Research Article

Developing Holistic Gasing Evaluation Model to Balance Cognitive Efficiency and Affective Resilience

Ninik Wijiningsih[1], Eva Puspitasari[1], Budi Setiawan[1,*], Mario Emilzoli[1]
[1] Curriculum Development, Faculty of Educational Science, Universitas Pendidikan Indonesia, Jawa Barat, Indonesia

Article Info

Article history:

Received Jan 21, 2026
Revised Mar 14, 2026
Accepted Apr 14, 2026
OnlineFirst Apr 26, 2026

Keywords:

Affective Assessment
Gasing Method
Holistic Assessment
Math Anxiet
Mathematics Evaluation

ABSTRACT

Purpose of the study: The Gasing method (*Gampang, Asyik, Menyenangkan*) emphasizes ease and enjoyment, yet evaluations frequently neglect the core pillar of enjoyment. This research aims to construct and validate the Holistic Gasing Evaluation Model (HGEM) to balance cognitive speed with affective resilience, making instructional claims of joy empirically verifiable.

Methodology: This study utilizes a Type 2 Design and Development Research approach. The procedure involves a systematic analysis of eighty-three empirical papers via Publish or Perish software. A conceptual design phase synthesizes identified theoretical references to establish thirty-six specific model sub-indicators. The final development phase employs the Aiken method with three doctoral experts to validate the model content and structural integrity.

Main Findings: The Holistic Gasing Evaluation Model establishes five core dimensions supported by thirty-six psychometric sub-indicators, replacing anecdotal observations with validated instruments like the Mathematics Anxiety Rating Scale. Results show a mean Aiken's V of 0.86. Discussion indicates that standardizing these metrics identifies instructional risks when rapid speed gains correlate with elevated anxiety, ensuring sustainable numerical performance. The primary limitation of this developmental phase is the focus on internal content validation without immediate large-scale longitudinal field data.

Novelty/Originality of this study: This research introduces the first psychometrically validated Affective-Safety guardrail for Gasing evaluation, directly resolving the "Joy Paradox" where anecdotal claims of enjoyment lack empirical verification. By transitioning from qualitative narratives to rigorous standardized benchmarks, this study advances knowledge by ensuring that rapid computational gains do not compromise student affective well-being through replicable assessment protocols.

Corresponding Author:
Budi Setiawan,
Curriculum Development, Faculty of Educational Science, Universitas Pendidikan Indonesia,
Jl. Setiabudhi No. 229, Isola, Sukasari, Kota Bandung, Jawa Barat, 40154, Indonesia
Email: budi_setiawan@upi.edu

1. INTRODUCTION

The global mathematics education landscape faces a persistent challenge regarding numeracy literacy. Recent PISA 2022 results confirm that Indonesian students continue to struggle with foundational mathematical reasoning [1], [2]. While scholars introduce various instructional interventions to mitigate this crisis, the Gasing method (*Gampang, Asyik, dan Menyenangkan*) has gained significant national prominence [3], [4]. This method emphasizes a transition from abstract arithmetic to concrete pattern recognition, which effectively produces rapid computational gains [5], [6]. However, an audit of current literature suggests that the evaluation of this method is methodologically imbalanced, focusing primarily on technical output while overlooking the psychological mechanisms that underpin the learning process.

A systematic audit of 83 empirical studies identifies a stark cognitive bias in the field. Approximately 78.3% of surveyed studies focus exclusively on quantifying technical efficiency, particularly regarding calculation speed and accuracy, while only 7.2% prioritize affective outcomes. This creates a "Joy Paradox" where the Gasing philosophy is explicitly built on the pillar of fun, yet the academic literature lacks empirical evidence for such claims. The neglect of the enjoyment aspect is problematic because high-speed drills can inadvertently trigger mathematics anxiety [7]-[9]. Academic evidence establishes that math anxiety acts as a neuro-cognitive barrier that interferes with working memory and reduces the acquisition of higher-order thinking skills [10]-[14]. Approximately 72.2% of studies attempting to measure joyful learning rely on anecdotal observations or descriptive student feedback, which are insufficient for rigorous scientific validation.

