http://dx.doi.org/10.18623/rvd.v23.5885

DEVELOPING AND VALIDATING A SECOND-ORDER REFLECTIVE FRAMEWORK FOR SCIENTIFIC CREATIVITY COMPETENCE IN HIGH SCHOOL STUDENTS: A DELPHI-SEM MIXED-METHODS APPROACH

DESENVOLVIMENTO E VALIDAÇÃO DE UM PARADIGMA REFLEXIVO DE SEGUNDA ORDEM PARA A COMPETÊNCIA EM CRIATIVIDADE CIENTÍFICA EM ALUNOS DO ENSINO MÉDIO: UMA ABORDAGEM DE MÉTODOS MISTOS DELPHI-SEM

Article received on: 12/24/2025
Article accepted on: 3/25/2026

Le Quang Chau*
***University of Science and Education - The University of Danang, Da Nang, Vietnam
Orcid: https://orcid.org/0009-0002-1401-7041
lequangchau@caothang.edu.vn

Nguyen Van Hoanh*
***University of Science and Education - The University of Danang, Da Nang, Vietnam
Orcid: https://orcid.org/0009-0006-1778-8465
nguyenvanhoanh@gmail.com

Phung Viet Hai*
***University of Science and Education - The University of Danang, Da Nang, Vietnam
Orcid: https://orcid.org/0009-0002-8604-9712
pvhai@ued.udn.vn

The authors declare that there is no conflict of interest

Abstract

Background and Purpose: Scientific creativity competence (SCC) is recognized as a core higher-order competence in contemporary science education, yet validated frameworks that specify observable behavioral indicators within culturally situated, physics-based STEM learning contexts remain underdeveloped. The present study aimed to develop and validate a second-order reflective SCC measurement framework for Vietnamese upper-secondary students. Design and Methods: A sequential mixed-methods design was employed across five procedural steps: (1) systematic bibliometric synthesis of 690 Scopus-indexed publications (1960-2026); (2) second-order reflective model specification; (3) two-round Delphi expert consultation (Round 1: N = 42; Round 2: N = 38, retention rate 90.5%); (4) PLS-SEM exploratory validation (n = 300, SmartPLS 4, 5,000 bootstrap resamples); and (5) CB-SEM confirmatory validation incorporating EFA and CFA (n = 1,200, SPSS 27 and AMOS 27). Quantitative participants were Grade 10 students recruited from schools in Ho Chi Minh City and Gia Lai Province via stratified sampling. Results: The proposed framework - comprising four first-order components (PRB: problem recognition;

Resumo

Antecedentes e Objetivo: A competência em criatividade científica (SCC) é reconhecida como uma competência fundamental de ordem superior no ensino de ciências contemporâneo; no entanto, os modelos validados que especificam indicadores comportamentais observáveis em contextos de aprendizagem STEM baseados na física e situados culturalmente continuam pouco desenvolvidos. O presente estudo teve como objetivo desenvolver e validar um modelo de avaliação reflexiva de segunda ordem da SCC para alunos do ensino médio vietnamitas. Desenho e métodos: Foi empregado um desenho sequencial de métodos mistos em cinco etapas procedimentais: (1) síntese bibliométrica sistemática de 690 publicações indexadas no Scopus (1960-2026); (2) especificação do modelo reflexivo de segunda ordem; (3) consulta a especialistas por Delphi em duas rodadas (Rodada 1: N = 42; Rodada 2: N = 38, taxa de retenção de 90,5%); (4) validação exploratória PLS-SEM (n = 300, SmartPLS 4, 5.000 reamostragens bootstrap); e (5) validação confirmatória CB-SEM incorporando EFA e CFA (n = 1.200, SPSS 27 e AMOS 27) . Os participantes quantitativos foram alunos do 10º

IDE: idea generation; DES: solution design; EVA: solution evaluation) operationalized through 16 behavioral indicators - achieved expert consensus rates of 86.8%-92.1% in Round 2. PLS-SEM confirmed adequate internal consistency (Cronbach's $\alpha = 0.752$-0.792; CR = 0.843-0.865), convergent validity (AVE = 0.575-0.617), and discriminant validity (HTMT = 0.605-0.771). CFA demonstrated excellent model fit - $\chi^2(100) = 97.892$, $p = 0.541$; $\chi^2/df = 0.979$; GFI = 0.990; CFI = 1.000; TLI = 1.000; RMSEA = 0.000; SRMR = 0.017; PCLOSE = 1.000 - and confirmed the second-order reflective structure, with DES and PRB as the dominant reflections of SCC ($\beta = 0.85$; $R^2 = 0.72$), followed by IDE ($\beta = 0.74$) and EVA ($\beta = 0.71$). All three research hypotheses were supported. Conclusions and Implications: The validated SCC framework constitutes, to the best of the authors' knowledge, the first psychometrically rigorous second-order measurement model for scientific creativity competence developed and normed within the Vietnamese secondary education context. The five-step Delphi-SEM procedure offers a replicable protocol for competence framework development in educational measurement. The framework provides a theoretically grounded foundation for designing assessment instruments and SCC-targeted STEM learning sequences in Physics instruction at the upper-secondary level.

Keywords: Scientific Creativity Competence. Second-Order Reflective Model. Delphi Method. PLS-SEM. CB-SEM. STEM Education.

ano recrutados em escolas da Cidade de Ho Chi Minh e da Província de Gia Lai por meio de amostragem estratificada. Resultados: A estrutura proposta — composta por quatro componentes de primeira ordem (PRB: reconhecimento do problema; IDE: geração de ideias; DES: projeto da solução; EVA: avaliação da solução) operacionalizados por meio de 16 indicadores comportamentais — alcançou taxas de consenso entre especialistas de 86,8% a 92,1% na Rodada 2. A PLS-SEM confirmou consistência interna adequada (α de Cronbach = 0,752-0,792; CR = 0,843-0,865), validade convergente (AVE = 0,575-0,617) e validade discriminante (HTMT = 0,605-0,771). A CFA demonstrou excelente ajuste do modelo - $\chi^2(100) = 97{,}892$, $p = 0{,}541$; $\chi^2/df = 0{,}979$; GFI = 0,990; CFI = 1,000; TLI = 1,000; RMSEA = 0,000; SRMR = 0,017; PCLOSE = 1,000 — e confirmou a estrutura reflexiva de segunda ordem, com DES e PRB como reflexões dominantes da SCC ($\beta = 0{,}85$; $R^2 = 0{,}72$), seguidas por IDE ($\beta = 0{,}74$) e EVA ($\beta = 0{,}71$). Todas as três hipóteses de pesquisa foram confirmadas. Conclusões e implicações: O quadro validado de SCC constitui, até onde é do conhecimento dos autores, o primeiro modelo de medição de segunda ordem psicometricamente rigoroso para a competência em criatividade científica desenvolvido e padronizado no contexto do ensino médio vietnamita. O procedimento Delphi-SEM de cinco etapas oferece um protocolo replicável para o desenvolvimento de quadros de competências na avaliação educacional. O quadro fornece uma base teórica sólida para a concepção de instrumentos de avaliação e sequências de aprendizagem STEM voltadas para a SCC no ensino de Física no nível do ensino médio.

