Enigma in Education
https://enigma.or.id/index.php/edu

e-ISSN: 3026-5827

ORIGINAL ARTICLE
Enigma in Education | Vol. 4, No. 1 (2026) | pp. 7–12

Development and Validation of a Creativity Assessment Rubric for Evaluating Student Projects in Higher Education

Vita Amanda[1,*], Sophia Lucille Rodriguez[2]

[1]*Department of Psychiatry, CMHC Research Center, Palembang, Indonesia*
[2]*Department of Pediatrics, Trinidad General Hospital, Mexico City, Mexico*

ARTICLE INFO

Article history
Received: March 20, 2026
Revised: April 19, 2026
Accepted: May 2, 2026
Available online: May 18, 2026
Published: May 30, 2026

Keywords
Assessment rubric; creativity; higher education; psychometric properties; rubric validation

Corresponding author
Vita Amanda
vitaamandava@gmail.com

DOI
https://doi.org/10.61996/edu.v4i1.121

ABSTRACT

Background. Educators in higher education lack psychometrically validated instruments for assessing student creative output, even though creativity is recognized as a core 21st-century competency.

Objective. This study aimed to develop and validate the Creative Output Assessment Rubric (COAR) for evaluating student creative projects in university settings.

Methods. We used a multi-phase instrument development design at a private university in Palembang, Indonesia, during the 2023/2024 academic year. The phases were item generation and expert review (n = 8; S-CVI/Ave = 0.92), pilot testing (n = 45), exploratory factor analysis (EFA; n = 215), and confirmatory factor analysis (CFA) in an independent sample (n = 220).

Results. A five-factor structure of 22 items (Originality, Elaboration, Flexibility, Aesthetic Quality, and Technical Execution) was confirmed and explained 67.4% of total variance. The confirmatory model showed good fit (CFI = 0.94; TLI = 0.93; RMSEA = 0.048). Internal consistency was strong (Cronbach's $\alpha = 0.93$; McDonald's $\omega = 0.94$), inter-rater reliability was adequate (ICC = 0.84; 95% CI: 0.79–0.88), and a significant correlation with independent expert ratings supported criterion validity ($r = 0.74$; $p < 0.001$). Configural measurement invariance held across four academic faculties.

Conclusion. The COAR showed good evidence of validity and reliability for assessing student creative output in this sample. It gives educators a validated scoring framework for the formative and summative evaluation of creative work, pending replication in other settings.

Copyright: © 2026 The Author(s). Published by Enigma Institute on behalf of Enigma in Education. The authors retain copyright of their work.

Access license: This is an open access article distributed under the terms of the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and any derivative work is distributed under the same license.

1. Introduction

Creativity is widely regarded as one of the key 21st-century competencies that universities are expected to develop and assess, and both the Partnership for 21st Century Learning (P21) framework and the OECD Learning Compass 2030 name it a core competency for future-ready graduates.[1–3] Assessing it is harder. Educators report real difficulty in judging student creative work in a way that is both valid and reliable, and many fall back on subjective impressions instead of standardized instruments.[4–6]

Creativity assessment in education has usually followed one of two routes: divergent thinking tests, which gauge creative potential, and product-based assessments, which evaluate what students actually create.[7–9] Divergent thinking tests such as the Torrance Tests of Creative Thinking have been examined extensively from a psychometric standpoint. Product instruments lag behind.[8,10] The VALUE rubrics of the Association of American Colleges and Universities offer a general framework. They lack, however, comprehensive psychometric validation through factor analysis and criterion-related validity evidence.[5,6]

Creativity is generally understood as a multidimensional construct. In Amabile's componential model, creative production is determined by domain-relevant skills, creativity-relevant processes, and intrinsic task motivation,[4,11] while Torrance's framework describes creative output as originality, fluency, flexibility, and elaboration.[8] The Four-C model of Kaufman and Beghetto separates mini-c, little-c, Pro-c, and Big-C creativity; the first two matter most for educational assessment.[12] Bloom's revised taxonomy places “create” at the top of the cognitive hierarchy, which shows how central creative production is to higher-order learning outcomes.[3] Vygotsky's zone of proximal development supplies a further theoretical basis: rubrics can act as scaffolds that make implicit creativity criteria explicit to learners.[13]