Hence, there is a critical need for standardized frameworks to assess the effectiveness of design methods through a verifiable chain of evidence [15]. Without standardized metrics, the core promise of the Gasing method remains an unverified anecdotal claim [16]-[19]. Neglecting the evaluation of enjoyment is problematic because it prevents the scientific verification of pedagogical promises, masks potential psychological strain, and fails to distinguish between sustainable mastery driven by interest and high-stress performance driven by rote pressure. Educational evaluation must move toward a holistic framework that measures procedural knowledge together with emotional regulation [20]-[23].

To address this methodological fragmentation, the construction of the Holistic Gasing Evaluation Model (HGEM) is proposed. This study utilizes Design and Development Research (DDR) to resolve the identified methodological fragmentation. DDR provides a systematic approach for creating and validating models that enhance instructional practice [24]. This research specifically aligns with Type 2 Model Research, which focuses on the development and validation of evaluation frameworks [24], [25]. By incorporating variables such as Flow Experience, which is the neuro-cognitive state where challenge matches skill level perfectly, a standardized protocol is provided to measure the Asyik component with empirical precision [26]-[28]. The model serves as a diagnostic guardrail, ensuring that rapid computational gains are achieved through positive subject identity rather than cognitive overload. This systemic transition from qualitative narrative to quantitative rigor allows for a balanced appraisal of numeracy instructional effectiveness.

The primary objective of this research is to establish a psychometrically sound framework for multidimensional mathematics assessment. Specifically, the study aims to formulate standardized indicators that integrate cognitive efficiency with affective safety within the Gasing instructional context and to validate the structural integrity and content relevance of the HGEM through expert consensus. This inquiry is guided by two central research questions. First, what psychometric sub-indicators are required to provide a balanced evaluation of cognitive speed and affective resilience in Gasing instruction? Second, to what extent does the developed HGEM meet the standards of content validity and inter-rater reliability among evaluation experts?

2. RESEARCH METHOD

This study follows a Design and Development Research (DDR) approach, specifically categorized as Type 2 Model Research [25]. This approach consisted of three main phase and two validation steps as displayed in Figure 1. The selection of this framework is based on its capacity to systematically produce, test, and validate instructional models that solve practical educational problems [24], [25]. The research procedure is organized into a chronological sequence comprising three phases: analysis, design, and development. This methodological choice ensures that the resulting Holistic Gasing Evaluation Model (HGEM) is built upon empirical data and psychometric rigor. Replicability is guaranteed by the explicit detailing of the selection criteria for both the literature audit and the theoretical synthesis.