Palavras-chave: *Competência em Criatividade Científica. Modelo Reflexivo de Segunda Ordem. Método Delphi. PLS-SEM. CB-SEM. Educação STEM.*

1 INTRODUCTION

The rapid advancement of artificial intelligence and the transformative pressures of the Fourth Industrial Revolution have fundamentally altered the knowledge demands placed on new generations of learners (World Economic Forum, 2016; UNESCO, 2023).

In this context, scientific knowledge is no longer a stable repository to be transmitted but a dynamic system requiring continual renewal, application, and generation. Consequently, science education can no longer limit its ambition to equipping students with accumulated knowledge; it must orient learning toward the formation of higher-order competencies, particularly scientific creativity competence (SCC) - broadly understood as the capacity to produce novel, appropriate solutions through science-specific cognitive processes (Eyring, 1959; Hu & Adey, 2002; Kaufman & Sternberg, 2019).

The imperative to develop SCC is explicitly recognized at the policy level in Vietnam. The National General Education Curriculum 2018 identifies "problem-solving and creativity" as a core general competence to be developed across all subjects and educational activities (Ministry of Education and Training [MOET], 2018). The Resolution No. 29-NQ/TW on fundamental and comprehensive educational reform mandates a decisive shift from content-transmission toward competence-based learning (Communist Party Central Committee, 2013), while the National Strategy on Artificial Intelligence Development to 2030 underscores the urgency of building scientific and creative foundations among the youth population (Prime Minister, 2021, 2022). The cited policy documents converge on a shared educational obligation: to develop in high school students the capacity not only to understand scientific knowledge but to apply and extend it creatively in complex, real-world contexts.

Within physics education specifically, SCC manifests as a structured cognitive process encompassing problem identification, hypothesis formation, solution design, experimental testing, and iterative refinement - a cycle that mirrors authentic scientific inquiry (Eyring, 1959; Saefan et al., 2026). STEM education, organized around the Engineering Design Process (EDP), provides a particularly productive pedagogical environment for this development, as it engages learners in cross-disciplinary problem-solving that simultaneously demands divergent ideation and convergent evaluation (Aguilera & Ortiz-Revilla, 2021; Pinar et al., 2025). The Grade 10 Physics content on Dynamics - characterized by high levels of mathematical modeling and rich real-world application contexts - is especially well suited to STEM-integrated instruction aimed at developing SCC.

Despite growing policy attention and pedagogical interest, three substantive research gaps remain. First, existing SCC frameworks either operate at a high level of theoretical abstraction or are developed for cultural and educational contexts that differ significantly from Vietnamese secondary schooling (Hu & Adey, 2002; Xu et al., 2024). Observable behavioral indicators tied to specific learning tasks in the physics classroom remain largely underdeveloped. Second, no study in the Vietnamese context has subjected an SCC framework to a rigorous multi-stage validation sequence combining expert consensus (Delphi) with structural equation modeling (SEM). Isolated applications of either Delphi or SEM exist in the educational measurement literature (Hair et al., 2019; Lucas, 2016), but their systematic integration into a five-step framework development process has not been reported for SCC in Vietnam. Third, the theoretical modeling choice between reflective and formative measurement structures has rarely been made explicit or justified in SCC research, resulting in conceptual ambiguity regarding the relationship between the latent construct and its observable indicators (Hair et al., 2019).

To address these gaps, the present study pursues two interconnected objectives: (1) to develop a theoretically grounded, four-component SCC framework - comprising problem recognition (PRB), idea generation (IDE), solution design (DES), and solution evaluation (EVA) - operationalized through 16 behavioral indicators; and (2) to validate this framework as a second-order reflective measurement model through a sequential mixed-methods procedure combining two-round Delphi expert consultation with PLS-SEM (exploratory; n = 300) and CB-SEM/CFA (confirmatory; n = 1,200), conducted with Vietnamese high school students. Accordingly, two research questions guide the study:

RQ1: What components and behavioral indicators constitute the SCC framework for Vietnamese high school students?

RQ2: Does the proposed SCC framework achieve expert consensus and demonstrate adequate fit with empirical data drawn from Vietnamese high school students?

The study contributes to the international literature on science education assessment by offering a methodologically rigorous, contextually situated SCC framework that is directly applicable to the design of learning activities and assessment instruments in Physics-based STEM instruction at the secondary level.

2 METHODS

2.1 Research design

The study employed a sequential mixed-methods design, progressing through five procedural steps: (1) systematic literature synthesis to construct the theoretical framework; (2) second-order reflective model specification; (3) two-round Delphi expert consultation for content validity; (4) exploratory quantitative validation via PLS-SEM; and (5) confirmatory validation via CB-SEM incorporating EFA and CFA. The sequential design follows the logic of construct development in educational measurement - establishing theoretical and content validity before proceeding to structural and empirical validation (Hair et al., 2019; Lucas, 2016). The sequential structure ensures that each stage produces evidence that informs and constrains the next, minimizing the risk of measurement artifacts propagating through the validation process.

2.2 Step 1 - systematic literature synthesis and theoretical framework construction

Relevant literature was retrieved from the Scopus database using in February, 2026 using “scientific creativity” as the primary search term, restricted to titles, abstracts, and keywords, covering the period 1960-2026 and limited to journal articles and reviews in English. After removing duplicates and irrelevant records, 690 publications were retained for analysis. Bibliometric analysis was conducted in RStudio using the *Bibliometrix* package (Aria & Cuccurullo, 2017), supplemented by keyword co-occurrence mapping in VOSviewer. Analyses included annual publication trends, institutional and national output distributions, citation impact profiles, and thematic clustering via co-word analysis.

The synthesis identified four recurrent cognitive phases in SCC across multiple theoretical traditions - problem recognition, idea generation, solution design, and solution evaluation - consistent with both the Scientific Creativity Structure Model (Hu & Adey, 2002) and the Engineering Design Process employed in STEM education (Aguilera & Ortiz-Revilla, 2021). These four phases were adopted as the first-order components of the proposed SCC framework. Sixteen behavioral indicators (four per component) were then formulated by mapping each component to observable student actions in the context of physics-based STEM learning tasks, anchored to the cognitive developmental characteristics of high school students (Benedek & Fink, 2019; Hu & Adey, 2002).

