International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 4, August 2026, pp. 3346~3361
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i4.38884

Establishing quality indicators for higher-order thinking skills: development and validation using factor analysis

Nia Kania[1], Florence Kyaruzi[2]
[1]Mathematics Education Study Program, Faculty of Teacher Training and Education, Universitas Majalengka, Majalengka, Indonesia
[2]Department of Educational Psychology and Curriculum Studies, Dar es Salaam University College of Education, Dar es Salaam, Tanzania

Article Info

Article history:

Received Jan 30, 2026
Revised Jun 25, 2026
Accepted Jul 7, 2026

Keywords:

Creative thinking
Critical thinking
Exploratory factor analysis
Higher-order thinking skills
Problem-solving

ABSTRACT

Assessing higher-order thinking skills (HOTS) in mathematics education remains challenging because available instruments often measure critical thinking (CT), creative thinking (CrT), and problem-solving (PS) separately, limiting a coherent evaluation of their interrelated nature. This study developed and empirically validated an integrated set of quality indicators for HOTS assessment using a mixed-methods design. In the qualitative phase, focus group discussions (FGDs) with five experts in mathematics education and educational evaluation generated theory-driven indicators grounded in instructional practice. In the quantitative phase, 70 participants (university lecturers, secondary school teachers, and undergraduate students) were selected through stratified purposive sampling to evaluate the indicators. An exploratory factor analysis (EFA) was performed using principal axis factoring with varimax rotation, and factorability was confirmed using Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity. The results showed satisfactory construct validity (KMO=0.75; Bartlett's $p<0.001$) and acceptable factor loadings (≥ 0.50), supporting a stable multidimensional HOTS structure aligned with theoretical expectations. The validated indicators provide a practical measurement framework to support cognitively demanding mathematics instruction and serve as a foundation for subsequent confirmatory factor analysis (CFA) in larger, more diverse samples.

Corresponding Author:

Nia Kania
Mathematics Education Study Program, Faculty of Teacher Training and Education
Universitas Majalengka
K.H. Abdul Halim Street No. 103, Majalengka, Indonesia
Email: niakania@unma.ac.id

1. INTRODUCTION

In the context of 21st-century education, the cultivation of higher-order thinking skills (HOTS) [1]–[4], comprising critical, creative, and problem-solving (PS) abilities [5], has emerged as a central objective for fostering learners' intellectual autonomy, adaptability, and lifelong learning capacity [6]–[8]. These cognitive competencies enable individuals not merely to retrieve factual knowledge but to engage in abstract reasoning, synthesize information from diverse sources, and develop innovative solutions to new and complex problems. Within mathematics education, HOTS are particularly indispensable [9]–[11], as mathematical proficiency transcends algorithmic execution toward conceptual understanding [12], [13], analytical reasoning [14], [15], and the creative application of principles to real-world contexts [16]–[18].

Despite the widespread acknowledgment of their importance, contemporary assessment systems remain limited in their capacity to capture the multidimensional and interactive nature of HOTS [19], [20].

Traditional examinations tend to prioritize procedural accuracy and factual recall over inferential reasoning or metacognitive engagement. Consequently, they provide only a partial representation of students' cognitive capabilities [21], [22], obscuring their ability to transfer knowledge [23]–[26] and reason flexibly across problem contexts [27]–[29]. This diagnostic insufficiency limits educators' capacity to design instructional strategies that nurture higher-order reasoning, reflective judgment, and creative PS.

Over the past two decades, several theoretical frameworks have been proposed to conceptualize HOTS. Ennis's framework, critical thinking (CT) refers to reasoned and reflective thinking used to decide what to believe or what to do [30]. Ennis operationalizes CT into core components: interpretation, analysis, evaluation, inference, and self-regulation. His framework emphasizes epistemic justification, argument evaluation, and reflective judgment, and has been widely adopted in educational assessment research. Ennis's model has informed numerous standardized instruments that assess students' abilities to analyze arguments, identify assumptions, and draw evidence-based conclusions.

In contrast, Torrance's creative thinking (CrT) framework conceptualizes creativity through four core dimensions: fluency, flexibility, originality, and elaboration [31]. Torrance operationalized these dimensions through the Torrance tests of creative thinking (TTCT), which remain among the most widely used creativity assessments globally. His work emphasizes divergent thinking processes, idea generation, and imaginative transformation. While Torrance's framework has been influential, critics argue that it focuses primarily on ideational production rather than domain-specific PS, particularly in mathematical contexts.

Similarly, Krulik and Rudnick [32] PS model conceptualizes PS as a structured cognitive process comprising four stages: reading and exploring, planning, implementing strategies, and reflecting. This framework emphasizes metacognitive regulation, the use of heuristic strategies, and reflective evaluation. Krulik's model has been widely applied in mathematics education to design instructional interventions and assessment rubrics for PS proficiency.

Building on and extending these classical frameworks, research by [33], [34] proposed an integrative HOTS framework specifically tailored to mathematics education. Unlike Ennis, Torrance, and Krulik [35], [36], who developed domain-general models, this study conceptualizes HOTS as an interconnected triadic construct comprising CT, CrT, and PS, which operate synergistically within mathematical contexts. Their framework integrates evaluative reasoning, generative ideation, and strategic PS processes into a unified cognitive system [37]–[39]. This study's empirical studies demonstrate that these dimensions are not independent but dynamically interact during mathematical reasoning tasks, thereby offering a more ecologically valid representation of students' cognitive processes in mathematics learning.

However, despite the conceptual richness of this study's framework, existing studies remain limited in the number of psychometrically validated measurement instruments that operationalize this integrated model. Most current tools still assess CT, creativity, and PS separately, in isolation [40]. This fragmentation risks oversimplifying HOTS and neglecting their interactive nature.

To address this theoretical and methodological gap, the present study develops a comprehensive set of quality indicators for assessing HOTS, grounded in cognitive theory and empirically validated. Drawing explicitly on CT taxonomy, CrT dimensions, and PS stages [41], [42] while integrating them within this study unified framework this study conceptualizes HOTS as an interrelated triad of cognitive processes: CT as the evaluative core, CrT as the generative mechanism, and PS as the integrative application of both. The proposed indicators are designed to capture these dimensions through observable, measurable behaviors in mathematics learning environments.

This study aims to develop and validate a set of quality indicators for assessing HOTS in mathematics education. Specifically, the study seeks to: i) identify theoretically grounded HOTS indicators derived from established cognitive and educational frameworks, and ii) empirically evaluate the construct validity and reliability of these indicators using rigorous psychometric methods. Accordingly, the study is guided by the following research questions:

- Which indicators validly represent HOTS in mathematics learning contexts?
- To what extent do these indicators demonstrate satisfactory construct validity and reliability?

By positioning HOTS measurement at the intersection of cognitive theory, psychometric validation, and pedagogical feasibility, this study adopts a mixed-methods design integrating qualitative synthesis (expert FGDs) and quantitative validation (confirmatory factor analysis). The findings are expected to contribute a theoretically coherent and empirically robust measurement framework, providing a foundation for more equitable, reflective, and intellectually demanding mathematics learning practices.

2. METHOD

2.1. Study design

This study adopted a sequential exploratory mixed-methods design, integrating qualitative exploration through FGDs and quantitative validation using structural equation modeling (SEM) with

LISREL [43], [44]. The qualitative phase aimed to generate theoretically grounded indicators for measuring HOTS, while the quantitative phase empirically tested their validity and reliability. This design was chosen to ensure both conceptual depth through expert elicitation and empirical rigor through statistical verification, aligning with the psychometric standards of instrument development in educational psychology.

2.2. Participants and sampling

Participants for the FGDs were purposively selected based on their recognized expertise and scholarly engagement with HOTS in mathematics education. The experts were purposively selected based on the following criteria: i) demonstrated academic or professional expertise in mathematics education and educational evaluation; ii) had more than 20 years of teaching experience of associate professor or full professor; and iii) had prior experience in instrument review, research, or professional practice relevant to the topic of the study. A total of five experts who met these criteria were invited to evaluate the instrument. The expert panel consisted of five academics: two full professors in mathematics education, one full professor in educational evaluation, and two associate professors in mathematics education. All participants had an established record of research and professional contributions in curriculum development, assessment design, and the evaluation of higher-order cognitive processes. This purposeful heterogeneity was intended to enhance content validity by capturing multidimensional perspectives on the constructs of CT, CrT, and PS.

In addition, quantitative data were collected from 70 respondents (lecturers, teachers, and students). These participants were selected using stratified purposive sampling to ensure representation across educational roles relevant to HOTS implementation and assessment. Participants within each group were recruited based on their experience with mathematics instruction or assessment. This sampling strategy enabled meaningful triangulation between expert qualitative insights and broader practitioner and learner perspectives.

Given that the quantitative phase involved 70 respondents, sample size considerations were taken into account in the application of structural equation modelling (SEM). In SEM, sample size may affect the stability of parameter estimates, standard errors, model fit, and statistical power. Therefore, the SEM results should be interpreted with caution, and the generalizability of the findings may be limited.

2.3. Data collection procedures and ethical considerations

Data collection was conducted in two sequential phases in accordance with the mixed-methods research design. In the qualitative phase, FGDs were organized across three iterative sessions, each lasting approximately 90 minutes. The FGDs followed semi-structured protocols designed to elicit in-depth expert perspectives on the dimensions and sub-dimensions of HOTS. Guiding questions encouraged participants to articulate observable indicators, assessment criteria, and instructional contexts relevant to HOTS measurement. All sessions were audio-recorded with participants' permission, transcribed verbatim, and anonymized to ensure data confidentiality.

In the quantitative phase, survey data were collected from 70 participants, including university lecturers, secondary school teachers, and undergraduate students. The survey aimed to corroborate and triangulate qualitative findings by examining the empirical structure of the proposed indicators. The integration of qualitative expert insights and quantitative stakeholder feedback enhanced methodological rigor and strengthened the trustworthiness of the resulting measurement framework.

Regarding ethical considerations, this study received formal ethical approval from the Institutional Review Board (IRB) of the affiliated university. Prior to participation, all respondents were provided with detailed information about the study objectives and procedures, and subsequently gave written informed consent. Respondents participated voluntarily and were advised that withdrawal was permitted at any point without any penalty. All personal identifiers were removed during data processing, and confidentiality was strictly maintained throughout data collection, analysis, and reporting.

2.4. Validity and trustworthiness

The facilitator adopted an interactive facilitation approach, employing probing questions, reflective summaries, and consensus-building techniques. To ensure trustworthiness, the study adhered to the four classical criteria outlined by Lincoln *et al.* [45]: credibility (through member checking), transferability (through detailed contextual descriptions), dependability (through systematic documentation), and confirmability (through audit trails and reflexivity notes).