Theory has advanced, yet four gaps persist. Most existing measures of creativity were built for psychological research, not for everyday classroom use.[6,9] Psychometrically validated rubrics for creative project outputs in higher education are scarce.[5,14] The few that exist have seldom been validated through both exploratory and confirmatory factor analysis.[8,15] Evidence on criterion-related validity and inter-rater reliability is often missing too.[7,10]

This study aimed to develop and validate the Creative Output Assessment Rubric (COAR) for evaluating student creative projects in higher education. A multi-phase design was used. It combined expert review, pilot testing, exploratory factor analysis, and confirmatory factor analysis. The rubric is meant for educators who assess creative project output across multiple academic disciplines in university settings. By supplying comprehensive validation evidence, the study adds to the educational assessment literature.

2. Methods

2.1 Study design and setting

A multi-phase design for instrument development and validation was used, following established psychometric procedures.[16,17] Data were collected at a private university in Palembang, Indonesia, during the 2023/2024 academic year.

2.2 Phase 1: item generation and content validation

Phase 1 (Item Generation and Content Validation) began with an initial pool of 30 items, drawn from a comprehensive review of creativity theory and existing assessment frameworks.[3,7,8,11,18] A construct-mapping approach to item development[19] organized the items into five theoretically derived dimensions: Originality (novelty and uniqueness), Elaboration (depth and detail), Flexibility (multiple perspectives), Aesthetic Quality (visual or conceptual appeal), and Technical Execution (skill and craftsmanship). "Usefulness" was deliberately not made a separate dimension. It is operationalized within Elaboration (practical applicability of ideas) and Technical Execution (functional quality of output).[11,20] An expert panel of eight specialists (four creativity researchers, four higher education instructors) evaluated content validity. The Content Validity Index (CVI) was calculated at the item level (I-CVI ≥ 0.78 for retention) and the scale level (S-CVI/Ave, the average of the item-level indices).[17]

2.3 Phase 2: pilot testing

In Phase 2 (Pilot Testing), the refined instrument was given to 45 university students who had recently completed creative projects. Items were rated on a five-point Likert-type scale (1 = not at all present to 5 = exceptionally present). Item analysis covered corrected item-total correlations (threshold ≥ 0.30) and reliability.

2.4 Phase 3: exploratory factor analysis

Phase 3 (Exploratory Factor Analysis, EFA) administered the 22-item COAR to 215 students from four faculties (Education, Arts and Design, Sciences, Social Sciences). Each project was scored independently by two trained raters. Principal axis factoring with oblimin rotation was conducted. Oblimin was chosen over promax because creativity dimensions were expected to be correlated.[16] Parallel analysis and eigenvalues greater than 1.0 determined factor retention.

2.5 Phase 4: confirmatory factor analysis and validity evidence

Phase 4 (Confirmatory Factor Analysis, CFA) recruited an independent sample of 220 students from different cohorts. The five-factor model was tested by CFA with maximum likelihood estimation in IBM SPSS AMOS version 26, and multivariate normality was checked with Mardia's coefficient. Model fit was judged against χ^2/df (acceptable < 3.0), the comparative fit index (CFI ≥ 0.90), the Tucker–Lewis index (TLI ≥ 0.90), the root mean square error of approximation (RMSEA ≤ 0.06), and the standardized root mean square residual (SRMR ≤ 0.08).[16] Modification indices were examined. None was applied, in order to preserve theoretical integrity. Reliability was assessed with Cronbach's alpha, McDonald's omega, and the intraclass correlation coefficient (ICC; two-way mixed, consistency model, with individual rater values reported). Convergent validity was evaluated through the average variance extracted (AVE ≥ 0.50), and discriminant validity through the Fornell-Larcker criterion and the heterotrait–monotrait (HTMT) ratio (< 0.85). For criterion-related validity, COAR scores were correlated with independent expert panel ratings (separate from the development panel) in the validation sample (n = 215). Configural measurement invariance was tested across the four faculties, and practical utility was assessed by recording scoring time per project. All analyses used IBM SPSS Statistics 26 and AMOS 26 at alpha = 0.05.

