Principal Component Analysis. Rotation: Varimax with Kaiser normalisation. KMO = .84; Bartlett's $\chi^2(45) = 512.34$, $p < .001$. h^2 = communality. Loadings ≥ .40 are considered meaningful.

Speed–Accuracy Relationships

Spearman rank-order correlations between level completion time and level score are presented in Table 4. A systematic pattern of increasingly negative correlations was observed across ascending task levels. Level 1 yielded a weak negative correlation ($\rho = -.169$, $p = .006$), consistent with dexterity-dominated performance. Higher-order levels demonstrated progressively stronger negative correlations (Levels 7–10: ρ range −.612 to −.692, all $p < .001$), indicating that at greater cognitive complexity, processing efficiency is more tightly coupled with accuracy. This gradient supports the validity of the composite speed-accuracy scoring model and is consistent with the speed-accuracy coupling component of H5.

Table 4 Spearman Rank-Order Correlations Between Level Completion Time and Level Score (N = 260)

Level	ρ	p	Interpretation
Level 1	−.169	.006	Weak ; dexterity-dominant; speed variance limited
Level 3	−.321	<.001	Moderate ; onset of reasoning-speed coupling
Level 5	−.569	<.001	Strong ; processing efficiency integral to accuracy
Level 7	−.612	<.001	Strong ; elevated cognitive load amplifies time penalty
Level 9	−.655	<.001	Strong ; near-ceiling reasoning demands
Level 10	−.692	<.001	Very strong ; dual burden of complexity and speed

Note. All reported correlations are statistically significant at $p \leq .006$ or below. Negative correlations indicate that longer completion times are associated with lower scores. ρ = Spearman's rho.

Demographic Group Differences

Table 5 presents all group comparison results. A Mann–Whitney U test revealed no statistically significant difference between male and female participants in total score ($U = 8241.0$, $p = .197$, $r = .07$), indicating a trivial effect size and supporting the fairness interpretation of H4. Similarly, no significant locality effect was observed ($U = 8112.5$, $p = .788$, $r = .01$), indicating rural and urban participants performed at statistically equivalent levels. A Kruskal–Wallis H test demonstrated a significant effect of academic grade level on total score ($H = 22.14$, $p = .001$, $\eta^2 = .08$, moderate effect), with a monotonic increase in performance from Grade 8 through undergraduate level. These findings support the demographic fairness and developmental sensitivity dimensions of H5.

Table 5 Demographic Group Differences in Total Score on the skilme Craft-the-Block Task (N = 260)

Variable	Category	n	Mean Rank	Statistic	p	Effect Size	Outcome
Gender	Male	130	125.4	U = 8241.0	.197	r = .07	H4: No bias
	Female	130	135.6				
Locality	Rural	146	131.8	U = 8112.5	.788	r = .01	H5: No bias
	Urban	114	128.9				
Grade	Gr. 8–Grad.	260	Monotonic ↑	H = 22.14	.001	$\eta^2 = .08$	H5: Significant

Note. Mann–Whitney U tests used for Gender and Locality; Kruskal–Wallis H test used for Academic Level. Effect sizes: $r = Z/\sqrt{N}$ for Mann–Whitney; $\eta^2 = H/(N-1)$ for Kruskal–Wallis. H4 = Hypothesis 4; H5 = Hypothesis 5.

Discussion

Structural Validity and the CHC Framework

The most theoretically significant finding was the emergence of a clean two-factor structure mirroring CHC predictions. The separation of foundational dexterity (Gps; Levels 1–2) from complex concept formation (Gf/Gv; Levels 3–10) represents a methodological advancement over traditional physical block design tasks, in which motor coordination and abstract reasoning are systematically confounded. By structurally embedding a dexterity calibration phase at the outset, the skilme Craft-the-Block Task functions as its own motor screener, ensuring that subsequent level performance can be attributed with greater confidence to Gf/Gv rather than interface learning artefacts. This finding directly addresses a validation concern raised by Lumsden et al. (2016), who noted that mixed-domain measurement is a common problem in gamified cognitive assessments.