2.3 Step 2 - second-order reflective model specification

The SCC framework was specified as a second-order reflective measurement model, in which the latent variable SCC is reflected by four first-order constructs - PRB (problem recognition), IDE (idea generation), DES (solution design), and EVA (solution evaluation) - each of which is in turn reflected by four observed indicators (PRB1-PRB4, IDE1-IDE4, DES1-DES4, EVA1-EVA4), yielding a total of 16 observed variables. The reflective specification was chosen because SCC is conceptualized as a latent disposition: the behavioral indicators are manifestations of the underlying construct rather than its constitutive parts, and a change in the level of SCC is expected to produce commensurate changes across all components and indicators (Jarvis et al., 2003; Hair et al., 2019). A formative specification would imply that indicators are independent and interchangeable, which is inconsistent with the system-theoretic basis of the SCC construct (Hu & Adey, 2002).

2.4 Step 3 - two-round delphi expert consultation

2.4.1 Expert panel composition

The Delphi panel was drawn from three professional groups: (i) academics and researchers in science education; (ii) specialists in educational measurement and assessment; and (iii) experienced high school science teachers with demonstrated competence in competence-based instruction or student research supervision. The tripartite panel composition was designed to secure perspectives on theoretical rigor, measurement feasibility, and classroom practicality simultaneously.

Round 1 included 42 panelists (science education: n = 18; measurement and assessment: n = 11; high school teachers: n = 13); of these, 38 continued to Round 2 (retention rate: 90.5%). The proportion holding doctoral qualifications was 64.3% in

Round 1 and 68.4% in Round 2; mean professional experience was 12.8 years (SD = 5.4) and 13.1 years (SD = 5.2), respectively.

2.4.2 Instrument and consensus criteria

The Round 1 questionnaire comprised three sections: (a) panelist professional background; (b) rating of the four first-order components; and (c) rating of each behavioral indicator on four criteria - *clarity*, *relevance*, *representativeness*, and *feasibility in the high school classroom context* - each scored on a five-point Likert scale (1 = *entirely inappropriate*; 5 = *entirely appropriate*), with open-ended prompts inviting qualitative feedback. The Round 2 questionnaire presented the revised framework with explicit notation of all changes made in response to Round 1 feedback, and asked panelists to re-evaluate the modified indicators.

Consensus was operationalized as $M \geq 4.0$ and a consensus rate (percentage of panelists scoring 4 or 5) of $\geq 75\%$ for a given indicator. Indicators approaching the threshold or exhibiting high dispersion ($IQR > 1.0$) were revised and resubmitted in Round 2 (Hsu & Sandford, 2007); indicators failing to reach consensus without a defensible theoretical rationale were considered for removal. Quantitative data were summarized using descriptive statistics (M, Md, SD, IQR, consensus rate); qualitative responses from the open-ended items were analyzed inductively to identify recurring themes and to inform precise revision of indicator wording (Endacott et al., 1999; Diamond et al., 2014).

2.5 Step 4 - PLS-SEM exploratory validation (round 1; n = 300)

A convenience sample of 300 Grade 10 students from three high schools in Ho Chi Minh City was recruited for the exploratory validation phase. Students completed a 16-item questionnaire with a five-point Likert response scale. Data were analyzed using *SmartPLS 4* to evaluate: (a) internal consistency reliability (Cronbach's $\alpha \geq 0.70$; composite reliability $CR \geq 0.70$); (b) convergent validity (average variance extracted $AVE \geq 0.50$); and (c) discriminant validity (heterotrait-monotrait ratio $HTMT < 0.85$). The statistical significance of factor loadings and structural paths was assessed via bootstrapping with 5,000 resamples. Indicators with standardized loadings below 0.50 or whose removal would substantially improve AVE were candidates for revision prior to the confirmatory phase (Henseler et al., 2015; Hair et al., 2019).

2.6 Step 5 - CB-SEM confirmatory validation (round 2; n = 1,200)

2.6.1 Participants

A stratified sample of 1,200 Grade 10 students was recruited from high schools in Ho Chi Minh City and Gia Lai Province, stratified by geographic location (urban vs. non-urban) to increase the representativeness of the normative data. The sample size of 1,200 satisfies the recommended minimum ratio of 10 observations per estimated parameter for CFA models of this complexity (Hair et al., 2019).

2.6.2 Exploratory factor analysis

EFA was conducted in SPSS 27 using principal axis factoring with promax rotation to examine the emergent factor structure prior to imposing the theoretically specified structure. Retention criteria were eigenvalue > 1, pattern coefficient ≥ 0.50, absence of cross-loadings ≥ 0.30, and total variance explained ≥ 50% (Hair et al., 2019). The suitability of the correlation matrix for factoring was confirmed prior to extraction.

2.6.3 Confirmatory factor analysis

CFA was conducted in AMOS 27 to test the hypothesized second-order reflective structure. Model fit was evaluated using a comprehensive set of indices: the non-significant chi-square test ($p > .05$); $\chi^2/df \leq 3.0$; Goodness of Fit Index (GFI) ≥ 0.90; Comparative Fit Index (CFI) ≥ 0.95; Tucker-Lewis Index (TLI) ≥ 0.95; Root Mean Square Error of Approximation (RMSEA) ≤ 0.06; Standardized Root Mean Square Residual (SRMR) ≤ 0.08; and p of close fit (PCLOSE) > 0.05 (Hu & Bentler, 1999; Steiger, 1990). Convergent validity was assessed via standardized factor loadings ($\lambda \geq 0.50$) and AVE (≥ 0.50); discriminant validity was assessed via HTMT (< 0.85). The second-order factor structure was estimated using the repeated indicators approach (Marsh et al., 1996).

2.6.4 Research hypotheses

Three hypotheses were formulated to guide the quantitative validation:

- **H1:** The SCC of Vietnamese high school students constitutes a second-order reflective structure reflected by four first-order components: PRB, IDE, DES, and EVA.
- **H2:** Each first-order component is meaningfully measured by its four corresponding behavioral indicators, with factor loadings and AVE meeting accepted SEM thresholds.
- **H3:** The second-order reflective measurement model achieves adequate fit with empirical data collected from Vietnamese high school students in science learning contexts.

Acceptance of all three hypotheses would confirm the structural and empirical validity of the proposed SCC framework as a measurement instrument suitable for deployment in Physics-based STEM instructional research.