2.6 Ethical considerations

No sensitive personal or health information was gathered. Participants were actively recruited. They took part voluntarily after giving informed consent, and all data were anonymized before analysis. The study was approved by the Ethical Committee of Social Research, Enigma Institute Indonesia (approval number 2023/189).

3. Results

3.1 Participant characteristics

Demographic characteristics of the CFA sample (n = 220) appear in Table 1. There were 92 males (41.8%) and 128 females (58.2%), and the mean age was 21.4 years (SD = 1.8). The students came from four faculties: Education (n = 78, 35.5%), Arts and Design (n = 62, 28.2%), Sciences (n = 48, 21.8%), and Social Sciences (n = 32, 14.5%).

Table 1. Demographic characteristics of study participants (CFA sample, N = 220).

Characteristic	n (%)
Gender	
Male	92 (41.8)
Female	128 (58.2)
Age, mean (SD)	21.4 (1.8)
Faculty	
Education	78 (35.5)
Arts and Design	62 (28.2)
Sciences	48 (21.8)
Social Sciences	32 (14.5)
Year of study	
Second year	95 (43.2)
Third year	82 (37.3)
Fourth year	43 (19.5)
Creative project type	
Visual arts / design	87 (39.5)
Written / research	68 (30.9)
Performance / multimedia	42 (19.1)

Characteristic	n (%)
STEM innovation	23 (10.5)

Notes: SD = standard deviation. Values are n (%) unless otherwise indicated; percentages are calculated within the total CFA sample (N = 220).

3.2 Content validity and pilot testing

The Phase 1 expert panel evaluation gave an S-CVI/Ave of 0.92, and every retained item reached an I-CVI of 0.78 or higher. Eight items were dropped, which reduced the pool from 30 to 22 items. Pilot testing in Phase 2 showed acceptable corrected item-total correlations (range: 0.38–0.72). No further items needed to be removed.

3.3 Exploratory factor analysis

Phase 3 produced a Kaiser-Meyer-Olkin value of 0.89, and Bartlett's test of sphericity was significant (χ^2 = 2847.62, df = 231, p < 0.001). Principal axis factoring extracted five factors that together explained 67.4% of total variance: Originality (5 items, 18.4%), Elaboration (5 items, 14.2%), Flexibility (4 items, 12.8%), Aesthetic Quality (4 items, 11.5%), and Technical Execution (4 items, 10.5%). Every item loaded above 0.40 on its intended factor. There were no problematic cross-loadings.

3.4 Confirmatory factor analysis

Table 2 lists the CFA model fit indices and reliability coefficients. Fit of the five-factor model was good: χ^2/df = 1.82, CFI = 0.94, TLI = 0.93, RMSEA = 0.048 (90% CI: 0.038–0.058), SRMR = 0.041. Mardia's coefficient (12.4) fell below the critical threshold, which supported maximum likelihood estimation. The five-factor model fitted substantially better than the one-factor model (χ^2/df = 4.56, CFI = 0.72, RMSEA = 0.112), the three-factor model (χ^2/df = 2.91, CFI = 0.84, RMSEA = 0.078), and a second-order five-factor specification (CFI = 0.92, TLI = 0.90, RMSEA = 0.056). No modification indices were applied.

Table 2. Confirmatory factor analysis model fit and reliability of the COAR (N = 220).