2.5. Data analysis

This study adopted a mixed-methods analytic design, combining qualitative and quantitative analyses to enhance rigor and triangulate findings. Qualitative data from the FGDs were analyzed first, followed by quantitative analysis of the survey data to validate and complement the emergent framework.

2.6. Qualitative data analysis

Qualitative data were analyzed in ATLAS.ti (version 23) using grounded theory procedures, including open, axial, and selective coding. During the open coding stage, initial codes were inductively generated from participants' verbatim statements. These codes were then clustered during axial coding based on conceptual similarities and guided by established theoretical frameworks of HOTS, particularly Bloom's revised taxonomy and contemporary models of CT, CrT, and PS. This theory-informed clustering ensured conceptual alignment between empirical data and existing scholarship.

In the selective coding phase, core categories were identified and integrated into a coherent theoretical model representing the dimensions of CT, CrT, and PS. The analytic process emphasized theoretical saturation, whereby no new categories emerged from subsequent coding, and conceptual coherence across categories. To enhance analytical reliability, a second independent coder with expertise in qualitative research methods was involved in the coding process. Inter-coder agreement was assessed through iterative comparison and discussion of coding discrepancies until consensus was achieved. This procedure strengthened the trustworthiness and dependability of the qualitative findings.

2.7. Quantitative data analysis

Quantitative survey data were analyzed using descriptive statistics and psychometric validation procedures, including confirmatory factor analysis (CFA). CFA to assess the construct validity and reliability of the identified HOTS indicators. Internal consistency was examined through Cronbach's alpha and composite reliability (CR) coefficients. The integration of qualitative and quantitative analyses enabled robust validation of the proposed indicator framework.

To verify the construct validity and reliability of the developed indicators, a confirmatory phase was conducted using LISREL 10.4. The analysis proceeded in two stages:

a. Construct validity testing

CFA was used to evaluate the factor structure of the three HOTS constructs, as indicated by the following model-fit indices [46] were used to evaluate adequacy:

- Chi-square divided by degrees of freedom (χ^2/df)≤3.00
- Comparative fit index (CFI)≥0.90
- Goodness of fit index (GFI)≥0.90
- Root mean square error of approximation (RMSEA)≤0.08
- Standardized root mean square residual (SRMR)≤0.08
- Factor loadings ≥0.50 were considered evidence of convergent representation.

b. Reliability and convergent validity

CR and average variance extracted (AVE) were computed from LISREL output. Constructs were deemed reliable when CR≥0.70 and AVE≥0.50, indicating internal consistency and adequate item convergence. Discriminant validity was evaluated by comparing the square root of each construct's AVE with the inter-construct correlations, indicating that each HOTS dimension accounted for distinct conceptual variance.

2.8. Integration of qualitative and quantitative findings

The final stage synthesized the qualitative insights with quantitative evidence to refine and validate the HOTS measurement indicators. This methodological triangulation ensured that the developed framework was both theoretically coherent and empirically verifiable, bridging conceptual understanding with measurable constructs.

3. RESULTS AND DISCUSSION

This section reports findings from both qualitative and quantitative analyses conducted to develop and validate the indicators of HOTS, encompassing CT, CrT, and PS. The qualitative phase generated theoretically grounded indicators from expert insights. In contrast, the quantitative phase statistically verified the construct validity and reliability of the proposed measurement model using exploratory factor analysis (EFA) and CFA.

3.1. Qualitative phase: synthesis of FGD results

The qualitative analysis using ATLAS.ti 23 produced a coherent conceptual framework for HOTS. Through iterative open, axial, and selective coding, themes were clustered into dimensions and behavioral indicators consistent with major theoretical perspectives, particularly Bloom's revised taxonomy [47], problem-based learning theory, and creativity studies [48]. The derived constructs were operationalized into empirically measurable indicators that align with established models of cognitive development, ensuring theoretical fidelity and contextual relevance. The synthesis yielded three primary domains: CT, CrT, and PS, each composed of five structured indicators. Table 1 presents the operational framework for the CT domain by mapping its key aspects and dimensions to the corresponding indicators. This mapping is important because it clarifies how the theoretical construct of CT is translated into observable elements, thereby strengthening the transparency and validity of the indicator set used in this study.

Table 1. Aspect, dimension, and indicator of CT

Code	Aspect	Dimension	Indicator
CT1	Goal-oriented	Engaging in rational reasoning and logical analysis to address problems	Ability to apply logical principles when considering potential solutions
CT2	Evidential basis	Formulating justified beliefs supported by credible evidence and argument validity	Ability to consider various relevant information sources before making decisions
CT3	Attention to detail	Demonstrating precision and accuracy in understanding and interpreting arguments	Ability to deconstruct arguments or information systematically and in detail
CT4	Justification and explanation	Employing appropriate criteria or standards to justify reasoning and explain conclusions	Ability to apply the most relevant standards when addressing a given problem
CT5	Evidence-based evaluation	Demonstrating critical reflection on acquired knowledge and beliefs from diverse experiences	Ability to critically evaluate information sources based on relevance, reliability, and consistency

Survey data collected from 70 respondents (comprising university lecturers, teachers, and students) supported the FGD findings. A comparative analysis revealed that this developed indicator was rated 6% more effective than those developed by Ennis, suggesting greater contextual applicability and alignment with real-world cognitive demands in educational assessment. This result suggests that, while Ennis's model remains theoretically robust, this proposed framework offers a more pragmatic and adaptable approach, particularly for assessing reasoning processes in authentic learning contexts.

To further examine the relative performance of the CT domain, we compared effectiveness across its constituent dimensions. Figure 1 illustrates this comparison by presenting the effectiveness score for each dimension of CT. This visualization is important because it highlights the most effective dimensions, providing empirical guidance for refining indicators and prioritizing instructional improvements or assessments in the domain of CT.

This result suggests that, while Ennis's model remains theoretically robust, the proposed framework offers a more pragmatic and adaptable approach, particularly for assessing reasoning processes in authentic learning contexts. Following the CT domain, we operationalized the CrT construct into measurable components to ensure that it can be assessed consistently within the HOTS framework. Table 2 summarizes the operational framework for the CrT domain by linking its aspects and dimensions to the proposed indicators. This table is essential because it demonstrates how CrT is translated from a theoretical concept into observable indicators, providing a clear basis for instrument development, result interpretation, and future refinement of the indicator set.

To evaluate the variation in the CrT domain, we compared the effectiveness of its constituent dimensions. Figure 2 presents this comparison by displaying the effectiveness scores for each dimension of CrT. This figure is important because it identifies dimensions that may require further reinforcement or refinement, thereby informing targeted improvements in the development of indicators, instructional practices, and assessments related to CrT.

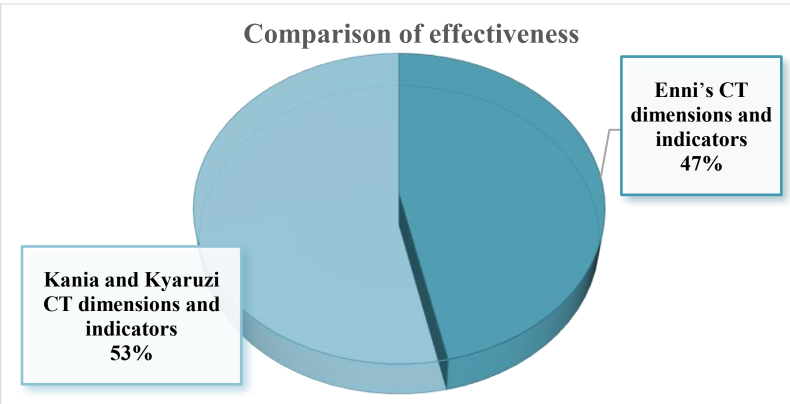

Figure 1. Comparison of the effectiveness of CT skills dimensions

Table 2. Aspect, dimension, and indicator of CrT

Code	Aspect	Dimension	Indicator
CrT1	Analytical thinking	Ability to evaluate information logically and systematically	Ability to identify essential elements within a given set of information
CrT2	Comparative thinking	Capacity to consider multiple perspectives or alternative possibilities	Ability to connect and integrate diverse information or concepts
CrT3	Reflective thinking	Ability to reflect on experiences or ideas to gain deeper insight	Ability to thoughtfully evaluate experiences or concepts for new understanding
CrT4	Corrective thinking	Capacity to refine arguments or approaches to enhance effectiveness	Ability to modify or adjust reasoning strategies for greater efficacy
CrT5	Inquisitive thinking	Motivation to question assumptions, explore new perspectives, and seek unconventional approaches	Ability to seek novel perspectives in understanding complex problems or situations

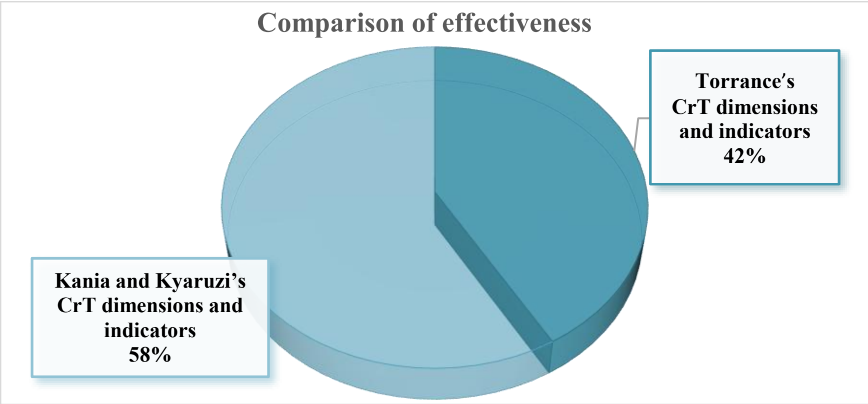

Figure 2. Comparison of the effectiveness of CrT skills dimensions

Empirical validation through the respondent survey confirmed that these proposed dimensions were 16% more effective than Torrance's traditional creativity model. While Torrance's framework prioritizes divergent production and imagination, this framework of Kania and Kyaruzi's work emphasize cognitive adaptability and contextual innovation, qualities more aligned with reflective academic creativity rather than purely artistic expression. To operationalize the PS construct, the study identified its main aspects, dimensions, and indicators based on the underlying theoretical framework. Table 3 summarizes these elements and illustrates the conceptual basis used to develop the PS component of the instrument.

In addition, the survey results distributed to university lecturers, teachers, and students indicated that the dimensions and indicators proposed by Kania and Kyaruzi's were perceived as more effective by 70 respondents. Most participants provided positive evaluations regarding the implementation of these dimensions and indicators in the learning process, suggesting that Kania and Kyaruzi's approach aligns well with their instructional needs and expectations.