3 RESULTS

3.1 Proposed SCC framework from literature synthesis

The systematic review of 690 Scopus-indexed publications identified four recurrent cognitive phases underlying scientific creativity across multiple theoretical traditions. Scientific creativity competence (SCC) was defined as the capacity to mobilize and coordinate higher-order cognitive processes - operating through a four-phase cycle (PRB → IDE → DES → EVA) - in the context of domain-specific scientific tasks (Benedek & Fink, 2019; Hu & Adey, 2002). Table 1 presents the coding matrix of 16 behavioral indicators assigned to the four first-order components.

Table 1

Four-component SCC framework with 16 behavioral indicators

Component	Code	Behavioral indicator (summary)
PRB - Problem recognition	PRB1	Identify scientific problems from real-world phenomena
	PRB2	Analyze conditions and constraints of the problem
	PRB3	Reformulate the problem in testable scientific terms
	PRB4	Relate the problem to existing scientific knowledge
IDE - Idea generation	IDE1	Propose multiple alternative solutions
	IDE2	Generate ideas with relative novelty compared to prior learning
	IDE3	Integrate cross-disciplinary knowledge into proposed ideas
	IDE4	Justify the scientific rationale for proposed ideas
DES - Solution design	DES1	Develop a structured plan for implementing the solution
	DES2	Select appropriate materials, tools, and methods
	DES3	Collect data and use empirical evidence to evaluate feasibility
	DES4	Construct models or prototypes to test the solution
EVA - Solution evaluation	EVA1	Evaluate the solution against specific criteria (accuracy, feasibility, effectiveness)
	EVA2	Identify strengths and limitations of the solution
	EVA3	Propose revisions based on evaluation outcomes
	EVA4	Communicate findings with scientific reasoning

Source: Authors' own research results

A second-order reflective specification was adopted because SCC is conceptualized as a latent disposition whose behavioral indicators are manifestations - not constitutive parts - of the underlying construct (Hair et al., 2019).

3.2 Delphi round 1 results (N = 42)

Round 1 results indicated broad agreement on the four-component structure. Mean ratings ranged from 4.12 to 4.29 on a 5-point scale; all medians equaled 4; standard deviations fell between 0.56 and 0.63; consensus rates ranged from 78.6% to 85.7% (Table 2).

Table 2

Delphi Round 1 results by component (N = 42)

Component	M	Md	SD	IQR	Consensus (%)
PRB	4.29	4	0.56	0.50	85.7
IDE	4.12	4	0.63	0.75	78.6
DES	4.21	4	0.59	0.50	81.0
EVA	4.18	4	0.61	0.50	80.9

Source: Authors' own research results

Qualitative feedback identified three indicators requiring revision. IDE2 was reframed as “relative novelty compared to prior learning” rather than “absolute novelty”, aligning the indicator with the cognitive level of upper-secondary students (Hu & Adey, 2002). DES3 was revised to emphasize empirical evidence in evaluating feasibility. EVA1 was refined by specifying evaluation criteria (accuracy, feasibility, effectiveness). The four-component structure was retained; revisions were confined to indicator wording.

3.3 Delphi round 2 results (N = 38)

Round 2 was conducted with 38 panelists (retention rate: 90.5%). Table 3 compares descriptive statistics across the two rounds.

Table 3

Comparison of Delphi Round 1 and Round 2 results

Component	R1 M	R1 SD	R1 Con. (%)	R2 M	R2 SD	R2 Con. (%)
PRB	4.29	0.56	85.7	4.37	0.49	89.5
IDE	4.12	0.63	78.6	4.28	0.52	86.8
DES	4.21	0.59	81.0	4.41	0.48	92.1
EVA	4.18	0.61	80.9	4.32	0.51	86.8

Source: Authors’ own research results

Consensus rates for the three revised indicators increased substantially: IDE2 rose from 73.8% to 86.8%; DES3 from 76.2% to 89.5%; EVA1 from 71.4% to 84.2%. All 16 indicators exceeded the 75% consensus threshold. Standard deviations decreased across all components (range: 0.48-0.52 vs. 0.56-0.63 in Round 1), confirming reduced dispersion after revision. The four-component, 16-indicator SCC framework was confirmed through expert consensus.

3.4 PLS-SEM exploratory validation (n = 300)

Table 4 reports the internal consistency reliability and convergent validity indices from PLS-SEM analysis (SmartPLS 4; 5,000 bootstrap resamples).

Table 4

PLS-SEM reliability and convergent validity (n = 300)

Construct	Cronbach’s α	ρA	CR	AVE
PRB	0.792	0.798	0.865	0.617
IDE	0.776	0.778	0.856	0.599
DES	0.792	0.793	0.865	0.615
EVA	0.752	0.758	0.843	0.575

Source: Authors’ own research results

Note. ρA = rho_A; CR = composite reliability; AVE = average variance extracted. Thresholds: α ≥ 0.70; CR ≥ 0.70; AVE ≥ 0.50 (Hair et al., 2019).

All Cronbach’s α values exceeded 0.70 (range: 0.752-0.792); composite reliability values surpassed 0.70 (range: 0.843-0.865); rho_A coefficients (range: 0.758-0.798) corroborated reliability under the consistent-reliability framework (Dijkstra & Henseler, 2015). AVE values ranged from 0.575 (EVA) to 0.617 (PRB), satisfying the 0.50 convergent validity threshold.

Discriminant validity was assessed via the heterotrait-monotrait ratio (HTMT). All pairwise HTMT values fell below the 0.85 threshold (range: 0.605-0.771), with the highest value observed for PRB-DES (HTMT = 0.771). Table 5 presents the HTMT matrix.

Table 5

HTMT discriminant validity matrix (PLS-SEM, n = 300)

	DES	EVA	IDE	PRB
DES	-	-	-	-
EVA	0.640	-	-	-
IDE	0.624	0.605	-	-
PRB	0.771	0.618	0.668	-

Source: Authors’ own research results

Bootstrapping (5,000 resamples) confirmed that all second-order path coefficients were significant ($p < .001$): SCC → PRB ($\beta = 0.833$, $t = 42.102$), SCC → DES ($\beta = 0.826$, $t = 40.798$), SCC → IDE ($\beta = 0.773$, $t = 29.490$), and SCC → EVA ($\beta = 0.749$, $t = 24.206$). Because all 16 indicators met reliability, convergent validity, and discriminant validity criteria, no items were removed prior to the confirmatory phase.

3.5 CB-SEM confirmatory validation (n = 1,200)

3.5.1 Exploratory factor analysis

EFA was conducted in SPSS 27 using principal axis factoring with promax rotation on the full confirmatory sample (n = 1,200). Prior to extraction, the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = 0.934) indicated excellent factorability, and Bartlett's test of sphericity was significant ($\chi^2(120) = 8252.442$, $p < 0.001$), confirming that the correlation matrix was suitable for principal axis factoring (Hair et al., 2019). Four factors with eigenvalues exceeding 1.0 were extracted (6.675, 1.503, 1.300, and 1.029), collectively accounting for 65.663% of total variance (Table 6). The scree plot supported the retention of four factors. Rotation converged in six iterations.