Parameter	Value	Criterion
Model fit indices		
χ^2/df	1.82	< 3.00
CFI	0.94	≥ 0.90
TLI	0.93	≥ 0.90
RMSEA (90% CI)	0.048 (0.038–0.058)	≤ 0.06
SRMR	0.041	≤ 0.08
Internal consistency		
Full scale (α / ω)	0.93 / 0.94	≥ 0.70
Originality (α / ω)	0.86 / 0.87	≥ 0.70
Elaboration (α / ω)	0.84 / 0.85	≥ 0.70
Flexibility (α / ω)	0.82 / 0.83	≥ 0.70
Aesthetic Quality (α / ω)	0.80 / 0.81	≥ 0.70
Technical Execution (α / ω)	0.79 / 0.80	≥ 0.70
Inter-rater reliability		
Overall ICC (95% CI)	0.84 (0.79–0.88)	≥ 0.75
Criterion validity		
Pearson's r with expert ratings	0.74 (p < 0.001)	Significant

Notes: α = Cronbach's alpha; ω = McDonald's omega; ICC = intraclass correlation coefficient (two-way mixed, consistency); CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual. Criterion validity was assessed in the validation sample (n = 215).

Standardized factor loadings for all 22 items ranged from 0.66 to 0.85, well above the 0.40 threshold (Figure 1). Flexibility item F3 (0.85) and Originality item O2 (0.82) had the highest loadings. The lowest acceptable loading belonged to Flexibility item F4 (0.66).

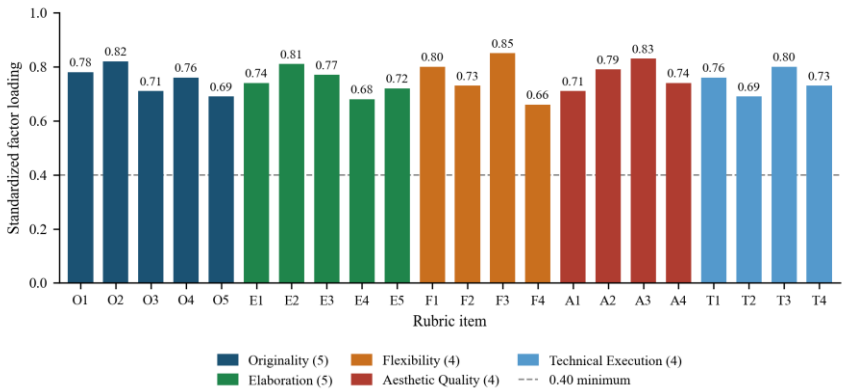


Figure 1. Standardized factor loadings from confirmatory factor analysis of the COAR (N = 220). The dashed line marks the 0.40 minimum retention threshold; bars are grouped and colored by dimension.

Figure 2 compares model fit across the alternative CFA models and confirms that the five-factor solution is superior. In the final model, CFI (0.94) and TLI (0.93) were above the 0.90 threshold and RMSEA (0.048) was below the 0.06 criterion. The one- and three-factor models did not meet these benchmarks.