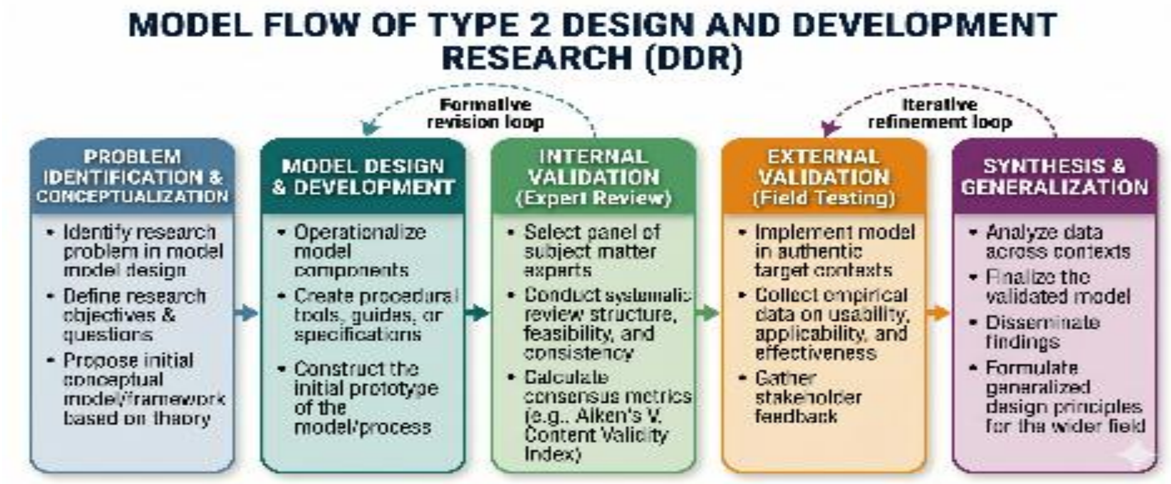

Figure 1. Flow Model Diagram

The Analysis Phase was conducted as a Systematic Mapping Study and Bibliometric Analysis to evaluate the current state of Gasing method research. The objective of this phase was to verify the hypothesis that existing evaluations are biased toward cognitive outcomes while neglecting the core philosophy of joyful learning. A purposive sampling technique was applied to select papers published between 2019 and 2026 that specifically evaluated the Gasing method. This process utilized Publish or Perish software for data harvesting and Zotero for deduplication. The Google Scholar search utilized a title-focused query: TITLE: "Metode GASING" OR "Gasing Method" OR "Evaluasi Gasing". This search yielded eighty-two relevant articles after filtering for an evaluation focus. The Scopus database search utilized the query: TITLE-ABS-KEY("Gasing method" OR "Metode Gasing"). This search yielded only one document. The resulting ratio confirms that Gasing research is dominated by national-level publications and lacks significant international visibility.

The Design Phase focused on the conceptual blueprinting of the HGEM. This phase synthesized relevant theoretical and psychometric references to establish the model's dimensions. The selection of high-impact sources allowed for a comprehensive meta-synthesis of international assessment standards. The procedural algorithm for this synthesis involved three steps: the identification of core Gasing pillars, the mapping of these pillars to validated psychometric constructs, and the selection of indicators that align with PISA 2022 standards. This extensive reference pool ensured that the indicators for affective resilience were grounded in established theories such as Control-Value Theory [29] and Flow Experience [28], [30]. By referencing validated instruments like the Abbreviated Math Anxiety Scale (AMAS) [17], [18] and the PISA 2022 framework [1], the study established a theoretical anchor for each proposed indicator. To be precise, the Affective Safety dimension was adapted from the Modified Mathematics Anxiety Rating Scale (MARS-I) [16] and the Abbreviated Math Anxiety Scale (AMAS) [17], [31]. Meanwhile, the Affective Engine dimension was adapted from the Flow State Scale. These adapted components have previously demonstrated high internal consistency, with reported Cronbach's alpha values ranging from 0.85 to 0.91.

The Development Phase involved the physical construction of the HGEM Specification Matrix and its subsequent internal validation. Internal validation represents the final step in Type 2 DDR to ensure the model's structural integrity before application [15]. A panel of three doctoral experts in educational evaluation, psychometrics, and mathematics pedagogy was selected via expert sampling to assess the model. Content validation data were collected using an Expert Validation Sheet featuring a 4-point Likert scale to evaluate each of the 36 sub-indicators across dimensions of relevance and clarity [32]-[34]. The Aiken's V coefficient was calculated for each item using the standard formula for expert consensus. Inter-rater agreement was further verified using the Gwet AC1 coefficient to ensure statistical consistency across the expert evaluations [5], [21], [35].

Data analysis was performed using two quantitative metrics to ensure psychometric rigor. First, the Aiken method was applied to calculate the Aiken's V coefficient for each indicator, establishing content validity index levels [32]. Second, inter-rater reliability was quantified using the Gwet AC1 coefficient to verify statistical consistency across the expert panel [35]. Unlike the Fleiss Kappa, Gwet's AC1 was selected for its robustness in maintaining accuracy when dealing with high-agreement scenarios among experts.