Reliability and Environmental Standardisation

The Cronbach's α of .812 is notably high for a brief, gamified screening instrument, meeting the $\geq .80$ threshold considered adequate for applied screening purposes (Nunnally & Bernstein, 1994). This reliability is attributable not only to the structural coherence of the item set but also to rigorous environmental standardisation: uniform devices, controlled settings, and administrator-delivered briefings eliminated hardware variance identified as a major source of error in digital assessment research. Sailer and Homner's (2020) meta-analytic findings that gamification produces significant small effects on cognitive learning outcomes ($g = .49$) are consistent with the present task's capacity to measure genuine cognitive variance rather than engagement artefacts.

Speed–Accuracy Integration

The progressively stronger negative Spearman correlations between completion time and score across ascending levels are consistent with established theory on mental processing speed (Jensen, 2006), which demonstrates that processing speed becomes increasingly salient as task complexity approaches the limits of working memory capacity. The composite scoring algorithm that awards higher marks for accurate fast responses thus capitalises on meaningful cognitive variance. However, the relationship must be interpreted cautiously, as faster performance may in some cases reflect strategic differences or prior exposure to similar game formats rather than pure cognitive efficiency.

Equity and the Digital Divide

The absence of significant rural-urban performance differentials ($p = .788$, $r = .01$) suggests that touch-based mobile interaction constitutes a genuinely culture-neutral interface modality for the Kerala population. This finding is particularly noteworthy given persistent theoretical concerns about the digital divide in South Asian education (Warschauer, 2004). The result is consistent with Landers et al.'s (2022) demonstration that theory-driven game-based assessments can achieve fairness across demographic groups, and with the broader evidence on gamification's potential to reduce barriers to assessment participation documented by Lumsden et al. (2016). However, it is important to contextualise this finding within Kerala's high literacy rate and smartphone penetration; generalisation to states with lower digital access requires caution.

Developmental Sensitivity

The significant monotonic increase in performance across academic levels ($H = 22.14$, $p = .001$, $\eta^2 = .08$) provides indirect construct validity evidence. Fluid reasoning is well established as a cognitive ability that develops throughout childhood and adolescence (McArdle et al., 2002), and the present findings mirror this trajectory, suggesting the instrument is sensitive to genuine cognitive developmental differences.

Limitations

Several limitations constrain the present findings. First, the absence of concurrent validity data against standardised instruments (e.g., Wechsler Intelligence Scale for Children, Fifth Edition; Raven's Progressive Matrices) precludes conclusions regarding criterion validity. Second, the use of PCA rather than Common Factor Analysis represents a methodological limitation; Confirmatory Factor Analysis (CFA) should be employed in replication studies. Third, the sample is restricted to a single Indian state with atypically high literacy and smartphone penetration; cross-state and cross-national samples are required for generalised normative standards. Fourth, test-retest reliability data were not collected, leaving temporal score stability unestablished. Fifth, no independent measure of fine motor skills was included to corroborate the dexterity factor interpretation.

Conclusion

The present study provides robust preliminary evidence that the skilme Craft-the-Block Task is a reliable, structurally valid, and demographically equitable mobile-based cognitive assessment tool for Indian adolescents. Aligning with CHC theory, the instrument successfully differentiates psychomotor speed (Gps) from fluid and visual-spatial reasoning (Gf/Gv), producing a two-factor solution consistent with theoretical predictions. Strong internal consistency (α = .812), adequate item discrimination, absence of demographic bias, and significant developmental sensitivity collectively establish a promising psychometric profile. Future research must prioritise concurrent validity, CFA-based structural replication, longitudinal assessment of score stability, Item Response Theory calibration, and cross-state validation to consolidate the evidence base for this instrument.

ACKNOWLEDGEMENTS
The authors express sincere gratitude to the school principals, teachers, and student participants across Kerala who contributed their time to this research. We acknowledge the technical support provided by the skilme platform development team at Master n Masters, Kozhikode.

