e-ISSN: 3026-5827

Enigma in Education

Journal website: https://enigma.or.id/index.php/edu

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

Keywords:
Assessment rubric
Creativity
Higher education
Psychometric properties
Rubric validation

***Corresponding author:**
Vita Amanda

E-mail address:
vitaamandava@gmail.com

All authors have reviewed and approved the final version of the manuscript.

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

A B S T R A C T

Educators in higher education lack psychometrically validated instruments for assessing student creative output, despite creativity being recognized as a core 21st-century competency. This study aimed to develop and validate the Creative Output Assessment Rubric (COAR) for evaluating student creative projects in university settings. A multi-phase instrument development design was employed at a private university in Palembang, Indonesia, during the 2023/2024 academic year, comprising item generation and expert review (n = 8; S-CVI/Ave = 0.92), pilot testing (n = 45), exploratory factor analysis (n = 215), and confirmatory factor analysis with an independent sample (n = 220). A five-factor structure—Originality, Elaboration, Flexibility, Aesthetic Quality, and Technical Execution—comprising 22 items was confirmed, explaining 67.4% of total variance. The confirmatory model demonstrated good fit (CFI = 0.94; TLI = 0.93; RMSEA = 0.048). Internal consistency was strong (Cronbach's α = 0.93; McDonald's ω = 0.94), inter-rater reliability was adequate (ICC = 0.84; 95% CI: 0.79–0.88), and criterion validity was supported by a significant correlation with independent expert ratings (r = 0.74; p < 0.001). Configural measurement invariance held across four academic faculties. The COAR is a valid and reliable instrument for assessing student creative output in higher education, offering educators a validated scoring framework for the formative and summative evaluation of creative work across academic disciplines.

1. Introduction

Creativity has been widely recognized as one of the most essential 21st-century competencies that higher education institutions must cultivate and assess. The Partnership for 21st Century Learning (P21) framework and the OECD Learning Compass 2030 both identify creativity as a core competency for future-ready graduates.[1-3] Despite this recognition, educators consistently report significant difficulty in assessing student creative output in a manner that is both valid and reliable, often relying on subjective impressions rather than standardized instruments.[4-6]

The assessment of creativity in educational settings has traditionally relied on two approaches: divergent thinking tests measuring creative potential, and product-based assessments evaluating tangible creative outputs.[7-9] While divergent thinking tests such as the Torrance Tests of Creative Thinking have undergone extensive psychometric examination, product-based creativity assessment instruments remain comparatively underdeveloped.[8,10] Existing rubrics such as the VALUE rubrics developed by the Association of American Colleges and Universities provide general frameworks but lack comprehensive psychometric validation with factor analysis and criterion-related validity evidence.[5,6]

From a theoretical perspective, creativity is understood as a multidimensional construct. Amabile's componential model identifies domain-

relevant skills, creativity-relevant processes, and intrinsic task motivation as determinants of creative production.[4,11] Torrance's framework conceptualizes creative output through originality, fluency, flexibility, and elaboration.[8] Kaufman and Beghetto's Four-C model distinguishes between mini-c, little-c, Pro-c, and Big-C creativity, with mini-c and little-c being most relevant to educational assessment.[12] Bloom's revised taxonomy positions "create" as the highest cognitive level, underscoring the centrality of creative production in higher-order educational outcomes.[3] Furthermore, Vygotsky's zone of proximal development provides a theoretical basis for rubrics as scaffolding tools that make implicit creativity criteria explicit for learners.[13]

Despite advances in creativity theory, several critical gaps remain. First, most existing creativity measurement tools were developed for psychological research rather than practical classroom use.[6,9] Second, there is a scarcity of psychometrically validated rubrics specifically designed for evaluating creative project outputs in higher education.[5,14] Third, the few existing rubrics have rarely undergone rigorous validation including both exploratory and confirmatory factor analysis.[8,15] Fourth, criterion-related validity and inter-rater reliability evidence is often absent.[7,10]

The aim of this study was to develop and validate the Creative Output Assessment Rubric (COAR) for evaluating student creative projects in higher education, using a multi-phase design encompassing expert review, pilot testing, exploratory factor analysis, and confirmatory factor analysis. This study contributes to the educational assessment literature by providing a comprehensively validated rubric designed for educators to assess creative project output across multiple academic disciplines in university settings.