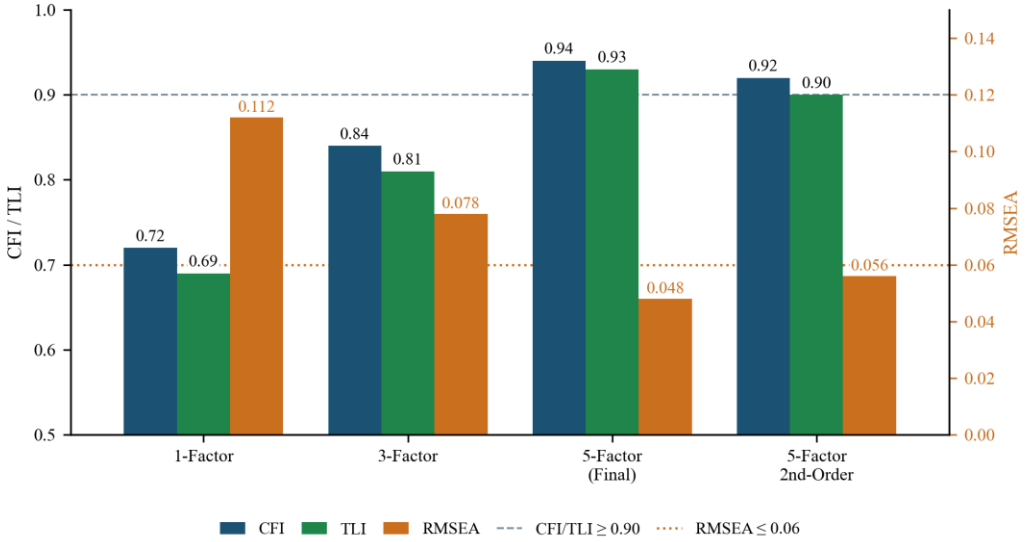


Figure 2. Comparison of model fit indices across alternative CFA models for the COAR. CFI and TLI are read on the left axis; RMSEA is read on the right axis.

3.5 Reliability, convergent validity, and discriminant validity

Inter-rater reliability was adequate (overall ICC = 0.84, 95% CI: 0.79–0.88; individual rater ICCs ranged from 0.78 to 0.87). Table 3 gives the evidence for convergent and discriminant validity. AVE values ranged from 0.51 to 0.58 and all exceeded 0.50. For every factor, the square root of AVE was larger than its inter-factor correlations, which confirms discriminant validity. HTMT ratios (0.58 to 0.82) all stayed below the 0.85 threshold.

Table 3. Convergent and discriminant validity: AVE and inter-factor correlations (CFA sample, N = 220).

Factor	AVE	√AVE	OR	EL	FL	AQ	TE
Originality (OR)	0.58	0.76	1.00				
Elaboration (EL)	0.56	0.75	0.62	1.00			
Flexibility (FL)	0.54	0.73	0.58	0.55	1.00		
Aesthetic Quality (AQ)	0.52	0.72	0.51	0.59	0.48	1.00	
Technical Execution (TE)	0.51	0.71	0.45	0.52	0.43	0.61	1.00

Notes: AVE = average variance extracted; √AVE = square root of AVE; OR = Originality; EL = Elaboration; FL = Flexibility; AQ = Aesthetic Quality; TE = Technical Execution. Off-diagonal values are inter-factor correlations. Discriminant validity is supported when √AVE exceeds the corresponding off-diagonal correlations.

3.6 Criterion validity, measurement invariance, and practical utility

In the validation sample (n = 215), COAR total scores correlated positively and significantly with independent expert panel ratings ($r = 0.74$, $p < 0.001$, 95% CI: 0.67–0.80, $R^2 = 0.55$), as Figure 3 shows. Configural measurement invariance was supported across the four faculties ($\Delta CFI < 0.01$ between the unconstrained and constrained models). Scoring took 14.6 minutes per project on average (SD = 3.2).

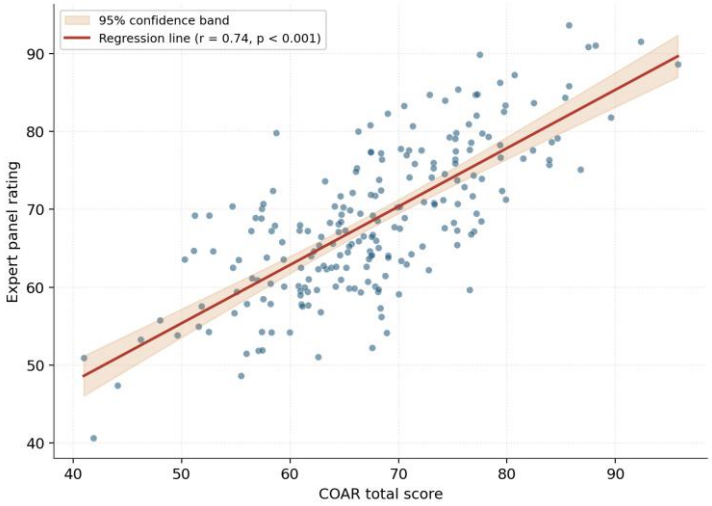


Figure 3. COAR total scores against expert panel ratings: criterion validity (N = 215, r = 0.74, p < 0.001). The shaded region is the 95% confidence band around the regression line.