CONFLICT OF INTEREST
Master n Masters LLP is the organization that developed and funded Skilme, the commercially developed gamified psychometric assessment tool examined in this study. The assessment was collaboratively developed by Dr. Ameer Hassan, Dr. Roshan Anie Alex, Shyam Kumar and Amal Tomy. Dr. Ameer Hassan, Head of the Department of Psychology at Amal College of Advanced Studies, conceptualized the Skilme CogniQues assessment tasks. Dr. Roshan Anie Alex (R&D Lead), Shyam Kumar (Research Associate), and Amal Tomy (Research Associate) as the team members of the Research and Development Department of Master n Masters contributed to the development and empirical evaluation of the skilme assessment, including item development, data collection, data management, statistical analysis, interpretation of findings, and standardization of the tool.

The authors acknowledge their professional and institutional affiliations with Master n Masters, as well as their involvement in the development, evaluation, and standardization of Skilme, as potential sources of conflict of interest. The authors confirm that these relationships did not influence the study design, data collection, analysis, interpretation, standardization procedures, or reporting of the findings.

ETHICAL STATEMENT
This study was conducted in accordance with the Declaration of Helsinki. Institutional Ethics Committee approval was obtained prior to data collection. Written informed consent was secured from all adult participants and from legal guardians of minor participants. Participation was voluntary, and data were anonymised prior to analysis.

References

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
2. Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312
3. Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. Contemporary Educational Psychology, 27(2), 270–295. https://doi.org/10.1006/ceps.2001.1094
4. Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54(1), 1–22. https://doi.org/10.1037/h0046743
5. Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In Proceedings of the 47th Hawaii International Conference on System Sciences (pp. 3025–3034). IEEE. https://doi.org/10.1109/HICSS.2014.377
6. Jensen, A. R. (2006). Clocking the mind: Mental chronometry and individual differences. Elsevier.
7. Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39(1), 31–36. https://doi.org/10.1007/BF02291575
8. Kohs, S. C. (1923). Intelligence measurement: A psychological and statistical study based upon the block-design tests. Macmillan.
9. Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. Simulation & Gaming, 45(6), 752–768. https://doi.org/10.1177/1046878114563660
10. Landers, R. N., Armstrong, M. B., Collmus, A. B., Mujcic, S., & Blaik, J. (2022). Theory-driven game-based assessment of general cognitive ability: Design theory, measurement, prediction of performance, and test fairness. Journal of Applied Psychology, 107(10), 1655–1677. https://doi.org/10.1037/apl0000954

11. Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. JMIR Serious Games, 4(2), e11. https://doi.org/10.2196/games.5888]
12. McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. Developmental Psychology, 38(1), 115–142. https://doi.org/10.1037/0012-1649.38.1.115
13. Nacke, L. E., & Deterding, S. (2017). The maturing of gamification research. Computers in Human Behavior, 71, 450–454. https://doi.org/10.1016/j.chb.2016.11.062
14. Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.
15. Putwain, D., & Remedios, R. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. Educational Studies, 40(5), 554–570. https://doi.org/10.1080/03055698.2014.953914
16. R Core Team. (2025). R: A language and environment for statistical computing (Version 4.5.0). R Foundation for Statistical Computing. https://www.R-project.org/
17. Rozencwajg, P. (1991). Analysis of problem solving strategies on the Kohs Block Design Test. European Journal of Psychology of Education, 6(1), 73–88. https://doi.org/10.1007/BF03172763
18. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68
19. Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. Educational Psychology Review, 32(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w
20. Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (3rd ed., pp. 99–144). Guilford Press.
21. Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. Journal of Educational Psychology, 101(4), 817–835. https://doi.org/10.1037/a0016127
22. Warschauer, M. (2004). Technology and social inclusion: Rethinking the digital divide. MIT Press.
23. Wechsler, D. (1939). The measurement of adult intelligence. Williams & Wilkins.
24. Wechsler, D. (2014). Wechsler Intelligence Scale for Children; Fifth Edition. NCS Pearson.
25. Zeidner, M. (1998). Test anxiety: The state of the art. Plenum Press.