The rotated pattern matrix (Table 7) revealed a clean simple structure: all 16 indicators loaded exclusively on their theoretically hypothesized factor, with pattern coefficients ranging from 0.622 (IDE2) to 0.835 (PRB3), all exceeding the 0.50 retention threshold (Hair et al., 2019). No cross-loading exceeded 0.30. Factor 1 captured all four EVA indicators (range: 0.655-0.784); Factor 2 captured all four DES indicators (range: 0.680-0.825); Factor 3 captured all four IDE indicators (range: 0.622-0.747); and Factor 4 captured all four PRB indicators (range: 0.638-0.835). The oblique rotation solution yielded balanced rotated sums of squared loadings across factors (4.253-4.929), reflecting the comparable contribution of each component to total explained variance. The emergent four-factor structure corresponded exactly to the theoretically specified PRB-IDE-DES-EVA framework, providing empirical justification for imposing the theoretically derived measurement structure in the subsequent CFA.

Table 6

EFA eigenvalues and variance explained (n = 1,200)

Factor	Initial Eigenvalue	% of Variance	Cumulative %	Extraction Sum of Squared Loadings
1	6.675	41.716	41.716	6.227
2	1.503	9.391	51.107	1.037
3	1.300	8.128	59.235	0.840
4	1.029	6.428	65.663	0.599
5	0.588	3.674	69.337	-

Source: Authors' own research results

Note. Extraction method: Principal axis factoring. Rotation method: Promax with Kaiser normalization. KMO = 0.934; Bartlett's test of sphericity: $\chi^2(120) = 8252.442$, $p < .001$. Only factors with eigenvalues > 1.0 were retained. Rotation sums of squared loadings for obliquely rotated factors cannot be summed to yield total variance (Hair et al., 2019).

Table 7

EFA pattern matrix - rotated factor loadings (n = 1,200)

Indicator	Factor 1 (EVA)	Factor 2 (DES)	Factor 3 (IDE)	Factor 4 (PRB)
EVA4	0.784			
EVA2	0.723			
EVA1	0.718			
EVA3	0.655			
DES1		0.825		
DES4		0.737		
DES2		0.706		
DES3		0.680		
IDE1			0.747	
IDE3			0.726	
IDE4			0.705	
IDE2			0.622	
PRB3				0.835
PRB1				0.700
PRB4				0.685
PRB2				0.638

Source: Authors' own research results

Note. Loadings < 0.30 suppressed for clarity. Extraction method: Principal axis factoring. Rotation method: Promax with Kaiser normalization (converged in 6 iterations). All retained loadings ≥ 0.50 (Hair et al., 2019). No cross-loading exceeded 0.30

3.5.2 Confirmatory factor analysis

CFA was conducted in AMOS 27 to test the hypothesized second-order reflective model using the independent validation sample. Table 8 presents the model fit indices.

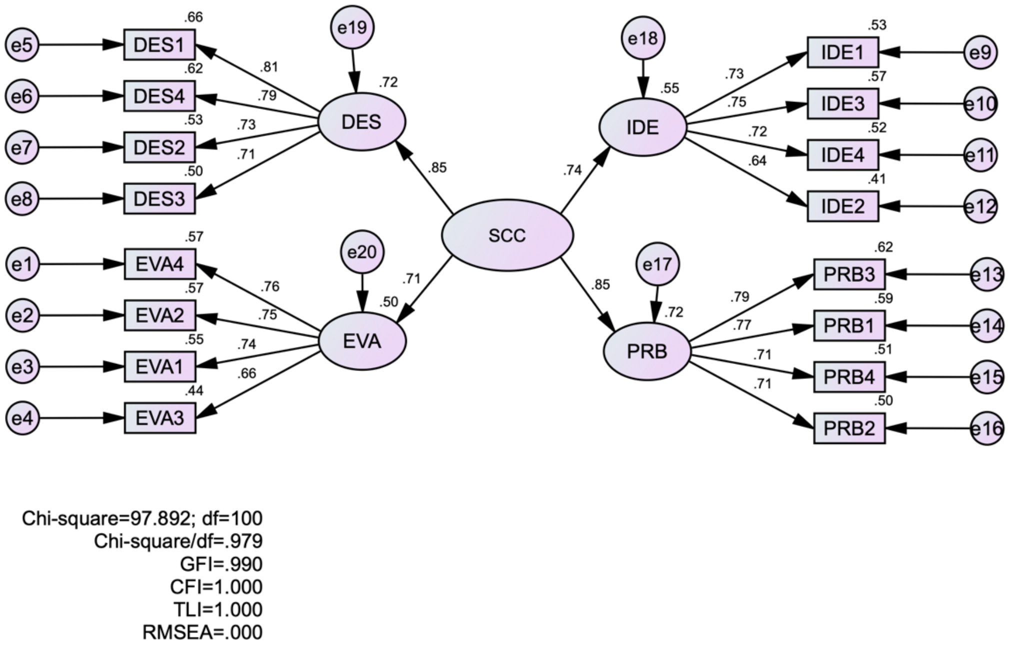
All fit indices satisfied the recommended thresholds (Hu & Bentler, 1999; Hair et al., 2019). The chi-square test was non-significant - $\chi^2(100) = 97.892$, $p = 0.541$ - indicating that the model-implied covariance matrix was statistically indistinguishable from the observed covariance matrix. The χ^2/df ratio of 0.979 fell marginally below unity, yielding RMSEA = 0.000 as a mathematically expected outcome when $\chi^2 < df$, not as a sign of model misspecification (Steiger, 1990). SRMR = 0.017 provided convergent evidence of minimal absolute residual misfit across all indicator pairs. The combination of a non-significant χ^2 test, GFI = 0.990, CFI = 1.000, TLI = 1.000, and SRMR = 0.017 confirmed that the second-order reflective structure was fully supported by the independent validation sample (n = 1,200). Table 9 presents the standardized loadings and construct-level reliability.

Standardized first-order loadings ranged from 0.64 (IDE2) to 0.81 (DES1). IDE2 ("generate ideas with relative novelty") exhibited the lowest loading at 0.64, remaining above the 0.50 minimum but below the more stringent 0.70 benchmark - a pattern suggesting that relative novelty is a weaker yet meaningful manifestation of Idea Generation. CR values ranged from 0.803 (IDE) to 0.846 (DES); AVE values ranged from 0.506 (IDE) to 0.579 (DES), all exceeding 0.50.