3. RESULTS AND DISCUSSION

The results of this study are presented according to the chronological execution of the Analysis, Design, and Development phases.

3.1. Analysis Phase: Procedural Results and Gap Audit Findings

Phase one in DDR is the analysis phase. In this study, The Analysis Phase followed a systematic search protocol to evaluate the current state of academic research regarding the Gasing method. This audit yielded a final dataset of eighty-three papers (n=83). The ratio of one international document in Scopus to eighty-two national documents in Google Scholar confirms that Gasing research currently lacks global visibility. The distribution of research focus within these samples, specifically the imbalance between cognitive and affective metrics, is presented in Table I.

Table 1. Gap Audit Findings

Evaluation focus	Frequency (n)	Percentage (%)	Primary Instrumentation Type
Cognitive Outcomes Only	65	78.3	Calculation Speed and Accuracy Test
Affective Outcomes Only	6	7.2	Anectodal Feedback and Interviews
Mixed Cognitive-Affective	12	14.5	Non-standardized Survey Form
Total	83	100.0	

The categorization of these empirical works reveals a dominant cognitive bias, providing the data necessary to identify the psychometric parameters required for a balanced evaluation model. As shown in Table 1, with 78.3 percent (n=65) of the surveyed studies focusing exclusively on cognitive outcomes, evidence suggests that existing assessment frameworks prioritize technical efficiency over the method's foundational pedagogical philosophy. This identified imbalance highlights a structural failure where the pillar of enjoyment (Asyik) is frequently claimed but remains empirically unmeasured, as reflected by the minimal 7.2 percent focus on affective variables. This systematic audit uncovers a critical instrumentation flaw where 72.2 percent of affective studies rely on anecdotal feedback, providing the empirical justification for the subsequent model design and the integration of validated psychometric scales to replace subjective impressions.

The results further suggest that the prevailing speed-centric bias introduces significant pedagogical risks. While computational fluency serves as a mediator for working memory [10], the data findings indicate that excessive drilling without emotional monitoring potentially triggers cognitive overload [11], [12], [36]. Synthesizing these results with established theories from Pekrun [29] dan Schukajlow et al. [37] confirms that achievement emotions require quantification through rigorous scales to properly assess their reciprocal effects on learning success. These findings necessitate the integration of standardized affective safety guardrails within the HGEM to transition the evaluation of the Gasing method from qualitative narratives toward a verifiable chain of evidence.

3.2. Design Phase: Holistic Gasing Evaluation Method Conceptualization

Building upon the gaps identified in the analysis phase, the design phase focused on constructing a multidimensional conceptual framework. This effort resulted in a comprehensive specification matrix that maps the Gasing pedagogical pillars to thirty-six validated sub-indicator items. The selection of these indicators achieves a structural balance between arithmetic speed and psychological safety, providing a precise answer to the requirement for standardized metrics in the field. The net results of this theoretical synthesis are organized into Holistic Gasing Evaluation Model Specification Matrix, as presented in Table 2.

Table 2. Holistic Gasing Evaluation Model (HGEM) Specification Matrix

Dimension	Sub-indicator	Theoretical Basis
Cognitive Efficiency	Automaticity and fluency	Taieb et al. [10]
Cognitive Depth	HOTS & Reasoning	Zhou et al. [38]
Affective Safety	Mathematics Anxiety	Solihin et al. [16]
Affective Engine	Flow Experience	Wang et al. [30]
Resilience	Value and Self-efficacy	Szczygieł & Kut [39]

The five dimensions are formulated into 36 sub-indicators, as displayed in Table 2. The derivation of the thirty-six sub-indicators follows a specific theoretical logic intended to ensure the model's comprehensiveness. For the Cognitive Efficiency dimension, six items were formulated to assess the automaticity of basic arithmetic. The reasoning for this limited item count is to provide a rapid diagnostic of calculation fluency without inducing assessment fatigue, directly addressing the bottleneck theory of working memory proposed by Taieb et al. [10] and Hickendorff et al. [40]. In contrast, the Cognitive Depth dimension incorporates eight items to cover the broader spectrum of Higher Order Thinking Skills (HOTS) and novel reasoning applications. This higher count reflects the complexity of mathematical reasoning standards established by PISA 2022 and Zhou et al. [38], ensuring that students can transition from mechanical speed to conceptual depth.