2. Methods

This study employed a multi-phase instrument development and validation design following established psychometric procedures.[16,17] The study was conducted at a private university in Palembang, Indonesia, during the 2023/2024 academic year. This study did not collect sensitive personal or health information. Participants were actively recruited and took part voluntarily after providing informed consent, and all data were anonymized prior to analysis. In accordance with institutional policy governing low-risk educational research, the study qualified for a documented institutional exemption from full research-ethics-committee review.

In Phase 1 (Item Generation and Content Validation), an initial pool of 30 items was developed based on a comprehensive literature review of creativity theory and existing assessment frameworks.[3,7,8,11,18] Items were organized 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), following a construct-mapping approach to item development. The exclusion of "usefulness" as a separate dimension was deliberate; this construct is operationalized within Elaboration (practical applicability of ideas) and Technical Execution (functional quality of output).[11,19] Content validity was evaluated by an expert panel of eight specialists (four creativity researchers, four higher education instructors). The Content Validity Index was calculated at item level (I-CVI ≥ 0.78 for retention) and scale level (S-CVI/Ave).[17]

In Phase 2 (Pilot Testing), the refined instrument was administered 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 included corrected item-total correlations (threshold ≥ 0.30) and reliability analysis. In Phase 3 (Exploratory Factor Analysis), the 22-item COAR was administered to 215 students from four faculties (Education, Arts and Design, Sciences, Social Sciences). Two trained raters independently scored each project. Principal axis factoring with

oblimin rotation was conducted; oblimin was selected over promax due to the expected correlated nature of creativity dimensions.[16] Factor retention was determined by parallel analysis and eigenvalues greater than 1.0.

In Phase 4 (Confirmatory Factor Analysis), an independent sample of 220 students from different cohorts was recruited. The five-factor model was tested using CFA with maximum likelihood estimation in AMOS version 26. Multivariate normality was assessed using Mardia's coefficient. Model fit was evaluated using chi-square/df (acceptable < 3.0), CFI (≥ 0.90), TLI (≥ 0.90), RMSEA (≤ 0.06), and SRMR (≤ 0.08).[16] Modification indices were examined but none were applied to preserve theoretical integrity. Reliability was assessed using Cronbach's alpha, McDonald's omega, and ICC (two-way mixed, consistency model, with individual rater values reported). Convergent validity was evaluated through AVE (≥ 0.50) and discriminant validity through the Fornell-Larcker criterion and the HTMT ratio (< 0.85). Criterion-related validity was assessed by correlating COAR scores 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. Practical utility was assessed by recording scoring time per project. All analyses were conducted in SPSS 26 and AMOS 26 at alpha = 0.05.

3. Results

Table 1 presents the demographic characteristics of participants in the CFA sample (n = 220). The sample comprised 92 males (41.8%) and 128 females (58.2%), with a mean age of 21.4 years (SD = 1.8). Students were drawn 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)
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).

In Phase 1, the expert panel evaluation yielded an S-CVI/Ave of 0.92, with all retained items achieving I-CVI values of 0.78 or higher. Eight items were removed, reducing the pool from 30 to 22 items. In Phase 2, pilot testing confirmed acceptable corrected item-total correlations (range: 0.38–0.72), with no additional items requiring removal.

In Phase 3, the Kaiser-Meyer-Olkin measure was 0.89, and Bartlett's test of sphericity was significant (x^2 = 2847.62, df = 231, p < 0.001). Principal axis factoring extracted five factors explaining 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%). All items loaded above 0.40 on their intended factors, with no problematic cross-loadings.