The second-order loadings (γ) from SCC to the four components were: DES and PRB (both $\gamma = 0.85$; $R^2 = 0.72$), IDE ($\gamma = 0.74$; $R^2 = 0.55$), and EVA ($\gamma = 0.71$; $R^2 = 0.50$). SCC explained 72% of the variance in DES and PRB, 55% in IDE, and 50% in EVA.

Figure 1

Second-order confirmatory factor analysis model of scientific creativity competence (SCC) with standardized factor loadings and global fit indices



Source: Authors' own research results

Table 8

CFA model fit indices (second-order model, n = 1,200)

Index	Value	Criterion	Decision
χ^2	97.892	-	-
df	100	-	Over-identified
χ^2/df	0.979	≤ 3.00	Met
GFI	0.990	≥ 0.90	Met
CFI	1.000	≥ 0.95	Met
TLI	1.000	≥ 0.95	Met
RMSEA	0.000	≤ 0.06	Met
SRMR	0.017	≤ 0.08	Met
PCLOSE	1.000	> 0.05	Met

Source: Authors' own research results

Note. df = 100 confirms substantial over-identification of the second-order model. RMSEA = 0.000 reflects the mathematical outcome of $\chi^2 < \text{df}$ (RMSEA $= \max[0, \sqrt{(\chi^2/\text{df} - 1)/n}]$); the expression under the radical is negative, yielding zero by convention; Steiger, 1990). SRMR = 0.017 provides convergent evidence of minimal absolute misfit (Hu & Bentler, 1999).

Table 9

CFA standardized loadings and construct reliability (n = 1,200)

Component	Indicator	λ	R²	CR	AVE
PRB	PRB3	0.79	0.62	0.833	0.556
	PRB1	0.77	0.59		
	PRB4	0.71	0.51		
	PRB2	0.71	0.50		
IDE	IDE3	0.75	0.57	0.803	0.506
	IDE1	0.73	0.53		
	IDE4	0.72	0.52		
	IDE2	0.64	0.41		
DES	DES1	0.81	0.66	0.846	0.579
	DES4	0.79	0.62		
	DES2	0.73	0.53		
	DES3	0.71	0.50		
EVA	EVA4	0.76	0.57	0.819	0.531
	EVA2	0.75	0.57		
	EVA1	0.74	0.55		
	EVA3	0.66	0.44		

Source: Authors' own research results

Note. λ = standardized loading; R² = squared multiple correlation; CR = composite reliability; AVE = average variance extracted. All loadings significant at $p < 0.001$.

3.5.3 Convergence across methods

The pattern of results was consistent across the two analytic stages. In PLS-SEM, PRB ($\beta = 0.833$) and DES ($\beta = 0.826$) were the two strongest reflections of SCC; in CFA, DES and PRB shared the highest second-order loadings ($\gamma = 0.85$). EVA consistently exhibited the lowest path weight in both analyses (PLS-SEM $\beta = 0.749$; CFA $\gamma = 0.71$).

3.5.4 Hypothesis testing

H1 was confirmed: all four second-order path coefficients were significant ($p < 0.001$) in both PLS-SEM (β: 0.749-0.833) and CFA (γ: 0.71-0.85), supporting SCC as a second-order reflective construct. H2 was confirmed: EFA identified a clean four-factor solution (total variance explained = 65.663%) in which all 16 indicators loaded exclusively on their theoretically designated factor (pattern coefficients: 0.622-0.835; no cross-loading > 0.30). CFA corroborated the structure, with all standardized first-order loadings ≥ 0.64 and all AVE values ≥ 0.50 (Hair et al., 2019). H3 was confirmed: the second-order reflective model achieved excellent fit - $\chi^2(100) = 97.892$, $p = 0.541$; χ^2/df

= 0.979; GFI = 0.990; CFI = 1.000; TLI = 1.000; RMSEA = 0.000; SRMR = 0.017; PCLOSE = 1.000 - with all indices satisfying the recommended thresholds (Hu & Bentler, 1999; Steiger, 1990). The combined PLS-SEM and CB-SEM/CFA evidence establishes the structural and empirical validity of the four-component, 16-indicator SCC framework for upper-secondary science education research.

4 DISCUSSION

4.1 Theoretical significance of the proposed SCC framework

The central finding of this study - that scientific creativity competence (SCC) in Vietnamese high school students constitutes a second-order reflective structure operating through four first-order components (PRB, IDE, DES, EVA) - carries substantive theoretical implications for the conceptualization and measurement of SCC in secondary science education. The CFA solution demonstrated excellent model fit - $\chi^2(100) = 97.892$, $p = 0.541$; $\chi^2/df = 0.979$; GFI = 0.990; CFI = 1.000; RMSEA = 0.000; SRMR = 0.017; PCLOSE = 1.000 - confirming that the four-component structure provides an accurate representation of the underlying SCC construct in the Vietnamese upper-secondary context. The finding is consistent with the theoretical position of Hu and Adey (2002), who argued that SCC is a domain-specific competence irreducible to general creative ability, and extends the Scientific Creativity Structure Model (SSCM) by providing a psychometrically validated second-order factor solution - a level of structural specification that the original SSCM did not subject to quantitative testing.

The ordering of second-order factor loadings carries substantive theoretical meaning. DES ($\beta = 0.85$, $R^2 = 0.72$) and PRB ($\beta = 0.85$, $R^2 = 0.72$) emerged as the two dominant reflections of SCC, followed by IDE ($\beta = 0.74$, $R^2 = 0.55$) and EVA ($\beta = 0.71$, $R^2 = 0.50$), a pattern fully replicated across PLS-SEM (PRB = 0.833; DES = 0.826) and CFA. The co-dominance of DES and PRB suggests that, at the upper-secondary developmental stage, SCC is most strongly expressed through the capacity to identify and reformulate scientific problems in testable terms and to translate ideas into structured, evidence-grounded plans - both of which demand concurrent activation of domain knowledge and procedural reasoning. The co-dominance pattern aligns with Benedek and

Fink's (2019) neurocognitive framework, which positions convergent elaboration as a defining feature of domain-embedded creative cognition.

The comparatively lower loading of EVA ($R^2 = 0.50$) indicates that self-directed evaluation and iterative revision of scientific solutions remain developing cognitive skills among upper-secondary students - consistent with prior evidence that metacognitive regulation of the creative process lags behind generative phases in adolescent populations (Sternberg et al., 2020). Similarly, IDE2 (*generate ideas with relative novelty compared to prior learning*) yielded the lowest first-order loading ($\lambda = 0.64$), remaining above the minimum 0.50 threshold while falling below the more stringent 0.70 benchmark. The loading pattern is theoretically informative: the reframing of originality as *relative novelty* after Delphi Round 1 was methodologically sound, yet the construct of novelty - even in its relativized form - remains the most difficult behavioral dimension of Idea Generation to observe and rate reliably in classroom settings (Kaufman & Beghetto, 2009).