Table 3. Aspect, dimension, and indicator of PS

Code	Aspect	Dimension	Indicator
PS1	Cognitive processes	Understanding problems through systematic information gathering and perspective analysis	Ability to identify problems in the level of detail required for effective resolution
PS2	Data gathering	Competence in collecting data from multiple and relevant sources	Ability to design strategic and stepwise plans to address identified problems
PS3	Efficient execution	The capability to implement plans efficiently to achieve problem resolution	Ability to evaluate the execution process and identify opportunities for improvement
PS4	Reflective evaluation	Ability to assess solutions and processes retrospectively	Ability to compare implemented solutions against alternative strategies to determine effectiveness
PS5	Analytical PS	Capacity to manage complexity by decomposing intricate problems into smaller, solvable components	Ability to develop precise and detailed solutions for each component of a complex problem

To examine how respondents evaluated the PS domain, we analyzed the effectiveness ratings across its constituent dimensions. Figure 3 presents a comparative distribution of respondents' effectiveness scores for each PS dimension. This visualization is important because it highlights relative strengths and weaknesses within the PS construct, providing empirical guidance for refining specific indicators and informing targeted improvements in assessment design and instructional support.

Survey findings indicate that this framework scored 10% higher than the Krulik and Rudnick model on perceived practical relevance. The comparative advantage likely stems from the contextual specificity and sequential structuring embedded within indicators, which enable clearer operationalization in instructional and evaluative settings. Collectively, Tables 1–3 establish a coherent conceptual foundation for measuring HOTS, ensuring semantic precision, theoretical distinctiveness, and pedagogical measurability. The qualitative synthesis substantiates the multidimensional yet unified structure of HOTS as an integrative cognitive construct.

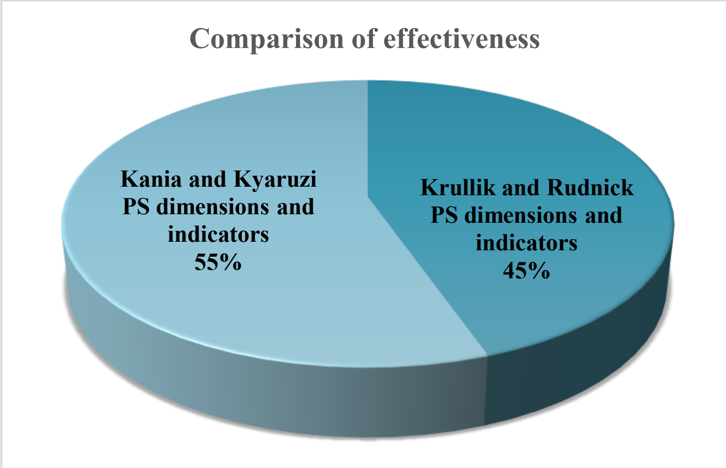

Figure 3. Comparison of the effectiveness of PS skills dimensions

3.2. Quantitative phase: validation of the measurement model

An EFA was undertaken to investigate the structural validity of the developed indicators based on quantitative survey data. The adequacy of the correlation matrix was checked before factor loadings were interpreted. Table 4 reports the measures of sampling adequacy (MSA) for each item, which indicate whether each indicator has sufficient shared variance with the others to be retained in the factor model. This table provides an item-level diagnostic for verifying that the data meet the requirements for EFA and for identifying any indicators that may need to be revised or removed before finalizing the factor structure.

Following the preliminary factor structure identified via EFA, we conducted CFA to test whether the proposed measurement model fit the data adequately. Figure 4 depicts the CFA model, showing the relationships between the latent HOTS constructs (domains/dimensions) and their observed indicators, together with the estimated factor loadings. This figure provides a visual and statistical representation of the validated measurement structure, supporting the construct validity of the developed indicators and confirming their alignment with the theoretical framework.

Table 4. Anti-image matrices

Anti-image matrices																
		CT1	CT2	CT3	CT4	CT5	CrT1	CrT2	CrT3	CrT4	CrT5	PS1	PS2	PS3	PS4	PS5
Anti-image correlation	CT1	.948[a]	-.162	-.028	.049	-.170	-.110	-.003	-.103	.012	-.058	.023	.005	-.143	-.099	-.062
	CT2	-.162	.889[a]	-.061	-.054	.019	-.036	.102	-.004	-.284	-.081	-.171	-.070	.116	.095	-.059
	CT3	-.028	-.061	.945[a]	-.123	-.225	-.050	-.063	-.126	-.131	-.029	.033	-.237	-.029	.016	-.023
	CT4	.049	-.054	-.123	.950[a]	-.038	-.205	-.109	.007	-.001	-.158	.026	-.150	-.051	-.085	-.091
	CT5	-.170	.019	-.225	-.038	.947[a]	-.049	-.059	.046	-.165	-.045	-.018	-.086	-.095	-.151	.035
	CrT1	-.110	-.036	-.050	-.205	-.049	.919[a]	-.072	.011	.011	-.073	-.226	.196	-.189	-.095	.036
	CrT2	-.003	.102	-.063	-.109	-.059	-.072	.941[a]	-.259	-.082	-.111	-.040	.013	.027	-.098	-.060
	CrT3	-.103	-.004	-.126	.007	.046	.011	-.259	.937[a]	-.007	-.126	-.063	-.023	-.047	-.063	.038
	CrT4	.012	-.284	-.131	-.001	-.165	.011	-.082	-.007	.924[a]	-.103	.094	.015	-.175	-.185	.004
	CrT5	-.058	-.081	-.029	-.158	-.045	-.073	-.111	-.126	-.103	.962[a]	-.065	-.028	-.043	-.153	-.110
	PS1	.023	-.171	.033	.026	-.018	-.226	-.040	-.063	.094	-.065	.901[a]	-.265	-.079	.020	-.234
	PS2	.005	-.070	-.237	-.150	-.086	.196	.013	-.023	.015	-.028	-.265	.907[a]	-.077	-.014	-.101
	PS3	-.143	.116	-.029	-.051	-.095	-.189	.027	-.047	-.175	-.043	-.079	-.077	.946[a]	-.050	-.151
	PS4	-.099	.095	.016	-.085	-.151	-.095	-.098	-.063	-.185	-.153	.020	-.014	-.050	.948[a]	.028
	PS5	-.062	-.059	-.023	-.091	.035	.036	-.060	.038	.004	-.110	-.234	-.101	-.151	.028	.939[a]

a. Measures of sampling adequacy (MSA)

To further validate the hypothesized measurement model, we used LISREL 8.80 with Maximum Likelihood (ML) estimation. ML was selected for its robustness under multivariate normality assumptions and its capacity to yield precise estimates of factor loadings and error variances. The measurement model delineates the relationships between the observed indicators and their respective latent constructs: CT, CrT, and PS, as identified in the preceding EFA. Each indicator's loading reflects the strength and direction of its contribution to the latent variable, while the associated error term represents the portion of variance unexplained by the construct [46]. Prior to factor analysis, data suitability was confirmed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity. The KMO value was 0.75, indicating adequate sampling adequacy, and Bartlett's test was significant, $\chi^2(105)=5460.41$, $p<0.001$, demonstrating sufficient inter-item correlations for factor extraction. Accordingly, the dataset was deemed appropriate for conducting EFA.

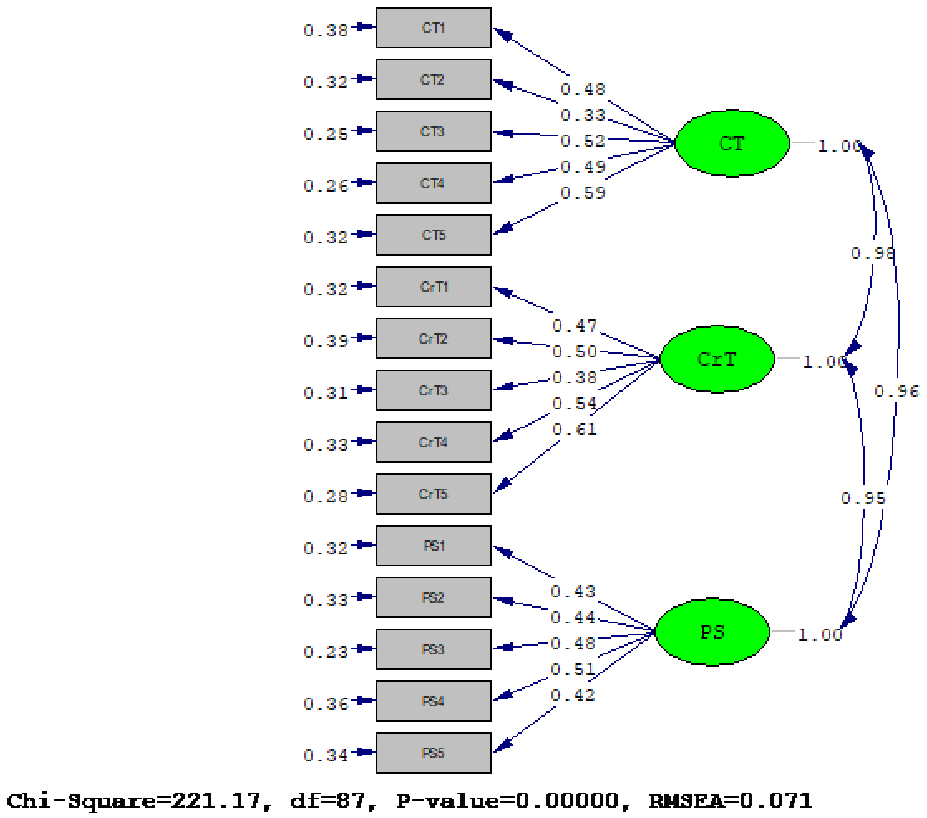

Figure 4. CFA model

3.3. Exploratory factor analysis

EFA was carried out using principal axis factoring with varimax rotation to examine the underlying structure of the developed indicators. The scree plot and eigenvalues greater than 1.0 supported a three-factor solution corresponding to the hypothesized HOTS dimensions. Table 5 reports the EFA results, including the extracted factors, item loadings for each indicator, and key summary statistics, which collectively demonstrate how the indicators cluster within the three-factor structure. This table is provide direct empirical evidence for the proposed dimensionality and helps identify indicators with weak or cross-loadings that may require revision before confirming the final measurement model.

All loadings exceeded 0.70, confirming both construct distinctiveness and theoretical alignment [49]. The emergent structure reflected a stable pattern of cognitive specialization across factors, consistent with the multidimensional HOTS framework [50]. No significant cross-loadings were observed, demonstrating strong discriminant validity. Although three or more indicators per construct are generally recommended in SEM, the two-item factor was retained in this study based on strong theoretical justification and robust empirical evidence. Both indicators demonstrated high standardized factor loadings (>0.70), satisfactory CR and acceptable AVE, indicating sufficient internal consistency and convergent validity.