Table 2 displays the CFA model fit indices and reliability coefficients. The five-factor model demonstrated good fit: x^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) was below the critical threshold, supporting the use of maximum likelihood estimation. This model showed substantially better fit than the one-factor (x^2/df = 4.56, CFI = 0.72, RMSEA = 0.112) and three-factor (x^2/df = 2.91, CFI = 0.84, RMSEA = 0.078) alternatives, as well as 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		
x^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.

All 22 items demonstrated standardized factor loadings ranging from 0.66 to 0.85, well above the 0.40 threshold, as illustrated in Figure 1. The highest loadings were observed for Flexibility item F3 (0.85) and Originality item O2 (0.82), whereas the lowest acceptable loading was for Flexibility item F4 (0.66).

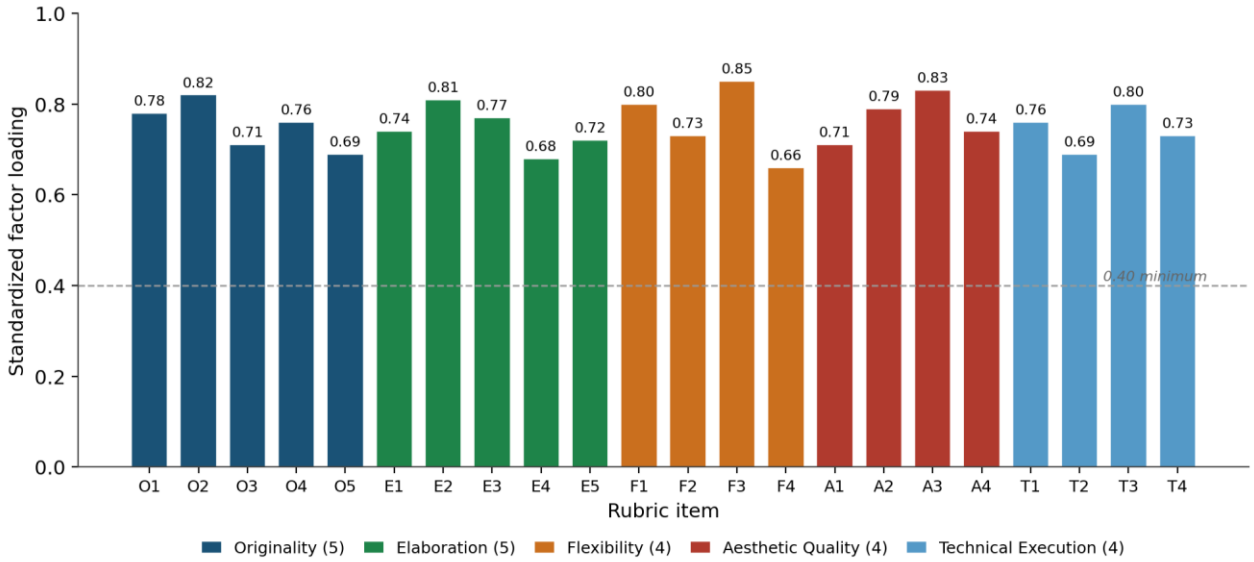

Figure 1. Standardized factor loadings from confirmatory factor analysis of the COAR (N = 220). The dashed line marks the 0.40 minimum retention threshold; bars are grouped and colored by dimension.

A comparison of model fit across the alternative CFA models confirms the superiority of the five-factor solution (Figure 2). For the final model, both CFI (0.94) and TLI (0.93) exceeded the 0.90 threshold and RMSEA (0.048) fell below the 0.06 criterion, whereas the one- and three-factor models failed to meet these benchmarks.