In comparison with existing instruments, the proposed framework offers two advances over the C-SAT (Ayas & Sak, 2014) and the C-SCA (Xu et al., 2024). First, neither instrument models the full PRB-to-EVA cycle as an integrated second-order structure; to the best of the authors' knowledge, the present framework is the first to formally specify the EVA component - iterative evaluation and refinement of scientific solutions - as a first-order latent construct within a validated second-order model. Second, whereas the SSCM (Hu & Adey, 2002) operates at the level of task performance, the present framework anchors each component to observable student behaviors in physics-based STEM learning contexts, enabling direct deployment in competence-based assessment.

4.2 Methodological significance

The five-step sequential validation procedure constitutes a methodologically rigorous contribution in its own right. A key outcome of the design is the *cross-stage consistency* observed between PLS-SEM and CB-SEM: the component ranking PRB = DES > IDE > EVA was reproduced across two independent samples (n = 300 and n = 1,200) and two distinct analytic frameworks. Such replication across methods substantially strengthens confidence in the stability of the second-order structure and mitigates concerns about sample-specific artifacts (Hair et al., 2019).

The χ^2/df ratio of 0.979 - marginally below unity - warrants explicit technical clarification for readers unfamiliar with this pattern. With $\chi^2(100) = 97.892$, the RMSEA formula (Steiger, 1990) produces a negative value under the radical - specifically, $(97.892/100 - 1)/1{,}200 = -0.0000176$ - yielding RMSEA = 0.000 by mathematical convention. The non-significant chi-square test ($p = 0.541$) confirms that the empirical covariance matrix cannot be statistically distinguished from the model-implied matrix. The df = 100 value confirms substantial over-identification of the model, ruling out any concern about just-identification. SRMR = 0.017 provides convergent evidence at the absolute fit level: on average, the model-reproduced correlations deviate from the observed correlations by only 1.7 percentage points. Collectively, the pattern of $\chi^2 < df$ alongside SRMR < 0.05 reflects a model that was rigorously pre-specified through both theoretical synthesis and two-stage empirical validation before the confirmatory phase, rather than an artifact of overfitting.

Specifically, the finding that the highest pairwise HTMT value (0.771, PRB-DES) remained below the conservative 0.85 threshold in both PLS-SEM and CFA confirms discriminant validity while also suggesting a theoretically interpretable proximity: both components require students to apply physics domain knowledge - one in framing the problem, the other in operationalizing a solution.

The Delphi procedure proved effective as a pre-quantitative content validity filter. Following two-round expert consultation, all 16 indicators entered PLS-SEM and none were subsequently removed - Cronbach's α ranged from 0.752 to 0.792, CR from 0.843 to 0.865, and AVE from 0.575 to 0.617. The sequential validation outcome demonstrates that establishing expert consensus before quantitative validation substantially reduces measurement noise in the factor-analytic pipeline (Endacott et al., 1999; Lucas, 2016). The three revised indicators (IDE2, DES3, EVA1) each showed substantial consensus gains of 12.8-13.3 percentage points between rounds, confirming the responsiveness of the framework to targeted wording refinement without altering the underlying conceptual structure.

The KMO value of 0.934 - classified as "marvelous" by Kaiser (1974) - further attests to the strong intercorrelations among the 16 indicators within each component, consistent with the high reliability coefficients observed at both the PLS-SEM (α: 0.752-

0.792) and CFA (CR: 0.803-0.846) stages.

One measurement outcome warrants methodological attention. IDE's AVE (0.506 in CFA) barely exceeded the 0.50 threshold, and IDE2 loaded at 0.64 - below the more stringent 0.70 criterion. While both values satisfy minimum psychometric requirements (Hair et al., 2019), they signal that the Idea Generation component, and specifically the novelty sub-dimension, may benefit from further indicator refinement. In particular, multi-source assessment approaches - combining student self-report with teacher-administered observation and product-based rubrics - may capture the novelty dimension of IDE more reliably than single-source self-ratings (Lucas, 2016).

4.3 Practical implications

The validated SCC framework has direct utility for two interrelated tasks in Physics-based STEM instruction: assessment design and lesson planning. For assessment, the 16 behavioral indicators underpin three complementary instruments - a five-point Likert self-report scale, a teacher-administered behavioral observation checklist, and a product-and-process rubric for evaluating STEM project outcomes. The behavioral specificity of the indicators (e.g., DES1: *develop a structured plan for implementing the solution*; EVA1: *evaluate the solution against criteria of accuracy, feasibility, and effectiveness*) enables assessors to assign ratings based on observable student actions rather than holistic impressions, improving inter-rater consistency and formative feedback quality (Lucas, 2016).

For instructional planning, the dominance of DES and PRB in the second-order model carries a direct pedagogical signal: STEM learning tasks in Physics should prioritize extended problem formulation and evidence-based solution design over loosely structured brainstorming alone. The lower SCC variance explained by EVA (50%) further suggests that evaluation and revision skills require explicit scaffolding - such as structured peer-review protocols or teacher-guided reflection rounds - rather than incidental development through project completion (Hmelo-Silver, 2004). Applied to Grade 10 Dynamics content, where students design experiments involving force analysis and motion constraints, these emphases translate directly into task design decisions about the relative weight of problem-scoping, prototyping, and iterative testing phases within each

STEM learning sequence (Aguilera & Ortiz-Revilla, 2021).

4.4 Limitations

Three limitations constrain the generalizability of the present findings. First, both quantitative validation samples were drawn from Ho Chi Minh City and Gia Lai Province. Although the stratified sampling design captured urban and non-urban school contexts, broader geographic coverage - particularly rural highland, Mekong Delta, and northern provincial settings - is necessary before the framework can function as a national normative standard. Second, the cross-sectional design provides a structural snapshot of SCC but cannot speak to the developmental trajectory of individual components across grade levels or over the course of a STEM instructional intervention (Creswell & Creswell, 2018). A longitudinal design with repeated SCC measurement would be required to establish whether PRB and DES develop earlier and more rapidly than EVA - a hypothesis that the current factor loading pattern renders theoretically plausible (Hu & Adey, 2002). Third, the present study establishes measurement validity in the absence of any specific pedagogical treatment. Whether the framework is sensitive to pre-to-post instructional change - a prerequisite for its use as an outcome measure in experimental and quasi-experimental STEM research - remains an open empirical question constituting the planned next stage of the broader research program.