Previous methodological studies [51], [52] suggest that two-item constructs may be acceptable when the indicators are highly correlated, theoretically coherent, and empirically stable. In this study, the retained items represent the core conceptual domain of the construct and emerged consistently from expert validation during the qualitative phase. Nevertheless, this limitation is acknowledged. Future research is recommended to expand the item pool to strengthen construct representation further and enhance measurement robustness across diverse samples.

Table 5. EFA

Indicator	Factor 1 (CT)	Factor 2 (CrT)	Factor 3 (PS)
Applying logical principles	0.78	—	—
Considering relevant information	0.82	—	—
Recognizing essential elements of information	—	0.85	—
Integrating diverse sources of information	—	0.76	—
Identifying problems in detail	—	—	0.81
Developing accurate and structured plans	—	—	0.79

3.4. Confirmatory factor analysis

The three-factor measurement model was subsequently validated using LISREL 8.80 with ML estimation. To evaluate how well the hypothesized model fits the observed data, we examined standard goodness-of-fit indices. Table 6 summarizes the CFA results, including key model-fit statistics (e.g., χ^2/df, RMSEA, CFI, Tucker–Lewis index (TLI)/non-normed fit index (NNFI), and SRMR) and the standardized factor loadings for each indicator. This table is provide the primary statistical evidence that the proposed three-factor structure adequately represents the data, thereby supporting the construct validity of the developed HOTS indicators and confirming their alignment with the theoretical framework.

All standardized factor loadings ranged between 0.33 and 0.61, and R^2 values ranged between 0.32 and 0.57, indicating acceptable indicator reliability [53]. CR values for CT, CrT, and PS exceeded 0.70, while AVE values surpassed 0.50, confirming internal consistency and convergent validity. Inter-factor correlations were high but theoretically justified: CT–CrT=0.98, CT–PS=0.96, and CrT–PS=0.95 (p<0.001). These values suggest strong interdependence among HOTS components while retaining conceptual distinction, reflecting the hierarchical integration of higher-order cognition.

Table 6. CFA

Goodness-of-fit index	Value	Interpretation
χ^2 (87)=218.58, p<0.001	—	Acceptable, given model complexity
RMSEA=0.071 (90% CI: 0.059–0.083)	<0.08	Good approximate fit
CFI=0.98	≥0.95	Excellent fit
NNFI=0.97	≥0.95	Excellent fit
GFI=0.91, AGFI=0.88	>0.90/>0.85	Acceptable global fit
SRMR=0.047	<0.05	Excellent residual fit

3.5. Structural consistency and theoretical coherence

The observed covariance among the three latent variables supports a synergistic model of higher-order thinking, where critical reasoning, creative integration, and problem-oriented application operate as interlocking cognitive systems. This pattern corroborates prior evidence that HOTS functions as a multidimensional yet unified construct influencing advanced learning outcomes [54], [55]. The model satisfies psychometric expectations for construct validity, reliability, and fit indices, suggesting that the developed indicators provide a robust empirical foundation for measuring HOTS in educational contexts.

3.6. Discussion

The validation results indicate that the HOTS indicators developed through the integration of expert-derived FGD insights and quantitative testing form a coherent multidimensional structure comprising three domains: CT, CrT, and PS. Rather than functioning as isolated competencies, these domains appear empirically connected, supporting the interpretation that HOTS is best represented as an interdependent system of higher cognition. This pattern is consistent with theoretical accounts positioning CT as an evaluative-analytical foundation, CrT as a generative process, and PS as an applied integration of both within task contexts [56]–[70].

The observed three-domain structure aligns with Bloom's revised taxonomy [71] insofar as it reflects differentiated cognitive processes. Importantly, our findings also suggest functional interdependence among these processes, an interpretation that extends prior conceptualizations by emphasizing that higher-order cognition in educational settings may involve coordinated engagement across analytical, generative, and applied reasoning [72]–[76]. Taken together, these results provide empirical support for the construct validity of the proposed indicator framework [77] and strengthen the view that HOTS represents a structured, multidimensional construct rather than a set of discrete abilities [78]–[84].

Our results are in line with prior work reporting HOTS as inherently multidimensional and dynamically enacted in learning tasks [85], [86]. In addition, the present study contributes by providing quantitative evidence that the three domains are not only distinguishable but also meaningfully related within a single measurement framework, thereby supporting arguments that educational assessment should capture integrated higher-order competence rather than fragmented skill components [87]–[90]. This integrated representation is pedagogically relevant because it suggests that growth in one domain (e.g., analytical evaluation) may be accompanied by or depend on development in complementary processes (e.g., idea generation and adaptive application) [91]–[96].

Within the context of competency-based curriculum implementation (e.g., the Merdeka Curriculum), the validated indicators provide a structured basis for developing assessment tools that address how learners reason across critical, creative, and PS demands [97]–[104]. Practically, the framework can serve two complementary functions: i) as a diagnostic reference for identifying relative strengths and weaknesses across HOTS domains and ii) as a design guide for aligning instruction with cognitively demanding learning goals [105]–[116]. These implications follow directly from the validated measurement structure and do not depend on assumptions beyond the data analyzed in this study.

Despite the robust psychometric signals observed in this sample, several limitations should be acknowledged. First, the respondent group was relatively small and drawn from specific educational settings, which may constrain generalizability [117]–[128]. Second, disciplinary composition and local academic norms may influence how HOTS dimensions manifest and are perceived, potentially affecting the observed balance among domains (e.g., greater salience of analytical aspects than generative ones) [129]–[137]. Accordingly, the present evidence should be interpreted as supporting the proposed structure under the current sample conditions rather than as a definitive, context-free model. Future research should extend validation through: i) confirmatory testing in larger and more diverse samples, ii) cross-disciplinary and cross-cultural replication to evaluate measurement stability, and iii) longitudinal designs to examine how HOTS profiles evolve over time and across instructional experiences [138]–[148]. Such work would strengthen external validity and clarify the indicator framework's contextual sensitivity, thereby enhancing its utility for educational evaluation and curriculum-informed assessment.

4. CONCLUSION

This study developed and validated an integrated set of quality indicators to measure HOTS in mathematics education across three core dimensions: CT, CrT, and PS. The indicator development combined expert FGDs with psychometric testing, resulting in a model that is both theoretically grounded and empirically supported. CFA provided strong evidence of model–data fit and alignment. These results support the construct validity of the three-factor HOTS framework and indicate that the indicators can represent analytical, generative, and applied cognitive processes in an integrated manner.

Despite the strong psychometric performance, the findings should be interpreted within the study's sample characteristics and context. Future research should replicate the model with larger, more diverse participants, test measurement invariance across educational levels and disciplines, and employ longitudinal designs to examine the stability and developmental sensitivity of the indicators over time. The validated indicators provide a practical foundation for designing HOTS-oriented assessments and supporting instruction that targets critical, creative, and PS competencies within mathematics learning.

In practice, the validated indicators provide educators, researchers, and policymakers with a diagnostic tool to evaluate and enhance learning processes oriented toward cognitive competence. The model provides a structured framework to identify which aspects of critical, creative, and PS skills need reinforcement in classroom settings. At the policy level, these findings contribute to the development of a national assessment framework aligned with the competencies demanded by 21st-century learning and the Merdeka Curriculum initiatives, emphasizing adaptive, inquiry-based, and student-centered learning paradigms. From a theoretical and philosophical standpoint, this study reinforces the notion that assessing HOTS transcends mere evaluation—it represents a shift in educational epistemology toward cultivating reflective, creative, and autonomous learners. By integrating the principles of the revised Bloom's taxonomy with contextually grounded empirical validation, this research advances the science of educational assessment in Indonesia. Ultimately, the validated HOTS indicators stand as a foundational step toward constructing an educational ecosystem that not only measures learning outcomes but also inspires the transformation of how learners think, create, and solve problems in an increasingly complex world.

FUNDING INFORMATION

This research was supported by Universitas Majalengka, Indonesia.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nia Kania	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓
Florence Kyaruzi				✓			✓	✓	✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

INFORMED CONSENT

Informed consent was obtained from all individuals included in this study.

DATA AVAILABILITY

The data underlying the findings of this study can be obtained from the corresponding author, [NK], upon reasonable request.

REFERENCES

[1] H. Retnawati, H. Djidu, K. Kartianom, E. Apino, and R. D. Anazifa, "Teachers' knowledge about higher-order thinking skills and its learning strategy," *Problems of Education in the 21st Century*, vol. 76, no. 2, pp. 215–230, 2018, doi: 10.33225/pec/18.76.215.

[2] A. H. Abdullah *et al.*, "Using active learning with smart board to enhance primary school students' higher order thinking skills in data handling," *Universal Journal of Educational Research*, vol. 8, no. 10, pp. 4421–4432, Oct. 2020, doi: 10.13189/ujer.2020.081009.

[3] L. Anwar, C. Sa'dijah, W. Murtafiah, and M. Huljannah, "Adversity quotient of Indonesian prospective mathematics teachers in solving geometry higher-order thinking skills problems," *Journal on Mathematics Education*, vol. 15, no. 1, pp. 79–98, Oct. 2023, doi: 10.22342/jme.v15i1.pp79-98.

[4] D. Juniati and I. K. Budayasa, "How the learning style and working memory capacity of prospective mathematics teachers affects their ability to solve higher order thinking problems," *European Journal of Educational Research*, vol. 13, no. 3, pp. 1043–1056, Jul. 2024, doi: 10.12973/eu-jer.13.3.1043.

[5] S. M. Wechsler *et al.*, "Creative and critical thinking: independent or overlapping components?" *Thinking Skills and Creativity*, vol. 27, pp. 114–122, Mar. 2018, doi: 10.1016/j.tsc.2017.12.003.

[6] A. Kosasih, T. Supriyadi, M. I. Firmansyah, and N. Rahminawati, "Higher-order thinking skills in primary school: teachers' perceptions of Islamic education," *Journal of Ethnic and Cultural Studies*, vol. 9, no. 1, pp. 56–76, 2022, doi: 10.29333/ejecs/994.

[7] Z. K. Szabo, P. Körtesi, J. Guncaga, D. Szabo, and R. Neag, "Examples of problem-solving strategies in mathematics education supporting the sustainability of 21st-century skills," *Sustainability*, vol. 12, no. 23, p. 10113, 2020, doi: 10.3390/su122310113.

[8] V. Kövecses-Gősi, "Theoretical and practical issues of learning support, in teacher training," *Acta Polytechnica Hungarica*, vol. 21, no. 3, pp. 205–223, 2024.