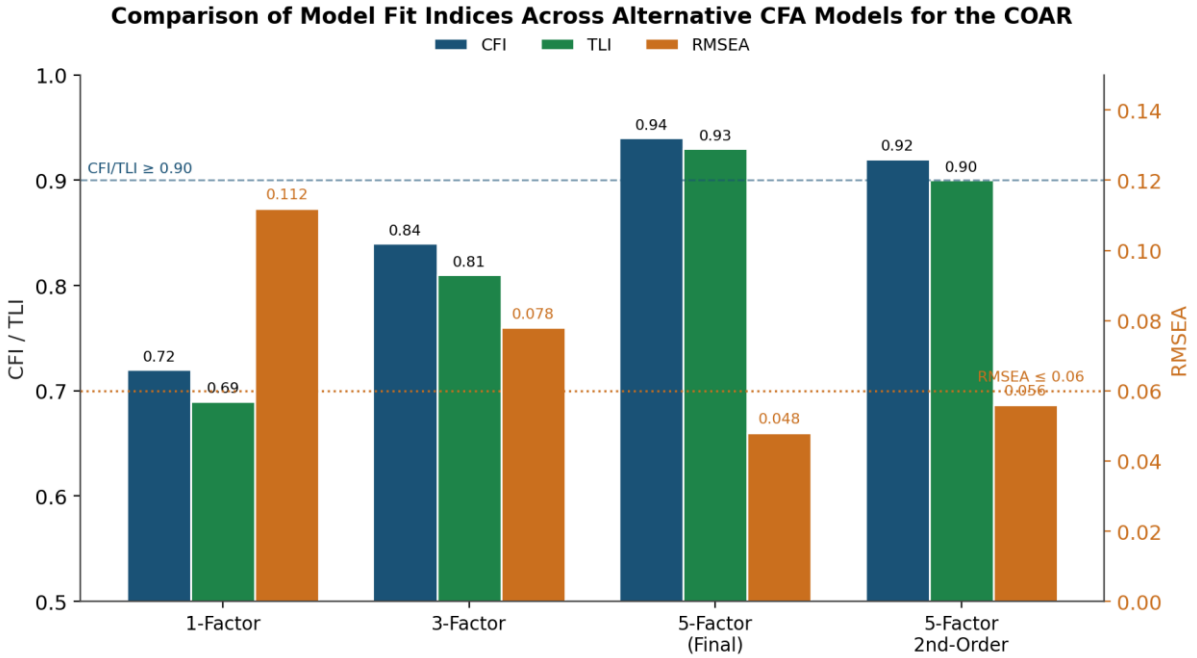

Figure 2. Comparison of model fit indices across alternative CFA models for the COAR. CFI and TLI are read on the left axis; RMSEA is read on the right axis.

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 presents the convergent and discriminant validity evidence. AVE values ranged from 0.51 to 0.58, all exceeding 0.50. The square root of AVE for each factor exceeded all corresponding inter-factor correlations, confirming discriminant validity. HTMT ratios ranged from 0.58 to 0.82, all below the 0.85 threshold.

Table 3. Convergent and discriminant validity: AVE and inter-factor correlations.

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. Off-diagonal values are inter-factor correlations. Discriminant validity is supported when √AVE exceeds the corresponding off-diagonal correlations.

Criterion validity, assessed in the validation sample (n = 215), revealed a significant positive correlation between COAR total scores and independent expert panel ratings (r = 0.74, p < 0.001, 95% CI: 0.67–0.80, R^2 = 0.55), as shown in Figure 3. Configural measurement invariance was supported across the four faculties (ΔCFI < 0.01 between the unconstrained and constrained models). Mean scoring time was 14.6 minutes per project (SD = 3.2).