5 CONCLUSION

The present study developed and validated a second-order reflective framework for scientific creativity competence (SCC) in Vietnamese high school students through a five-step sequential procedure combining systematic bibliometric synthesis, two-round Delphi expert consultation ($N_1 = 42$; $N_2 = 38$), PLS-SEM exploratory validation ($n = 300$), and CB-SEM confirmatory validation ($n = 1,200$). The proposed framework - comprising four first-order components (PRB, IDE, DES, EVA) operationalized through 16 behavioral indicators - achieved expert consensus rates of 86.8%-92.1% in Round 2 and demonstrated excellent structural fit in CFA - $\chi^2(100) = 97.892$, $p = 0.541$; $\chi^2/df = 0.979$; GFI = 0.990; CFI = 1.000; RMSEA = 0.000; SRMR = 0.017; PCLOSE = 1.000 - with all reliability and validity indices satisfying accepted psychometric thresholds. Cross-stage consistency between PLS-SEM and CFA - with DES and PRB as the dominant second-order reflections of SCC across both analytic frameworks - further corroborates the structural stability of the model.

The study makes two principal contributions. Theoretically, the validated second-order reflective model is, to the best of the authors' knowledge, the first psychometrically rigorous SCC framework developed and normed within the Vietnamese secondary education context, extending the Scientific Creativity Structure Model (Hu & Adey, 2002) with an empirically confirmed four-component factor structure and domain-anchored behavioral indicators. Methodologically, the five-step Delphi-SEM protocol demonstrates that pre-quantitative expert consensus validation substantially reduces measurement artifacts in subsequent factor-analytic procedures, offering a replicable framework development procedure applicable across competence domains in educational measurement research.

The validated framework provides a theoretically grounded and practically deployable foundation for designing assessment instruments and SCC-targeted STEM learning sequences in Physics instruction at the upper-secondary level. Future research should extend normative validation to broader regional samples, employ longitudinal designs to examine the developmental trajectory of individual SCC components across instructional cycles, and conduct quasi-experimental studies to establish the framework's sensitivity to change under structured Physics-based STEM interventions.

REFERENCES

Aguilera, D., & Ortiz-Revilla, J. (2021). STEM vs. STEAM education and student creativity: A systematic literature review. *Education Sciences*, *11*(7), 331. https://doi.org/10.3390/educsci11070331

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: Psychometric validity of the Creative Scientific Ability Test. *Thinking Skills and Creativity*, *13*, 195–205. https://doi.org/10.1016/j.tsc.2014.06.001

Benedek, M., & Fink, A. (2019). Toward a neurocognitive framework of creative cognition: The role of memory, attention, and cognitive control. *Current Opinion in Behavioral Sciences*, *27*, 116–122. https://doi.org/10.1016/j.cobeha.2018.12.009

Communist Party Central Committee. (2013). *Nghị quyết số 29-NQ/TW ngày 04/11/2013 về đổi mới căn bản, toàn diện giáo dục và đào tạo, đáp ứng yêu cầu công nghiệp hóa, hiện đại hóa trong điều kiện kinh tế thị trường định hướng xã hội chủ nghĩa và hội nhập quốc tế* [Resolution No. 29-NQ/TW on fundamental and comprehensive educational reform]. Communist Party of Vietnam.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, *39*(2), 297–316. https://doi.org/10.25300/MISQ/2015/39.2.02

Endacott, R., Pearson, M., & Clifton, S. (1999). Generating consensus for practice using the Delphi method: Practical considerations for nurse educators. *Nurse Education Today*, *19*(7), 553–558. https://doi.org/10.1054/nedt.1999.0366

Eyring, H. (1959). Scientific creativity. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development* (pp. 1–10). Wiley.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation*, *12*(10), 1–8. https://doi.org/10.7275/pdz9-th90

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, *24*(4), 389–403. https://doi.org/10.1080/09500690110098912

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(2), 199–218. https://doi.org/10.1086/376806

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology*, *13*(1), 1–12. https://doi.org/10.1037/a0013688

Kaufman, J. C., & Sternberg, R. J. (Eds.). (2019). *The Cambridge handbook of creativity* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781316979839

Lucas, B. (2016). A five-dimensional model of creativity and its assessment in schools. *Applied Measurement in Education*, *29*(4), 278–290. https://doi.org/10.1080/08957347.2016.1209205

Marsh, H. W., Balla, J. R., & Hau, K.-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 315–353). Lawrence Erlbaum.

Ministry of Education and Training. (2018). *Chương trình giáo dục phổ thông 2018 - Chương trình tổng thể* [National General Education Curriculum 2018 - General programme]. Ministry of Education and Training of Vietnam.

Pinar, A., Akçay, H., & Tüysüz, C. (2025). Effects of instructional interventions on scientific creativity in secondary science education: A systematic review. *Journal of Science Education and Technology*, *34*(1), 45–62. https://doi.org/10.1007/s10956-024-10118-4

Prime Minister of Vietnam. (2021). *Quyết định số 127/QĐ-TTg ngày 26/01/2021 phê duyệt Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng Trí tuệ nhân tạo đến năm 2030* [Decision No. 127/QĐ-TTg approving the National Strategy on Research, Development and Application of Artificial Intelligence to 2030]. Government of Vietnam.

Prime Minister of Vietnam. (2022). *Quyết định số 411/QĐ-TTg ngày 31/3/2022 phê duyệt Chiến lược quốc gia phát triển kinh tế số và xã hội số đến năm 2025, định hướng đến năm 2030* [Decision No. 411/QĐ-TTg approving the National Strategy for Digital

Economy and Digital Society to 2025, with orientation to 2030]. Government of Vietnam.

Saefan, J., Priyono, D., & Wibowo, S. W. (2026). STEM-integrated physics instruction and the development of scientific creativity in upper-secondary students. *Journal of Physics: Conference Series*, *2891*, Article 012042. https://doi.org/10.1088/1742-6596/2891/1/012042

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Sternberg, R. J., Glăveanu, V., Karami, S., Kaufman, J. C., Phillipson, S. N., & Preiss, D. D. (2020). Resource-based theory of creative contributions. *Creativity Research Journal*, *33*(2–3), 214–232. https://doi.org/10.1080/10400419.2020.1855321

UNESCO. (2023). *Guidance for generative AI in education and research*. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

World Economic Forum. (2016). *The future of jobs: Employment, skills and workforce strategy for the Fourth Industrial Revolution*. https://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

Xu, W., Chen, G., Zhao, L., & Lian, Z. (2024). Development and validation of a comprehensive scientific creativity assessment for high school students. *Journal of Research in Science Teaching*, *61*(4), 1102–1135. https://doi.org/10.1002/tea.21903

Authors' Contribution

All authors contributed equally to the development of this article.

Data availability

All datasets relevant to this study's findings are fully available within the article.