[9] B. Tanujaya, R. C. I. Prahmana, and J. Mumu, "Mathematics instruction to promote mathematics higher-order thinking skills of students in Indonesia: moving forward," *TEM Journal*, vol. 10, no. 4, pp. 1945–1954, Nov. 2021, doi: 10.18421/TEM104-60.

[10] F. M. Zana, C. Sa'dijah, Susiswo, L. Anwar, and H. Zulnaidi, "Curriculum and teacher assessment practices in mathematics learning: alignment with higher order thinking skills in Indonesian secondary schools," *Journal on Mathematics Education*, vol. 15, no. 4, pp. 1311–1334, Nov. 2024, doi: 10.22342/jme.v15i4.pp1311-1334.

[11] N. Kania, Y. S. Kusumah, J. A. Dahlan, E. Nurlaelah, and F. Kyaruzi, "Decoding student struggles in geometry: Newman error analysis of higher-order thinking skills," *International Journal of Geometry Research and Inventions in Education (Gradient)*, vol. 1, no. 1, pp. 31–47, Jun. 2024, doi: 10.56855/gradient.v1i01.1146.

[12] B. Jonsson, C. Granberg, and J. Lithner, "Gaining mathematical understanding: the effects of creative mathematical reasoning and cognitive proficiency," *Frontiers in Psychology*, vol. 11, p. 574366, Dec. 2020, doi: 10.3389/fpsyg.2020.574366.

[13] Z. Yang, X. Yang, K. Wang, Y. Zhang, G. Pei, and B. Xu, "The emergence of mathematical understanding: connecting to the closest superordinate and convertible concepts," *Frontiers in Psychology*, vol. 12, p. 525493, Nov. 2021, doi: 10.3389/fpsyg.2021.525493.

[14] K. Härmä, S. Kärkkäinen, and E. Jeronen, “The dramatic arc in the development of argumentation skills of upper secondary school students in geography education,” *Education Sciences*, vol. 11, no. 11, p. 734, Nov. 2021, doi: 10.3390/educsci11110734.
[15] Y. A. Tuyakov, “Enhancing the preparation of future mathematics teachers in a digital learning environment,” *International Journal of Information and Education Technology*, vol. 15, no. 6, pp. 1193–1200, 2025, doi: 10.18178/ijiet.2025.15.6.2322.
[16] N. Kania and Y. S. Kusumah, “The measurement of higher-order thinking skills: a systematic literature review,” *Malaysian Journal of Learning and Instruction*, vol. 22, no. 1, pp. 97–116, Jan. 2025, doi: 10.32890/mjli2025.22.1.6.
[17] K. Kusaeri, A. S. Hamdani, and S. Suprananto, “Student readiness and challenge in completing higher order thinking skill test type for mathematics,” *Infinity Journal*, vol. 8, no. 1, pp. 75–86, Feb. 2019, doi: 10.22460/infinity.v8i1.p75-86.
[18] Edwar, R. I. I. Putri, Zulkardi, and Darmawijoyo, “Developing a workshop model for high school mathematics teachers in constructing HOTS questions through the pendidikan matematika realistik Indonesia approach,” *Journal on Mathematics Education*, vol. 14, no. 4, pp. 603–626, Jul. 2023, doi: 10.22342/jme.v14i4.pp603-626.
[19] J. Jailani *et al.*, “A phenomenological study of challenges that prospective mathematics teachers face in developing mathematical problems that require higher-order thinking skills,” *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 19, no. 10, p. em2339, Oct. 2023, doi: 10.29333/ejmste/13631.
[20] I. C. Suazo-Galdames and A. M. Chaple-Gil, “AI-driven assessment systems in higher education: effectiveness for enhancing critical thinking and creativity,” *Ingénierie des Systèmes d'Information*, vol. 30, no. 6, p. 1665, 2025, doi: 10.18280/isi.300624.
[21] N. Kania, A. Saepudin, and F. Gürbüz, “Assessing cognitive obstacles in learning number concepts: insights from preservice mathematics teachers,” *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, vol. 10, no. 3, pp. 146–166, Jul. 2025, doi: 10.23917/jramathedu.v10i3.8638.
[22] J. C. Vázquez-Parra, L. C. Henao-Rodriguez, J. P. Lis-Gutierrez, and A. Medina-Vidal, “Dynamics of thinking: exploring the interactions of cognitive sub-competencies and their effect on complex thinking in university students,” *Journal of Applied Research in Higher Education*, vol. 18, no. 5, pp. 1424–1441, May 2026, doi: 10.1108/JARHE-03-2025-0209.
[23] A. C. Issac, A. Dhir, and M. Christofi, “True knowledge vs empowering knowledge: conceptualizing a theory of mindfulness and knowledge transfer (TMKT),” *Journal of Managerial Psychology*, vol. 39, no. 3, pp. 264–286, Apr. 2024, doi: 10.1108/JMP-05-2022-0217.
[24] J. A. Selvi A., A. B, P. Rajan Y., and M. Geetha, “The mask of hidden knowledge behind advanced technologies: a case of distress, unlearning and knowledge transfer,” *Journal of Information Technology Teaching Cases*, vol. 15, no. 1, pp. 81–89, May 2025, doi: 10.1177/20438869231210577.
[25] J. Sulistiawan, M. Moslehpour, F. Diana, and P.-K. Lin, “Why and when do employees hide their knowledge?” *Behavioral Sciences*, vol. 12, no. 2, p. 56, Feb. 2022, doi: 10.3390/bs12020056.
[26] E. de W.-de Vries, W. A. Dolfsma, H. J. van der Windt, and M. P. Gerkema, “Knowledge transfer in university–industry research partnerships: a review,” *The Journal of Technology Transfer*, vol. 44, no. 4, p. 1236, 2019, doi: 10.1007/s10961-018-9660-x.
[27] H. Hamzah, M. I. Hamzah, and H. Zulkifli, “Systematic literature review on the elements of metacognition-based higher order thinking skills (HOTS) teaching and learning modules,” *Sustainability*, vol. 14, no. 2, p. 813, Jan. 2022, doi: 10.3390/su14020813.
[28] J. F. Luesia, I. Benítez, R. Company-Córdoba, I. Gómez-Gómez, and M. Sánchez-Martín, “Assessing the relevance of academic competencies in college admission tests from a higher-order thinking perspective: a systematic review,” *Thinking Skills and Creativity*, vol. 48, p. 101251, Jun. 2023, doi: 10.1016/j.tsc.2023.101251.
[29] S. Greiff and P. Kyllonen, “Contemporary assessment challenges: the measurement of 21st century skills,” *Applied Measurement in Education*, vol. 29, no. 4, pp. 243–244, Oct. 2016, doi: 10.1080/08957347.2016.1209209.
[30] B. Derks, F. Manzi, C. van Laar, and N. Ellemers, “Don’t rock the boat! Do men prefer women leaders who support the status quo?” *British Journal of Social Psychology*, vol. 65, no. 2, p. e70053, Apr. 2026, doi: 10.1111/bjso.70053.
[31] Gunawan, Kosim, N. Nisrina, A. Busyairi, A. Harjono, and L. Herayanti, “Fostering creative thinking in pre-service physics teachers: a project-based learning model with e-modules,” *Educational Process International Journal*, vol. 19, no. 1, p. e2025532, 2025, doi: 10.22521/edupij.2025.19.532.
[32] S. Krulik and J. A. Rudnick, *The new sourcebook for teaching reasoning and problem solving in elementary school*. Boston, MA: Allyn and Bacon, 1995.
[33] B. C. Runck, M. P. Brakke, and P. M. Porter, “The extended classroom framework for teaching systems analysis of food systems,” *Natural Sciences Education*, vol. 44, no. 1, pp. 101–111, Dec. 2015, doi: 10.4195/nse2015.04.0004.
[34] I. W. E. Mahendra, N. N. Parmithi, E. Hermawan, D. P. Juwana, and I. W. Gunartha, “Teachers’ formative assessment: accessing students’ high order thinking skills (HOTS),” *International Journal of Innovation, Creativity and Change*, vol. 12, no. 12, pp. 180–202, 2020.
[35] R. H. Ennis, “Critical thinking,” in *The Palgrave Handbook of Critical Thinking in Higher Education*, M. Davies and R. Barnett, Eds., New York, NY: Palgrave Macmillan, 2015, pp. 31–47, doi: 10.1057/9781137378057.0005.
[36] E. P. Torrance, “Creativity research in education,” in *Perspectives in Creativity*, I. A. Taylor, Ed., New York, NY: Routledge, 2017, pp. 278–296, doi: 10.4324/9781315126265-12.
[37] E. P. Torrance, “Retooling education for creative talent: how goes it?” *Gifted Child Quarterly*, vol. 18, no. 4, pp. 233–239, Dec. 1974, doi: 10.1177/001698627401800401.
[38] K. Khotimah, M. K. Budiarto, A. Amarulloh, S. W. M. Diningrat, A. N. Carreza, and J. H. Son, “Fostering creativity through meta virtual project-based networked learning: an in-depth examination,” *Electronic Journal of e-Learning*, vol. 24, no. 1, pp. 93–108, Jan. 2026, doi: 10.34190/ejel.24.1.4302.
[39] R. B. Kusumandari, H. Triluqman BS, and Martitah, “Enhancing critical thinking and student independence through Interactive math media: an articulate storyline-based study in Indonesian elementary education,” *Journal of Educational and Social Research*, vol. 15, no. 5, pp. 312–325, Sep. 2025, doi: 10.36941/jesr-2025-0178.
[40] I. Setiasih, N. Winarno, E. C. Prima, F. Qamila, R. A. Firdaus, and N. J. Ahmad, “Fostering students’ critical thinking through STEM project-based learning: designing earthquake-resistant buildings,” *STEM Education*, vol. 6, no. 2, pp. 284–306, 2026, doi: 10.3934/steme.2026012.
[41] R. H. Ennis, “Critical thinking and subject specificity: clarification and needed research,” *Educational Researcher*, vol. 18, no. 3, pp. 4–10, Apr. 1989, doi: 10.3102/0013189X018003004.
[42] J. T. Manoy, S. Hartono, S. L. Yusrina, and R. Nikmah, “Ethnomathematics of Bandar Kediri’s Ikat weaving: mathematical concepts in a traditional east Javanese textile craft,” *Multidisciplinary Science Journal*, vol. 8, no. 6, p. 2026384, Dec. 2025, doi: 10.31893/multiscience.2026384.
[43] M. T. McCrudden and E. M. McTigue, “Implementing integration in an explanatory sequential mixed methods study of belief bias about climate change with high school students,” *Journal of Mixed Methods Research*, vol. 13, no. 3, pp. 381–400, Jul. 2019, doi: 10.1177/1558689818762576.