4. Discussion

4.1 Principal findings

The study produced the Creative Output Assessment Rubric (COAR), a 22-item instrument with five factors for evaluating student creative projects in higher education, and tested its psychometric quality. Content validity was high (S-CVI/Ave = 0.92). Both EFA and CFA supported the five-factor structure, and internal consistency was strong (α = 0.93, ω = 0.94). Inter-rater reliability was adequate (ICC = 0.84). Criterion-related validity was significant (r = 0.74).

The five factors (Originality, Elaboration, Flexibility, Aesthetic Quality, and Technical Execution) fit well with established creativity frameworks. Originality and Flexibility correspond to Torrance's classical conceptualization,[8] and Elaboration reflects the depth stressed in Amabile's componential model.[4] Aesthetic Quality and Technical Execution go beyond the traditional divergent thinking dimensions. Evaluating tangible creative products makes practical demands of this kind.[3,20] The multidimensional solution agrees with Besemer and O'Quin, who confirmed a multidimensional structure for creative products,[5] and with the meta-confirmatory factor analysis of Said-Metwaly et al., which favored multidimensional models over unidimensional ones.[8] Leaving "usefulness" out as a separate dimension was a theoretically grounded choice. The construct is covered by Elaboration (practical development of ideas) and Technical Execution (functional quality), in line with how educational assessors implicitly judge creative work.[11,20]

4.2 Reliability and criterion validity

Internal consistency (α = 0.93, ω = 0.94) is higher than that of many existing creativity assessment instruments. Reiter-Palmon et al. observed that many instruments reach reliabilities below 0.80,[7] which implies that the rubric-based approach of the COAR, with its clear behavioral anchors, may carry a measurement advantage. For context, inter-rater reliability (ICC = 0.84) is comparable to that reported for recent large-language-model-based automated creativity scoring by Organisciak et al.,[21] although the two approaches differ in what they measure, and it is above the level usually reported for holistic creativity ratings.[9,10] A structured format may help to limit rater subjectivity.[6,15]

The criterion validity coefficient (r = 0.74) lies above the r = 0.45–0.60 range usually reported for creativity tests against expert evaluations.[14,22] The expert panel that served as the criterion was separate from the panel that developed the instrument. This reduces the risk of circularity in the validation.[17] Convergent validity (AVE = 0.51–0.58) and discriminant validity (Fornell-Larcker and HTMT criteria met) add further support for construct validity.

4.3 Theoretical, cultural, and practical implications

In terms of educational theory, the COAR operationalizes the highest level of Bloom's revised taxonomy, "create," through the Originality and Flexibility dimensions. Elaboration and Technical Execution capture lower-order but essential skills of application and analysis.[3] In line with Vygotsky's zone of proximal development, the rubric scaffolds students' grasp of what makes creative work good. It makes implicit assessment criteria explicit and so supports learning.[13] The COAR also fits the principles of authentic assessment, since it evaluates artifacts that students actually produced in genuine educational contexts and not under artificial test conditions.[1,23]

Cultural context may also matter. Indonesian collectivist values, in which group harmony and respect for authority may shape how originality is expressed and evaluated, could influence creativity assessment.[24] Configural measurement invariance across four faculties gives initial evidence that the factor structure of the COAR holds across disciplinary contexts. Full scalar invariance, however, has yet to be tested.