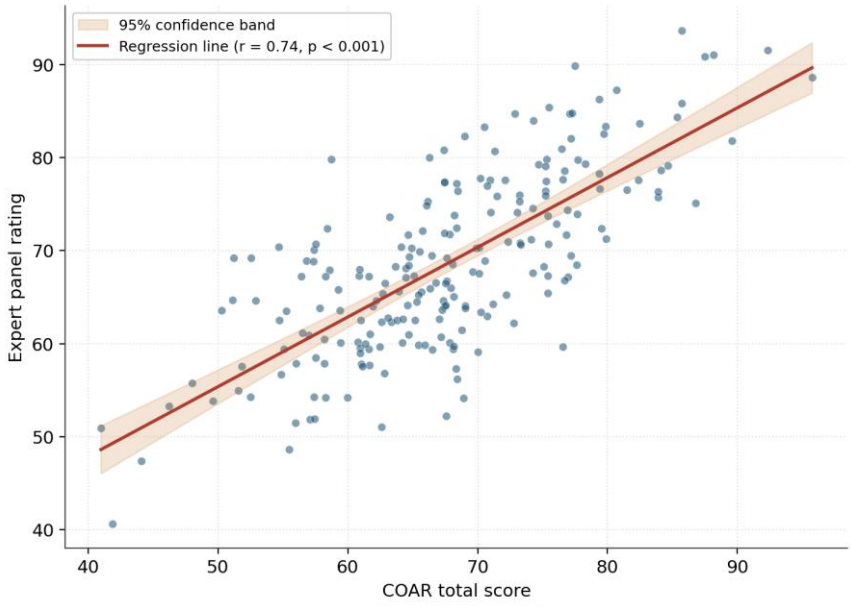

Figure 3. COAR total scores against expert panel ratings: criterion validity (N = 215, r = 0.74, p < 0.001). The shaded region is the 95% confidence band around the regression line.

4. Discussion

This study developed and validated the Creative Output Assessment Rubric (COAR), a 22-item, five-factor instrument for evaluating student creative projects in higher education. The principal findings demonstrated strong psychometric properties: excellent content validity (S-CVI/Ave = 0.92), a robust five-factor structure confirmed by both EFA and CFA, strong internal consistency (α = 0.93, ω = 0.94), adequate inter-rater reliability (ICC = 0.84), and significant criterion-related validity (r = 0.74).

The five-factor structure—Originality, Elaboration, Flexibility, Aesthetic Quality, and Technical Execution—aligns well with established creativity frameworks. The Originality and Flexibility dimensions correspond to Torrance's classical conceptualization,[8] while Elaboration captures the depth emphasized in Amabile's componential model.[4] Aesthetic Quality and Technical Execution extend beyond traditional divergent thinking dimensions, reflecting the practical demands of evaluating tangible creative

products.[3,19] This multidimensional solution is consistent with Besemer and O'Quin, who confirmed a multidimensional structure for creative products,[5] and with Said-Metwaly et al.'s meta-confirmatory factor analysis supporting multidimensional models over unidimensional solutions.[8] The deliberate exclusion of "usefulness" as a standalone dimension was theoretically grounded: this construct is captured within Elaboration (practical development of ideas) and Technical Execution (functional quality), consistent with how educational assessors implicitly evaluate creative work.[11,19]

The strong internal consistency (α = 0.93, ω = 0.94) exceeds that of most existing creativity assessment instruments. Reiter-Palmon et al. noted that many instruments achieve reliabilities below 0.80,[7] suggesting that the COAR's rubric-based approach with clear behavioral anchors offers measurement advantages. The inter-rater reliability (ICC = 0.84) compares favorably with the reliability reported for recent large-language-model-based automated creativity scoring by Organisciak et al.,[20] and is substantially higher than that of typical holistic creativity ratings.[9,10] This suggests that the COAR's structured format effectively reduces rater subjectivity.[6,15]

The criterion validity (r = 0.74) is notably higher than the r = 0.45–0.60 range typically reported between creativity tests and expert evaluations.[14,21] Importantly, the criterion expert panel was independent from the instrument development panel, eliminating potential circularity in validation.[17] The convergent validity (AVE = 0.51–0.58) and discriminant validity (Fornell-Larcker and HTMT criteria met) further strengthen the construct validity evidence.

From an educational theory perspective, the COAR operationalizes the highest level of Bloom's revised taxonomy—"create"—through the Originality and Flexibility dimensions, while Elaboration and Technical Execution capture lower-order but essential skills of application and analysis.[3] Consistent with Vygotsky's zone of proximal development, the rubric scaffolds students' understanding of what constitutes quality creative work, making implicit assessment criteria explicit and thereby supporting learning.[13] The COAR also aligns with principles of authentic assessment, as it evaluates actual student-produced creative artifacts within genuine educational contexts rather than artificial test conditions.[1,22]

Regarding cultural context, creativity assessment may be influenced by Indonesian collectivist values, in which group harmony and respect for authority may shape how originality is expressed and evaluated.[23] The configural measurement invariance across four faculties provides initial evidence that the COAR's factor structure holds across disciplinary contexts, though full scalar invariance remains to be tested.