[44] V. J. Owan *et al.*, "Lecturers' job satisfaction questionnaire (LJSQ): development, construct validity, and bifactor modelling in Nigerian universities," *Multidisciplinary Reviews*, vol. 9, no. 6, p. 2026256, Nov. 2025, doi: 10.31893/multirev.2026256.

[45] Y. S. Lincoln, E. G. Guba, and J. J. Pilotta, "Book Review. Naturalistic inquiry: Beverly Hills, CA: Sage Publications, 1985," *International Journal of Intercultural Relations*, vol. 9, no. 4, pp. 438–439, Jan. 1985, doi: 10.1016/0147-1767(85)90062-8.

[46] J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks, CA: SAGE Publications, Inc., 2014.

[47] M. D. Laddha, V. T. Lokare, A. W. Kiwelekar, and L. D. Netak, "Classifications of the summative assessment for revised Bloom's taxonomy by using deep learning," *International Journal of Engineering Trends and Technology*, vol. 69, no. 3, pp. 211–218, Mar. 2021, doi: 10.14445/22315381/IJETT-V69I3P232.

[48] S. M. Brookhart, *How to assess higher-order thinking skills in your classroom*. Alexandria, VA: ASCD, 2010.

[49] L. M. Dehkordi, M. Houshmand, S. Alahdadi, and A. Yaghoobzadeh, "Psychometric properties of the learning organization perception scale in Iranian universities of medical sciences," *Scientific Reports*, vol. 16, no. 1, p. 2368, Dec. 2025, doi: 10.1038/s41598-025-32166-9.

[50] G. A. Begimbetova, H. Retnawati, G. K. Kassymova, M. B. Triyono, M. N. A. Azman, and R. A. Sodikin, "Developing students' HOTS digital literacy profile by using IRT," *Journal of Curriculum Studies Research*, vol. 7, no. 2, pp. 155–176, Aug. 2025, doi: 10.46303/jcsr.2025.16.

[51] M. Sarstedt, J. F. Hair, J.-H. Cheah, J.-M. Becker, and C. M. Ringle, "How to specify, estimate, and validate higher-order constructs in PLS-SEM," *Australasian Marketing Journal*, vol. 27, no. 3, pp. 197–211, 2019, doi: 10.1016/j.ausmj.2019.05.003.

[52] R. L. Worthington and T. A. Whittaker, "Scale development research," *The Counseling Psychologist*, vol. 34, no. 6, pp. 806–838, Nov. 2006, doi: 10.1177/0011000006288127.

[53] J. F. Hair *et al.*, "Executing and interpreting applications of PLS-SEM: updates for family business researchers," *Journal of Family Business Strategy*, vol. 12, no. 3, p. 100392, Sep. 2021, doi: 10.1016/j.jfbs.2020.100392.

[54] D. T. Tiruneh, M. De Cock, J. M. Spector, X. Gu, and J. Elen, "Toward a systematic and model-based approach to design learning environments for critical thinking," in *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, J. M. Spector, B. B. Lockee, and M. D. Childress, Eds., Cham: Springer International Publishing, 2017, pp. 1–37, doi: 10.1007/978-3-319-17727-4_79-1.

[55] F. Wang, L. M. Zeng, and R. B. King, "University students' socio-emotional skills: the role of the teaching and learning environment," *Studies in Higher Education*, vol. 50, no. 8, pp. 1670–1687, Aug. 2025, doi: 10.1080/03075079.2024.2389447.

[56] H. Song and L. Cai, "Interactive learning environment as a source of critical thinking skills for college students," *BMC Medical Education*, vol. 24, no. 1, p. 270, Mar. 2024, doi: 10.1186/s12909-024-05247-y.

[57] C. Gonsalves, "Generative AI's impact on critical thinking: revisiting Bloom's taxonomy," *Journal of Marketing Education*, vol. 48, no. 1, pp. 4–19, Apr. 2026, doi: 10.1177/02734753241305980.

[58] A. Essien, O. T. Bukoye, X. O'Dea, and M. Kremantzis, "The influence of AI text generators on critical thinking skills in UK business schools," *Studies in Higher Education*, vol. 49, no. 5, pp. 865–882, May 2024, doi: 10.1080/03075079.2024.2316881.

[59] M. Gerlich, "AI tools in society: impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, Jan. 2025, doi: 10.3390/soc15010006.

[60] N. Hidayati, S. Zubaidah, E. Suarsini, and H. Praherdhiono, "Examining the relationship between creativity and critical thinking through integrated problem-based learning and digital mind maps," *Universal Journal of Educational Research*, vol. 7, no. 9A, pp. 171–179, Sep. 2019, doi: 10.13189/ujer.2019.071620.

[61] O. E. Tamayo Alzate, "Critical thinking and epistemic sophistication in science education," *Journal of Intelligence*, vol. 13, no. 8, p. 93, Jul. 2025, doi: 10.3390/jintelligence13080093.

[62] B. A. Saeed and T. Ramdane, "The effect of implementation of a creative thinking model on the development of creative thinking skills in high school students: a systematic review," *Review of Education*, vol. 10, no. 3, p. e3379, 2022, doi: 10.1002/rev3.3379.

[63] T. Raj, P. Chauhan, R. Mehrotra, and M. Sharma, "Importance of critical thinking in the education," *World Journal of English Language*, vol. 12, no. 3, pp. 126–133, Apr. 2022, doi: 10.5430/wjel.v12n3p126.

[64] K. A. bin Abdulrahman, F. A. Alateeq, and I. Siddiqui, "Critical and creative thinking in the context of education in the health professions," in *Essential Learning Skills for Health Professions Students*, K. A. bin Abdulrahman and H. Darami, Eds., Singapore: Springer Nature Singapore, 2025, pp. 93–103, doi: 10.1007/978-981-96-5670-7_9.

[65] A. Koçoğlu and S. Kanadlı, "Critical thinking disposition as a mediator in creativity and problem-solving: a MASEM study," *Intelligence*, vol. 113, p. 101951, Nov. 2025, doi: 10.1016/j.intell.2025.101951.

[66] C. N. Blundell, L. Cameron, M. Mukherjee, D. Tran, B. Zunica, and B. Tregenza, "Systems thinking, design thinking, computational thinking, and critical thinking," in *Creative Technologies Education: Students as Digital Designers*, 1st ed., M. Bower and B. von Mengersen, Eds., London: Routledge, 2025, pp. 44–58, doi: 10.4324/9781003490715-4.

[67] M. N. R. Jauhariyah *et al.*, "Exploring the research trend and interplay of critical thinking, creative thinking, and problem solving: a bibliometric and literature review analysis," *Multidisciplinary Reviews*, vol. 9, no. 8, 2026, doi: 10.31893/multirev.2026407.

[68] H. Saikia and N. R. Roy, "Fostering critical thinking skills in modern learners: a peep into higher educational institutions," in *Digital Skill Development for Industry 4.0*, P. Kaliraj, G. Singaravelu, and T. Devi, Eds., Boca Raton, FL: Auerbach Publications, 2024, pp. 46–51, doi: 10.1201/9781003504894-4.

[69] R. Adila, Y. Hartono, Indaryanti, Scristia, and M. Yusup, "The effect of creative problem-solving models on students' higher level thinking skills in linear programming," in *Journal of Physics: Conference Series*, Mar. 2020, p. 012047, doi: 10.1088/1742-6596/1480/1/012047.

[70] U. Akpur, "Critical, reflective, creative thinking and their reflections on academic achievement," *Thinking Skills and Creativity*, vol. 37, p. 100683, Sep. 2020, doi: 10.1016/j.tsc.2020.100683.

[71] P. T. Sowden, A. Pringle, and L. Gabora, "The shifting sands of creative thinking: connections to dual-process theory," *Thinking & Reasoning*, vol. 21, no. 1, pp. 40–60, Jan. 2015, doi: 10.1080/13546783.2014.885464.

[72] O. Gun and M. J. Bossé, "Synthesizing cognitive mathematics learning taxonomies," *Thinking Skills and Creativity*, vol. 57, p. 101796, Sep. 2025, doi: 10.1016/j.tsc.2025.101796.

[73] M. Alkandari, "Perspective chapter: mathematical definitions – a key to conceptual clarity and theoretical rigor," in *Mathematics - Teaching and Learning: Teaching and Learning*, M. Tezer, Ed., London, 2025, pp. 47–64, doi: 10.5772/intechopen.1011078.

[74] S. Yu and Q. Xue, "Designing for deep thinking: a theory-driven inquiry into problem-chain teaching in primary mathematics," *Research Square Preprint*, pp. 1–24, Jul. 2025, doi: 10.21203/rs.3.rs-6782865/v1.

[75] A. Nguyen *et al.*, "Conceptualizing and enhancing metaverse literacy for education," *Education and Information Technologies*, vol. 30, no. 12, pp. 17133–17153, Aug. 2025, doi: 10.1007/s10639-025-13486-9.