The practical implications are considerable. A mean scoring time of 14.6 minutes per project makes the COAR workable in the classroom. Because the five dimensions are modular, instructors can give formative feedback and pinpoint particular areas of creative strength and weakness. A scoring manual with exemplars was prepared to support rater training. It needs about two hours of preparation. Applicability across the four faculties studied (Education, Arts and Design, Sciences, Social Sciences), supported at the configural level only, points to potential utility in higher education more widely.[23,25]

4.4 Strengths and limitations

The study has several clear strengths. Its multi-phase design, which follows established procedures,[16,17] yields wide-ranging validity evidence. Separate EFA and CFA samples support the independence of the confirmatory test. The range of validity types examined (content, construct, convergent, discriminant, and criterion), together with the measurement invariance testing, is broader than that of many published educational rubrics. Limitations also need to be acknowledged. First, the study took place at a single private university in Indonesia, which restricts cross-cultural generalizability. Second, predictive validity linking COAR scores to later creative achievements was not assessed. Third, full measurement invariance across demographic groups has not yet been tested. Finally, all AVE values (0.51–0.58) exceeded the 0.50 threshold but reflect borderline convergent validity. For that reason, the Fornell-Larcker and HTMT criteria were relied upon as complementary evidence for discriminant validity.

5. Conclusion

The Creative Output Assessment Rubric (COAR) was developed in this study, and evidence for its validity was gathered. It is a 22-item, five-factor instrument for evaluating student creative projects in higher education, covering Originality, Elaboration, Flexibility, Aesthetic Quality, and Technical Execution. Its psychometric properties were good. Content validity was high (S-CVI/Ave = 0.92), CFA model fit was good (CFI = 0.94, RMSEA = 0.048), internal consistency was strong (α = 0.93), inter-rater reliability was adequate (ICC = 0.84, 95% CI: 0.79–0.88), and criterion validity was significant (r = 0.74, p < 0.001). Universities and their faculty may consider piloting the COAR as an assessment tool once the evidence has been replicated in other settings. Building it into institutional quality-assurance frameworks could then make creativity evaluation more objective and consistent. Future research should address cross-cultural validity, measurement invariance across demographic groups, predictive validity for long-term outcomes, and technology-enhanced scoring applications.

Declaration

Artificial Intelligence Use Statement

No generative artificial intelligence tool was used to write or revise the text, analyze the data, create images, or generate any other scholarly content of this manuscript.

Author Contribution Statement

Conceptualization, V.A.; methodology, V.A. and S.L.R.; formal analysis, V.A.; investigation, V.A.; data curation, V.A.; writing – original draft, V.A.; writing – review and editing, V.A. and S.L.R.; supervision, S.L.R.; project administration, V.A. All authors read and approved the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest / Competing Interests

The authors declare that they have no competing financial or non-financial interests related to this work.

Ethics Approval and Consent to Participate

The study was approved by the Ethical Committee of Social Research, Enigma Institute Indonesia (approval number 2023/189). Participation was voluntary, informed consent was obtained, and all data were anonymized before analysis.

Consent for Publication

Not applicable. No identifiable personal information is reported.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author on reasonable request.

Acknowledgments

None.