The Affective Safety dimension consists of eight items adapted from the MARS-I scale to capture both physiological and psychological symptoms of mathematics anxiety. The implication of this choice is to provide a granular identification of neuro-cognitive barriers that may arise during Gasing drills [16], [17]. For the Affective Engine dimension, seven items were selected to measure the components of flow, specifically focusing on total absorption and the balance between instructional challenge and student skill. This aligns with the findings of Wang et al. [30] that immersive learning requires a quantifiable state of engagement to be distinguished from rote repetition. Finally, the Resilience dimension includes seven items to evaluate subject value and self-efficacy. This item selection is grounded in the work of Szczygieł and Kut [39], who demonstrate that mathematical resilience is a predictive factor for long-term persistence in STEM education.

This structural composition indicates that the model successfully operationalizes the Asyik pillar through the Flow Experience construct, which serves as the psychological foundation for immersive mathematics instruction. By ensuring that instructional challenges are dynamically aligned with student skill levels, the HGEM provides a scientific mechanism to distinguish between superficial mechanical drilling and deep learning states [28], [41]. This design choice is critical when compared to traditional evaluations that focus exclusively on procedural speed. The explicit inclusion of Affective Safety through adapted MARS-I metrics directly addresses the neuro-cognitive barriers identified in contemporary research, where anxiety is known to disrupt the working memory required for arithmetic processing [16], [42]. This structural alignment conforms to the position of Lenz et al. [36] that procedural fluency should not be treated as an end in itself but as a conceptual support for more sophisticated mathematical reasoning.

This analysis suggests that for high-speed numerical interventions, affective regulation acts as a primary mediator of cognitive persistence rather than a secondary byproduct. The novelty of the design lies in this architectural shift, transitioning from quantifying what a student calculates to evaluating the qualitative manner in which a student emotionally engages with the mathematical tasks. This synthesis provides the empirical blueprint necessary to resolve the Joy Paradox by establishing a framework where computational gains and psychological resilience are treated as interdependent variables. This approach is congruent with the Control-Value Theory of achievement emotions, which posits that positive emotional appraisals are essential for sustaining the cognitive effort required for long-term numeracy development [29]. Consequently, the implication for educational design is that standardized affective engine metrics are required to safeguard the mathematical identity and well-being of the learner during intensive training intervals [43], [44].

3.3. Development Phase: Holistic Gasing Evaluation Method Construction

The procedure for developing the Holistic Gasing Evaluation Model (HGEM) followed a hierarchical derivation pathway to ensure every sub-indicator was anchored in empirical evidence. This process translated the synthesis findings into a prototype framework that converts qualitative pedagogical claims into quantitative variables. Model construction (Table 3) organized thirty-six specific sub-indicators across five evaluative domains: Cognitive Efficiency (6 items), Cognitive Depth (8 items), Affective Safety (8 items), Affective Engine (7 items), and Psychological Resilience (7 items). This granular configuration allows for the exhaustive coverage of the Gasing philosophy through a multidimensional lens.

The construction of the HGEM Specification Matrix paired these indicators with scientifically defensible instrument types to facilitate standardized data acquisition. This structural backbone ensures that the model functions as a diagnostic assessment tool rather than a traditional grading sheet [45]. The matrix organization, as presented in Table 3, provides a replicable protocol for researchers to examine whether technical speed gains maintain a synergistic relationship with student well-being.