The practical implications are substantial. The mean scoring time of 14.6 minutes per project makes the COAR feasible for classroom use. The modular five-dimension structure enables formative feedback, allowing instructors to identify specific areas of creative strength and weakness. A scoring manual with exemplars was developed to facilitate rater training, requiring approximately two hours of preparation. The cross-disciplinary applicability across four faculties (Education, Arts and Design, Sciences, Social Sciences) suggests broad utility in higher education.[22,24,25]

This study has several notable strengths. The multi-phase design following established procedures[16,17] provides comprehensive validity evidence. The use of separate EFA and CFA samples strengthens generalizability. The inclusion of multiple validity types (content, construct, convergent, discriminant, and criterion) and measurement invariance testing exceeds published standards for educational rubrics. Several limitations should also be acknowledged. First, the study was conducted at a single private university in Indonesia, which limits cross-cultural generalizability. Second, predictive validity linking COAR scores to future creative achievements was

not assessed. Third, full measurement invariance across demographic groups remains to be tested. Finally, although all AVE values (0.51–0.58) exceeded the 0.50 threshold, they represent borderline convergent validity; accordingly, the Fornell-Larcker and HTMT criteria were relied upon to provide complementary discriminant validity evidence.

5. Conclusion

This study successfully developed and validated the Creative Output Assessment Rubric (COAR), a 22-item, five-factor instrument comprising Originality, Elaboration, Flexibility, Aesthetic Quality, and Technical Execution for evaluating student creative projects in higher education. The COAR demonstrated excellent psychometric properties: content validity (S-CVI/Ave = 0.92), good CFA model fit (CFI = 0.94, RMSEA = 0.048), strong internal consistency (α = 0.93), adequate inter-rater reliability (ICC = 0.84, 95% CI: 0.79–0.88), and significant criterion validity ($r = 0.74$, $p < 0.001$). Higher education institutions and faculty are encouraged to adopt the COAR as a standardized assessment tool, integrating it into institutional quality-assurance frameworks to strengthen the objectivity and consistency of creativity evaluation. Future research should examine cross-cultural validity, measurement invariance across demographic groups, predictive validity for long-term outcomes, and technology-enhanced scoring applications.

6. References

1. Lucas B. A five-dimensional model of creativity and its assessment in schools. Appl Meas Educ 2016;29(4):278-290. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037/aca0000196
11. Acar S, Burnett C, Cabra JF. Ingredients of creativity: originality and more. Creat Res J 2017;29(2):133-144. https://doi.org/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. https://doi.org/10.1037/a0013688
13. Bereczki EO, Karpati A. Technology-enhanced creativity: a multiple case study of digital technology-integration expert teachers' beliefs and practices. Think Skills Creat 2021;39:100791. https://doi.org/10.1016/j.tsc.2021.100791
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. https://doi.org/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. https://doi.org/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, Melgar-Quinonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149. https://doi.org/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. https://doi.org/10.1017/9781316979839.029
19. 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. https://doi.org/10.1177/0954406218778311
20. 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. https://doi.org/10.1016/j.tsc.2023.101356
21. 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. https://doi.org/10.1080/10400419.2018.1411436
22. 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. https://doi.org/10.1080/02602938.2017.1412396
23. Jankowska DM, Karwowski M. Measuring creative imagery abilities. Front Psychol 2015;6:1591. https://doi.org/10.3389/fpsyg.2015.01591

24. An D, Youn N. The inspirational power of arts on creativity. J Bus Res 2018;85:467-475. https://doi.org/10.1016/j.jbusres.2017.10.025

25. Wilson M. Constructing Measures: An Item Response Modeling Approach. 2nd ed. New York: Routledge; 2023.