References

1. Lucas B. A five-dimensional model of creativity and its assessment in schools. Appl Meas Educ 2016;29(4):278-290. doi:10.1080/08957347.2016.1209206
2. Gajda A, Karwowski M, Beghetto RA. Creativity and academic achievement: a meta-analysis. J Educ Psychol 2017;109(2):269-299. doi:10.1037/edu0000133
3. Patston TJ, Kaufman JC, Cropley AJ, et al. What is creativity in education? A qualitative study of international curricula. J Adv Acad 2021;32(2):207-230. doi:10.1177/1932202X20978356
4. Barbot B, Besancon M, Lubart T. Creative potential in educational settings: its nature, measure, and nurture. Education 3-13 2015;43(4):371-381. doi:10.1080/03004279.2015.1020643
5. Besemer SP, O'Quin K. Confirming the three-factor Creative Product Analysis Matrix model in an American sample. Creat Res J 1999;12(4):287-296. doi:10.1207/s15326934crj1204_6
6. Said-Metwaly S, Van den Noortgate W, Kyndt E. Approaches to measuring creativity: a systematic literature review. Creat Theor Res Appl 2017;4(2):238-275. doi:10.1515/ctra-2017-0013
7. Reiter-Palmon R, Forthmann B, Barbot B. Scoring divergent thinking tests: a review and systematic framework. Psychol Aesthet Creat Arts 2019;13(2):144-152. doi:10.1037/aca0000227
8. Said-Metwaly S, Fernandez-Castilla B, Kyndt E, et al. The factor structure of the figural Torrance Tests of Creative Thinking: a meta-confirmatory factor analysis. Creat Res J 2018;30(4):352-360. doi:10.1080/10400419.2018.1530534
9. Zeng L, Proctor RW, Salvendy G. Can traditional divergent thinking tests be trusted in measuring and predicting real-world creativity? Creat Res J 2011;23(1):24-37.
   doi:10.1080/10400419.2011.545713
10. Forthmann B, Szardenings C, Holling H. Understanding the confounding effect of fluency in divergent thinking scores: revisiting average scores to quantify artifactual correlation. Psychol Aesthet Creat Arts 2020;14(1):94-112. doi:10.1037/aca0000196
11. Acar S, Burnett C, Cabra JF. Ingredients of creativity: originality and more. Creat Res J 2017;29(2):133-144.
    doi:10.1080/10400419.2017.1302776
12. Kaufman JC, Beghetto RA. Beyond big and little: the Four C model of creativity. Rev Gen Psychol 2009;13(1):1-12. doi:10.1037/a0013688
13. Panadero E, Jonsson A. The use of scoring rubrics for formative assessment purposes revisited: a review. Educ Res Rev 2013;9:129-144. doi:10.1016/j.edurev.2013.01.002
14. Cropley DH, Kaufman JC. Measuring functional creativity: non-expert raters and the Creative Solution Diagnosis Scale. J Creat Behav 2012;46(2):119-137. doi:10.1002/jocb.9
15. Cseh GM, Jeffries KK. A scattered CAT: a critical evaluation of the consensual assessment technique for creativity research. Psychol Aesthet Creat Arts 2019;13(2):159-166.
    doi:10.1037/aca0000220
16. DeVellis RF, Thorpe CT. Scale Development: Theory and Applications. 5th ed. Thousand Oaks: SAGE Publications; 2022.
17. Boateng GO, Neilands TB, Frongillo EA, et al. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149. doi:10.3389/fpubh.2018.00149
18. Beghetto RA. Creativity in classrooms. In: Kaufman JC, Sternberg RJ, editors. The Cambridge Handbook of Creativity. 2nd ed. Cambridge: Cambridge University Press; 2019. p. 587-606. doi:10.1017/9781316979839.029
19. Wilson M. Constructing Measures: An Item Response Modeling Approach. 2nd ed. New York: Routledge; 2023.
20. Cropley DH, Kaufman JC. The siren song of aesthetics? Domain differences and creativity in engineering and design. Proc Inst Mech Eng C J Mech Eng Sci 2019;233(2):451-464. doi:10.1177/0954406218778311
21. Organisciak P, Acar S, Dumas D, et al. Beyond semantic distance: automated scoring of divergent thinking greatly improves with large language models. Think Skills Creat 2023;49:101356. doi:10.1016/j.tsc.2023.101356
22. Haase J, Hoff EV, Hanel PHP, et al. A meta-analysis of the relation between creative self-efficacy and different creativity measurements. Creat Res J 2018;30(1):1-16.
    doi:10.1080/10400419.2018.1411436
23. Villarroel V, Bloxham S, Bruna D, et al. Authentic assessment: creating a blueprint for course design. Assess Eval High Educ 2018;43(5):840-854. doi:10.1080/02602938.2017.1412396
24. Niu W, Sternberg RJ. Contemporary studies on the concept of creativity: the East and the West. J Creat Behav 2002;36(4):269-288. doi:10.1002/j.2162-6057.2002.tb01069.x
25. An D, Youn N. The inspirational power of arts on creativity. J Bus Res 2018;85:467-475. doi:10.1016/j.jbusres.2017.10.025