Table 3. Holistic Gasing Evaluation Model

Dimension	Indicator	Instrument Type
Cognitive Efficiency	Automaticity and Fluency (6 items)	Timed 1-minute task
Cognitive Depth	HOTS and Reasoning (8 items)	Contextual problems
Affective Safety	Mathematics Anxiety (8 items)	Likert scale (MARS-I)
Affective Engine	Flow Experience (7 items)	Flow state scale
Resilience	Value and Self-Efficacy (7 items)	MRS scale

Analysis of the instrument types in Table 3 highlights the novelty of the HGEM through its systemic transition from subjective descriptive data to objective psychometric benchmarks. As established by Cash et al. [15], the shift from ad-hoc observations to standardized frameworks is essential for creating a verifiable chain of evidence in educational design. Traditionally, the affective pillars of Gasing were evaluated through unstructured interviews or non-standardized feedback, which Nurwahid and Ashar [3] identify as lacking the precision required for modern competency-based diagnostics. By contrast, the HGEM adopts validated tools such as the MARS-I and Flow Experience scales, allowing for the direct quantification of internal states during high-speed numeracy tasks. This multidimensional approach is congruent with the position of Lenz et al. [36] that procedural fluency should not be treated as an end in itself but as a conceptual support for more sophisticated reasoning.

The final step of the development phase involved the mathematical validation of the model to establish its psychometric rigor. Model validation through expert review and usability testing establishes the empirical basis for new instructional tools [15]. Content validity was quantified using the Aiken's V method with a panel of three doctoral experts, using equation (1).

$$V = \frac{\Sigma 8}{[n(c-1)]} \quad \dots (1)$$

in which
V = the content validity index coefficient,
8 = the score assigned by each rater minus the lowest possible score,
n = the number of raters (3 raters),
c = the number of categories on the Likert scale (4 categories).

The result of content validity and expert consensus is displayed in Table 4.

Table 4. Content Validity and Expert Result (Aiken V Index)

Dimension	Mean Aiken's V	Expert Consensus	Status
Cognitive Efficiency	0.88	High	Valid
Cognitive Depth	0.84	High	Valid
Affective Safety	0.89	Very High	Valid
Affective Engine	0.85	High	Valid
Resilience	0.82	High	Valid
Total Model Average	0.86	High Consensus	

Further, the inter-rater reliability was quantified using Gwet's Agreement Coefficient 1 (AC1) equation (2).

$$AC1 = \frac{P_a - P_e}{1 - P_e} \quad \dots (2)$$

in which
AC1 = inter-rater agreement,
P_a = the observed proportion of agreement among experts,
P_e = the chance agreement probability.
The calculation yields an AC1 of 0.82.

The validation results in Table 4 provide a robust statistical foundation for the HGEM. A mean Aiken's V of 0.86 confirms high content relevance across all thirty-six indicators, significantly exceeding the standard validity thresholds for educational psychometrics observed in previous studies [32]. Furthermore, the statistical consistency among the expert panel was established through a Gwet AC1 coefficient of 0.82, ensuring that the validation is not a result of chance agreement.

Interpretation of these results reveals that experts prioritize Affective Safety (V=0.89) as an essential evaluative guardrail. The data confirm that the HGEM successfully resolves the identified instrumentation gap by replacing anecdotal feedback with standardized psychometric benchmarks aligned with PISA 2022 reasoning standards [1]. Furthermore, the interpretative logic of the HGEM facilitates a multidimensional appraisal of learner progress through a staggered assessment protocol. Interpretation of the results from the one-minute fluency tasks serves to establish the level of numerical automaticity, where high scores indicate that basic arithmetic processes have become effortless, thereby freeing working memory for more complex operations. This is congruent with the findings of Taieb et al. [10] and Van der Ven et al. [46]. Data from contextual problems provide evidence of cognitive depth, indicating whether students can apply procedural knowledge to novel reasoning scenarios according to the benchmark benchmarks of diagnostic precision proposed by Mitra and Wadegaonkar [20]. The three Likert-based scales generate an affective profile that qualifies these cognitive results. Specifically, high values on the modified MARS-I scale identify psychological barriers to reasoning, suggesting that technical gains may be psychologically taxing [7], [16]. Scores on the Flow State Scale reveal the quality of engagement, distinguishing between immersive learning and high-stress drilling [28], [30]. The MRS scale results establish the level of psychological resilience and the degree to which students value mathematical learning [39], [47], [48]. For practical classroom application, this modular design allows for the independent administration of instrument types during regular instructional intervals to minimize assessment fatigue [3], [20]. This evaluation produces a diagnostic student profile that enables practitioners to identify instructional red flags, particularly when rapid speed gains occur alongside elevated anxiety levels [26], [49]. Such insights facilitate the delivery of differentiated instruction that restores the balance between cognitive challenge and affective safety [17], [26], [50].