[76] Y. J. Wu, C.-H. Wu, and K.-L. Peng, “Effects of creativity styles on learning engagement and motivation in STEAM education,” *Sustainability*, vol. 17, no. 6, p. 2755, Mar. 2025, doi: 10.3390/su17062755.
[77] J. Liu, Z. Liu, C. Wang, Y. Xu, J. Chen, and Y. Cheng, “K-12 students' higher-order thinking skills: conceptualization, components, and evaluation indicators,” *Thinking Skills and Creativity*, vol. 52, p. 101551, 2024, doi: 10.1016/j.tsc.2024.101551.
[78] A. Derakhshan and M. S. Taghizadeh, “Does artificial intelligence (AI) nurture or hinder language learners' higher-order thinking skills (HOTS)? A phenomenological study on L2 learners' perspectives and lived experiences,” *International Journal of Applied Linguistics*, Jul. 2025, doi: 10.1111/ijal.12824.
[79] T.-T. Wu, L. M. Silitonga, and A. T. Murti, “Enhancing English writing and higher-order thinking skills through computational thinking,” *Computers & Education*, vol. 213, p. 105012, May 2024, doi: 10.1016/j.compedu.2024.105012.
[80] Z. Aba-Oli, K. Koyas, and A. Husen, “Higher-order thinking skills-oriented problem-based learning interventions in mathematics: a systematic literature review,” *School Science and Mathematics*, vol. 125, no. 3, pp. 214–231, 2025, doi: 10.1111/ssm.12676.
[81] C. Azuuga, P. Agmabire Akendita, and F. Abunkurug, “Classroom environment and pedagogical practices as determinants of mathematics anxiety in senior high school: evidence from Ghana,” *International Journal of Mathematics and Mathematics Education*, vol. 4, no. 2, pp. 85–101, Jun. 2026, doi: 10.56855/ijmme.v4i1.1840.
[82] Suherman and T. Vidákovich, “Assessment of mathematical creative thinking: a systematic review,” *Thinking Skills and Creativity*, vol. 44, p. 101019, Jun. 2022, doi: 10.1016/j.tsc.2022.101019.
[83] J. Nathaniel, S. S. Oyelere, J. Suhonen, and M. Tedre, “Investigating the impact of generative AI integration on the sustenance of higher-order thinking skills and understanding of programming logic,” *Computers and Education: Artificial Intelligence*, vol. 9, p. 100460, Dec. 2025, doi: 10.1016/j.caeai.2025.100460.
[84] R. P. T. K. How, H. Zulnaidi, and S. S. B. A. Rahim, “Development of higher-order thinking skills test instrument on quadratic equation (HOTS-QE) for secondary school students,” *Pegem Journal of Education and Instruction*, vol. 13, no. 1, pp. 379–394, Jan. 2023, doi: 10.47750/pegegog.13.01.41.
[85] X. Du, M. Du, Z. Zhou, and Y. Bai, “Facilitator or hindrance? The impact of AI on university students' higher-order thinking skills in complex problem solving,” *International Journal of Educational Technology in Higher Education*, vol. 22, no. 1, p. 39, Jul. 2025, doi: 10.1186/s41239-025-00534-0.
[86] A. A. Alanazi, K. Osman, and L. Halim, “Effect of scaffolding strategies and guided discovery on higher-order thinking skills in physics education,” *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 20, no. 9, p. em2496, Sep. 2024, doi: 10.29333/ejmste/14980.
[87] P. Liamruk, N. Onwong, K. Amornrat, A. Arayapipatkul, and K. Sipiyaruk, “Development and evaluation of an augmented reality serious game to enhance 21st century skills in cultural tourism,” *Scientific Reports*, vol. 15, no. 1, p. 13492, Apr. 2025, doi: 10.1038/s41598-025-95615-5.
[88] R. P. Zúñiga, M. M. García, M. Campos-Valdez, and L. V. Sánchez-Orozco, “Pedagogical and evaluative competencies for the 21st century in molecular biology education and their relationship with honors programs: a systematic literature review,” *Frontiers in Education*, vol. 10, p. 1616858, Aug. 2025, doi: 10.3389/feduc.2025.1616858.
[89] S. N. Dvoryatkina, V. S. Karapetyan, A. M. Dallakyan, S. A. Rozanova, and E. I. Smirnov, “Synergetic effects manifestation by founding complexes deployment of mathematical tasks on the chessboard,” *Problems of Education in the 21st Century*, vol. 77, no. 1, pp. 8–21, Feb. 2019, doi: 10.33225/pec/19.77.08.
[90] I. Xenakis and A. Arnellos, “Relating creativity to aesthetics through learning and development: an interactivist-constructivist framework,” *Phenomenology and the Cognitive Sciences*, pp. 1–36, Jun. 2025, doi: 10.1007/s11097-025-10084-5.
[91] Y. Wu, X. Lu, and C. Lin, “Bridging disciplines: enhancing integrative thinking via collaborative problem-based learning in higher education,” *Thinking Skills and Creativity*, vol. 58, p. 101939, Dec. 2025, doi: 10.1016/j.tsc.2025.101939.
[92] S. Magistretti, C. T. A. Pham, and C. Dell'Era, “The creative process of problem framing for innovation: an integrative review and research agenda,” *Journal of Product Innovation Management*, vol. 42, no. 6, pp. 987–1018, 2025, doi: 10.1111/jpim.12783.
[93] Y. Jie and X. Zhou, “Research on the evaluation of college English teaching based on data mining technology from the perspective of ‘internet plus education,'” *International Journal of Web-Based Learning and Teaching Technologies*, vol. 19, no. 1, pp. 1–14, Oct. 2024, doi: 10.4018/IJWLTT.357247.
[94] Y. Wang, B. Xie, L. Zhou, L. Wang, and H. Jin, “The influence of learning interest on complex mathematical problem-solving ability: the mediating effect of classroom disruptive behavior and self-efficacy,” *Frontiers in Psychology*, vol. 16, p. 1638695, Oct. 2025, doi: 10.3389/fpsyg.2025.1638695.
[95] D. Wu and J. Zhang, “Generative artificial intelligence in secondary education: applications and effects on students' innovation skills and digital literacy,” *PLoS ONE*, vol. 20, no. 5, p. e0323349, May 2025, doi: 10.1371/journal.pone.0323349.
[96] Z. Pineda-Rico, “Implementing complex thinking as a formative framework in engineering education: a longitudinal study across two integrative project-based courses,” *International Journal of Mechanical Engineering Education*, pp. 1–24, Mar. 2026, doi: 10.1177/03064190261434595.
[97] A. Montefusco and F. Angeli, “Turning complexity into a delight to the mind: an integrative framework for teaching and learning complex reasoning,” *Management Learning*, vol. 56, no. 3, pp. 484–511, Jul. 2025, doi: 10.1177/13505076241258932.
[98] Z. Zhang *et al.*, “Understanding and improving fairness in cognitive diagnosis,” *Science China Information Sciences*, vol. 67, no. 5, p. 152106, May 2024, doi: 10.1007/s11432-022-3852-0.
[99] Y. Lee, G. Migut, and M. Specht, “Unveiling cognitive processes in digital reading through behavioural cues: a hybrid intelligence (HI) approach,” *British Journal of Educational Technology*, vol. 56, no. 2, pp. 678–711, Mar. 2025, doi: 10.1111/bjet.13551.
[100] J. Zou, M. Li, and D. Yang, “Cold start and hot feedback: a knowledge sharing and governance model considering individual willingness from a prospect theory perspective,” *Journal of Knowledge Management*, vol. 29, no. 2, pp. 512–540, Jan. 2025, doi: 10.1108/JKM-02-2024-0164.
[101] D. R. van Nederpelt *et al.*, “User requirements for quantitative radiological reports in multiple sclerosis,” *European Radiology*, vol. 35, no. 10, pp. 5967–5978, Apr. 2025, doi: 10.1007/s00330-025-11544-x.
[102] O. L. Aiyegbusi *et al.*, “Recommendations to address respondent burden associated with patient-reported outcome assessment,” *Nature Medicine*, vol. 30, no. 3, pp. 650–659, Mar. 2024, doi: 10.1038/s41591-024-02827-9.
[103] T. Toheri, W. Winarso, and A. A. Haqq, “Where exactly for enhance critical and creative thinking: the use of problem posing or contextual learning,” *European Journal of Educational Research*, vol. 9, no. 2, pp. 877–887, 2020, doi: 10.12973/eu-jer.9.2.877.
[104] S. Fathonah, E. Cahyono, S. Haryani, S. Sarwi, and N. Hayati Lestari, “Application of multirepresentation-based creative problem-solving learning models to improve critical and creative thinking skills for students,” *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, vol. 12, no. 1, pp. 185–200, Apr. 2024, doi: 10.23947/2334-8496-2024-12-1-185-200.

[105] X. Shou, K. Han, and S. Luo, "Primary students' core dimension of higher order thinking in science: measurement and reflections," *Psychology in the Schools*, vol. 62, no. 11, pp. 4760–4778, Nov. 2025, doi: 10.1002/pits.70035.

[106] C. Y. Piaw, L. F. Ying, and L. F. Chiat, "Unveiling key e-learning ingredients for enhancing higher-order thinking skills," *Discover Education*, vol. 4, no. 1, p. 202, Jun. 2025, doi: 10.1007/s44217-025-00600-9.

[107] Oknaryana *et al.*, "Improving students' higher-order thinking skills: a comparison between flipped learning and traditional teaching approach," *European Journal of Educational Research*, vol. 14, no. 4, p. 1245, 2025, doi: 10.12973/eu-jer.14.4.1245.

[108] Z. Cui, O. Ng, M. S. Jong, and X. Weng, "Middle school students' in-moment engagement in synchronous online learning: an activity–community of inquiry approach," *Journal of Computer Assisted Learning*, vol. 41, no. 4, p. e70081, Aug. 2025, doi: 10.1111/jcal.70081.

[109] A. Solano *et al.*, "Work route for the inclusion of learning analytics in the development of interactive multimedia experiences for elementary education," *Applied Sciences*, vol. 14, no. 17, p. 7645, Aug. 2024, doi: 10.3390/app14177645.

[110] R. S. Retno, Purnomo, A. Hidayat, and A. Mashfufah, "Conceptual framework design for STEM-integrated project-based learning (PjBL-STEM) for elementary schools," *Asian Education and Development Studies*, vol. 14, no. 3, pp. 579–604, May 2025, doi: 10.1108/AEDS-08-2024-0188.

[111] E. Rezaei and S. S. B. Shirazi, "The impact of thinking-aloud peer assessment (TAPA) on the development of high order thinking skills (HOTS) in math e-learning," *Education and Information Technologies*, vol. 29, no. 14, pp. 18883–18900, Oct. 2024, doi: 10.1007/s10639-024-12607-0.

[112] A. Fatahillah, V. A. Liyandri, and L. A. Monalisa, "Senior high school students' higher order thinking skills in solving combinatorics problems," *Jurnal Pendidikan Matematika*, vol. 16, no. 1, pp. 89–102, 2021, doi: 10.22342/jpm.16.1.15121.89-102.

[113] C. Hikayat, S. Suparman, Y. Hairun, and H. Suharna, "Design of realistic mathematics education approach to improve critical thinking skills," *Universal Journal of Educational Research*, vol. 8, no. 6, pp. 2232–2244, 2020, doi: 10.13189/ujer.2020.080606.

[114] H. Hyytinen, J. Ursin, K. Silvennoinen, K. Kleemola, and A. Toom, "The dynamic relationship between response processes and self-regulation in critical thinking assessments," *Studies in Educational Evaluation*, vol. 71, p. 101090, Dec. 2021, doi: 10.1016/j.stueduc.2021.101090.

[115] K. S. Double, J. A. McGrane, and T. N. Hopfenbeck, "The impact of peer assessment on academic performance: a meta-analysis of control group studies," *Educational Psychology Review*, vol. 32, no. 2, pp. 481–509, 2020, doi: 10.1007/s10648-019-09510-3.

[116] M. Alqassab, J.-W. Strijbos, and S. Ufer, "Training peer-feedback skills on geometric construction tasks: role of domain knowledge and peer-feedback levels," *European Journal of Psychology of Education*, vol. 33, no. 1, pp. 11–30, Jan. 2018, doi: 10.1007/s10212-017-0342-0.