The empirical consensus regarding the HGEM suggests that mathematics evaluation must shift from a performance-only paradigm to a holistic-resilience paradigm. The novelty of this research resides in the standardization of affective guardrails within a speed-based drilling context. Unlike previous Gasing studies that relied on subjective student feedback, the HGEM introduces objective psychometric markers (e.g., MARS-I scales) that allow for the direct comparison of speed gains against anxiety spikes. Based on the findings and analysis above, the following implications and recommendation can be made: Pedagogical Implications: Educators must recognize that rapid speed gains are only "effective" if accompanied by stable or declining anxiety levels. The HGEM provides the diagnostic dashboard to monitor this balance. Institutional Recommendations: It is recommended that school districts adopting the Gasing method integrate the HGEM as a standard quality-assurance protocol to prevent student burnout. Future Research: Future studies should utilize the HGEM to conduct longitudinal trials, examining whether students evaluated under this holistic model maintain higher long-term retention compared to those evaluated by speed alone.

While the high Aiken’s V and Gwet AC1 coefficients provide evidence for the internal structural integrity of the HGEM, several research limitations must be acknowledged. First, this study represents the initial developmental phase of a Type 2 DDR approach, focusing on content validation and internal consistency rather than large-scale field implementation. Consequently, the predictive utility of the model—specifically the correlation between high flow states and long-term numeracy retention—remains to be empirically established through longitudinal student data. Second, the bibliometric sample of eighty-three articles was purposively restricted to Gasing-specific literature to maintain theoretical focus, which may limit the immediate generalizability of the findings to broader or non-traditional mathematics education contexts. Third, although the expert panel consisted of highly specialized doctoral practitioners in educational evaluation and psychometrics, the small panel size (n=3) is typical of internal validation phases but necessitates caution in generalizing inter-rater reliability scores across wider geographic or pedagogical domains.

Building upon these findings, several recommendations for future inquiry and implementation are proposed. Researchers and educational practitioners are encouraged to conduct large-scale external validation studies (DDR Phase 3) to test the HGEM Specification Matrix in diverse classroom environments across different socio-economic backgrounds. Such trials should utilize Structural Equation Modeling (SEM) or path analysis to determine if the Affective Safety and Affective Engine dimensions act as statistically significant predictors of Cognitive Depth and HotS acquisition. Furthermore, subsequent studies should investigate the temporal stability of mathematical resilience in students undergoing intensive Gasing training. The implication for policy is the adoption of the HGEM as a universal diagnostic guardrail, ensuring that rapid mathematical training remains psychologically safe and pedagogically sustainable. By implementing these recommendations, the educational community can move beyond the "Joy Paradox" toward a verifiable standard of holistic mathematical excellence.