[117] L. Li, B. Zhang, T. Sun, and F. Drasgow, "The more contextualized, the more valid: effects of contextualization strategies on forced-choice measurement," *Journal of Business and Psychology*, vol. 40, no. 3, pp. 711–729, Jun. 2025, doi: 10.1007/s10869-024-09983-2.

[118] A. Glover, C. Holman, and P. Boise, "Patient-centered respectful maternity care: a factor analysis contextualizing marginalized identities, trust, and informed choice," *BMC Pregnancy and Childbirth*, vol. 24, no. 1, p. 267, Apr. 2024, doi: 10.1186/s12884-024-06491-2.

[119] J. Park, Y. Ryu, and Y. Kim, "Factors influencing air passengers' intention to purchase voluntary carbon offsetting programs: the moderating role of environmental knowledge," *Journal of Air Transport Management*, vol. 118, p. 102619, Jul. 2024, doi: 10.1016/j.jairtraman.2024.102619.

[120] A. P. Jaciw, F. Unlu, and T. Nguyen, "A within-study approach to evaluating the role of moderators of impact in limiting generalizations from 'large to small,'" *American Journal of Evaluation*, vol. 43, no. 1, pp. 108–131, Mar. 2022, doi: 10.1177/10982140211030552.

[121] E. Tipton and R. B. Olsen, "A review of statistical methods for generalizing from evaluations of educational interventions," *Educational Researcher*, vol. 47, no. 8, pp. 516–524, Nov. 2018, doi: 10.3102/0013189X18781522.

[122] H. L. Kern, E. A. Stuart, J. Hill, and D. P. Green, "Assessing methods for generalizing experimental impact estimates to target populations," *Journal of Research on Educational Effectiveness*, vol. 9, no. 1, pp. 103–127, Jan. 2016, doi: 10.1080/19345747.2015.1060282.

[123] A. M. Briesch, S. M. Chafouleas, and A. Johnson, "Use of generalizability theory within K–12 school-based assessment: a critical review and analysis of the empirical literature," *Applied Measurement in Education*, vol. 29, no. 2, pp. 83–107, Apr. 2016, doi: 10.1080/08957347.2016.1138955.

[124] K. Bottema-Beutel, B. Lloyd, E. W. Carter, and J. M. Asmus, "Generalizability and decision studies to inform observational and experimental research in classroom settings," *American Journal on Intellectual and Developmental Disabilities*, vol. 119, no. 6, pp. 589–605, Nov. 2014, doi: 10.1352/1944-7558-119.6.589.

[125] R. Maynard, "Improving the usefulness and use of meta-analysis to inform policy and practice," *Evaluation Review*, vol. 48, no. 3, pp. 515–543, Jun. 2024, doi: 10.1177/0193841X241229885.

[126] E. Tipton and B. J. Matlen, "improved generalizability through improved recruitment: lessons learned from a large-scale randomized trial," *American Journal of Evaluation*, vol. 40, no. 3, pp. 414–430, Sep. 2019, doi: 10.1177/1098214018810519.

[127] M. D. Merlin, S. Lavoie, and F. Gallagher, "Elements of group dynamics that influence learning in small groups in undergraduate students: a scoping review," *Nurse Education Today*, vol. 87, p. 104362, Apr. 2020, doi: 10.1016/j.nedt.2020.104362.

[128] J. T. Tang, D. Mo, and W. C. Lan, "Exploring the integration of cooperative learning in blended teaching environments," *Education and Information Technologies*, vol. 30, no. 18, pp. 26931–26959, Dec. 2025, doi: 10.1007/s10639-025-13791-3.

[129] A. Alam and A. Mohanty, "Cultural beliefs and equity in educational institutions: exploring the social and philosophical notions of ability groupings in teaching and learning of mathematics," *International Journal of Adolescence and Youth*, vol. 28, no. 1, p. 2270662, Dec. 2023, doi: 10.1080/02673843.2023.2270662.

[130] L. Barón-Birchenall, A. Sánchez-Vallejo, C. Toro-Silva, and A. Folleco-Eraso, "Creativity: an individual or collective phenomenon? A historical-psychological perspective," *Adaptive Behavior*, vol. 33, no. 4, pp. 249–269, Aug. 2025, doi: 10.1177/10597123251343849.

[131] H. B. Carlone, A. K. Mercier, and S. R. Metzger, "The production of epistemic culture and agency during a first-grade engineering design unit in an urban emergent school," *Journal of Pre-College Engineering Education Research (J-PEER)*, vol. 11, no. 1, pp. 1–22, Jun. 2021, doi: 10.7771/2157-9288.1295.

[132] N. Norbi, S. R. S. Aris, and N. Abdullah, "Managing the implementation of higher-order thinking skills in Malaysian primary science classrooms: teachers' knowledge, practices and administrative support," *The Malaysian Journal of Qualitative Research*, vol. 12, no. 1, pp. 233–255, May 2026, doi: 10.61211/mjqr120116.

[133] I. Wigati, M. Mardeli, M. Astuti, Y. Yuniar, and Z. Ramdani, “Perception of religious lecturers of higher order thinking skills and students’ academic performance in online learning,” *International Journal of Learning, Teaching and Educational Research*, vol. 22, no. 4, pp. 124–140, Apr. 2023, doi: 10.26803/ijlter.22.4.8.
[134] K. K. Abdulbaki, M. Suhaimi, A. Alsaqqaf, and W. Jawad, “The use of the discussion method at university: enhancement of teaching and learning,” *International Journal of Higher Education*, vol. 7, no. 6, pp. 118–128, 2018, doi: 10.5430/ijhe.v7n6p118.
[135] D. Gong, H. H. Yang, and J. Cai, “Exploring the key influencing factors on college students’ computational thinking skills through flipped-classroom instruction,” *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, p. 19, Dec. 2020, doi: 10.1186/s41239-020-00196-0.
[136] R. A. Razak, S. M. Yusoff, C. H. Leng, and A. F. M. Marzaini, “Evaluating teachers’ pedagogical content knowledge in implementing classroom-based assessment: a case study among ESL secondary school teachers in Selangor, Malaysia,” *PLoS ONE*, vol. 18, no. 12, p. e0293325, Dec. 2023, doi: 10.1371/journal.pone.0293325.
[137] H. Akram, Y. Yingxiu, A. S. Al-Adwan, and A. Alkhalifah, “Technology integration in higher education during COVID-19: an assessment of online teaching competencies through technological pedagogical content knowledge model,” *Frontiers in Psychology*, vol. 12, p. 736522, Aug. 2021, doi: 10.3389/fpsyg.2021.736522.
[138] H. Abu-Rasheed, C. Weber, and M. Fathi, “Context based learning: a survey of contextual indicators for personalized and adaptive learning recommendations – a pedagogical and technical perspective,” *Frontiers in Education*, vol. 8, p. 1210968, Jul. 2023, doi: 10.3389/feduc.2023.1210968.
[139] L. N. Yeganeh, N. S. Fenty, Y. Chen, A. Simpson, and M. Hatami, “The future of education: a multi-layered metaverse classroom model for immersive and inclusive learning,” *Future Internet*, vol. 17, no. 2, p. 63, Feb. 2025, doi: 10.3390/fi17020063.
[140] D. Kwak, M. C. Bell, K.-S. C. Blair, and S. E. Bloom, “Cultural responsiveness in assessment, implementer training, and intervention in school, home, and community settings: a systematic review,” *Journal of Behavioral Education*, vol. 34, no. 3, pp. 526–573, Sep. 2025, doi: 10.1007/s10864-024-09547-7.
[141] F. Naseer, R. Tariq, H. M. Alshahrani, N. Alruwais, and F. N. Al-Wesabi, “Project based learning framework integrating industry collaboration to enhance student future readiness in higher education,” *Scientific Reports*, vol. 15, no. 1, p. 24985, Jul. 2025, doi: 10.1038/s41598-025-10385-4.
[142] O. Østerlie and C. M. Killian, “Flipped learning in physical education: evolutions over a decade and future horizons,” *Physical Education and Sport Pedagogy*, pp. 1–16, Apr. 2025, doi: 10.1080/17408989.2025.2497326.
[143] Y. Qu, M. X. Y. Tan, and J. Wang, “Disciplinary differences in undergraduate students’ engagement with generative artificial intelligence,” *Smart Learning Environments*, vol. 11, no. 1, p. 51, Nov. 2024, doi: 10.1186/s40561-024-00341-6.
[144] K. H. Arar, H. Özen, G. Polat, and S. Turan, “Artificial intelligence, generative artificial intelligence and research integrity: a hybrid systemic review,” *Smart Learning Environments*, vol. 12, no. 1, p. 44, Jul. 2025, doi: 10.1186/s40561-025-00403-3.
[145] S. Zang, P. Lin, X. Xia, Y. Bai, and H. Deng, “Mediation mechanisms in the hierarchical evolution of high-order thinking skills: a longitudinal study in Chinese science education,” *Thinking Skills and Creativity*, vol. 59, p. 101996, Mar. 2026, doi: 10.1016/j.tsc.2025.101996.
[146] X. Huang, M. Wilson, and L. Wang, “Exploring plausible causes of differential item functioning in the PISA science assessment: language, curriculum or culture,” *Educational Psychology*, vol. 36, no. 2, pp. 378–390, Feb. 2016, doi: 10.1080/01443410.2014.946890.
[147] S. Al-Busaidi and F. Al-Seyabi, “Project-based learning as a tool for student-teachers’ professional development: a study in an Omani EFL teacher education program,” *International Journal of Learning, Teaching and Educational Research*, vol. 20, no. 4, pp. 116–136, Apr. 2021, doi: 10.26803/ijlter.20.4.7.
[148] R. Belwal, S. Belwal, A. B. Sufian, and A. Al Badi, “Project-based learning (PBL): outcomes of students’ engagement in an external consultancy project in Oman,” *Education + Training*, vol. 63, no. 3, pp. 336–359, 2021, doi: 10.1108/ET-01-2020-0006.

BIOGRAPHIES OF AUTHORS

Nia Kania is an associate professor at the Mathematics Education Study Program, Universitas Majalengka. She received her Ph.D. in Mathematics Education from Universitas Pendidikan Indonesia. She has conducted research and published numerous works in the field of mathematics education, particularly in improving mathematical abilities. She can be contacted at email: niakania@unma.ac.id.

Florence Kyaruzi is an associate professor in Psychometry and Mathematics Education, Department of Educational Psychology and Curriculum Studies at the University of Dar es Salaam. He holds a Ph.D. in the Learning Sciences from Ludwig-Maximilians-Universität München (LMU), Germany. His research expertise includes student and teacher conceptions of assessment, mathematics education, gender issues in education, teacher education, feedback, and mentorship. He can be contacted at email: sakyaruzi@gmail.com; florence.kyaruzi@duce.ac.tz.