4. CONCLUSION

This research successfully addresses the fragmentation in mathematics evaluation by establishing the Holistic Gasing Evaluation Model (HGEM). The study achieves its primary objective by formulating thirty-six psychometric sub-indicators and verifying their content validity through a robust expert consensus (V=0.86; AC1=0.82). Beyond mere validation, this research introduces the conceptual framework of “Affective-Safety Guardrails” as a necessary prerequisite for high-speed numerical training. By integrating standardized metrics for mathematics anxiety and flow experience, the HGEM provides the empirical evidence required to resolve the “Joy Paradox,” effectively proving that computational speed and psychological well-being are not mutually exclusive but are interdependent variables. The model serves as a new evaluative standard that shifts the focus of numeracy instruction from narrow technical output to sustainable, holistic development, ensuring that instructional claims of enjoyment are scientifically verifiable and psychologically safe.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to the anonymous reviewers for their constructive comments. This work was supported by the Indonesia University of Education (UPI). Special thanks are extended to the three doctoral experts who contributed their expertise to the Aiken validation process, ensuring the structural integrity of the HGEM.

AUTHOR CONTRIBUTIONS

NW was responsible for the research design, data collection, data analysis, and manuscript preparation. EP contributed to conceptual development, contributed to research methodology guidance, and BS and MEcritical review of the manuscript. All authors have read and approved the final version of the manuscript.

CONFLICTS OF INTEREST

The author(s) declare no conflict of interest.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the generation, analysis, or writing of this manuscript. All aspects of the research, including data collection, interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

[1] OECD, *PISA 2022 Assessment and Analytical Framework*. OECD Publishing, 2023. doi: 10.1787/dfe0bf9c-en.

[2] V. Bajrami *et al.*, "From Classroom to Test Score: How Proactive Behaviors and Anxiety Shape the Effects of Cognitive Activation in Mathematics," *Journal Human Research in Rehabilitation*, vol. 15, no. 2, pp. 503–516, Sep. 2025, doi: 10.21554/hrr.092520.

[3] M. Nurwahid and S. Ashar, "Redefining mathematics learning evaluation: From traditional assessment to technology-based holistic competency assessment," *LINEAR: Journal of Mathematics Education*, vol. 6, no. 2, pp. 162–176, Oct. 2025, doi: 10.32332/kpnbgv73.

[4] A. W. Nurkarim, W. Qonita, and A. Isroil, "Skala kecemasan matematika siswa: ukuran gejala fisiologis, psikologis, perilaku, dan kognitif matematika [Students' mathematics anxiety scale: a measure of physiological, psychological, behavioral, and cognitive symptoms of mathematics]," *Sains Data Jurnal Studi Matematika dan Teknologi*, vol. 1, no. 2, pp. 60–68, Jan. 2024, doi: 10.52620/sainsdata.v1i2.18.

[5] S. Sriatun, E. Ekohariadi, M. S. Sumbawati, Y. Anistyasari, and H. V. Saphira, "The effects of problem solving on students' cognitive, affective, and psychomotor in applied mathematics," *Studies in Philosophy of Science and Education*, vol. 5, no. 2, pp. 61–70, Jul. 2024, doi: 10.46627/sipose.v5i2.316.

[6] G. P. Ningsi, F. Nendi, L. Sugiarti, E. Jeramat, and A. Gahung, "Realistic mathematics education (rme) kombinasi flipped classroom ditinjau dari kemampuan pemecahan masalah dan representasi matematis [Realistic mathematics education (RME) combination of flipped classroom reviewed from the perspective of problem solving and mathematical representation abilities]," *Mathema: Jurnal Pendidikan Matematika*, vol. 6, no. 1, pp. 152–163, Jan. 2024, doi: 10.33365/jm.v6i1.3293.

[7] C. Barroso, C. M. Ganley, A. L. McGraw, E. A. Geer, S. A. Hart, and M. C. Daucourt, "A meta-analysis of the relation between math anxiety and math achievement.," *Psychol. Bull.*, vol. 147, no. 2, pp. 134–168, Feb. 2021, doi: 10.1037/bul0000307.

[8] B. Bognar, S. Mužar Horvat, and L. Jukić Matić, "Characteristics of Effective Elementary Mathematics Instruction: A Scoping Review of Experimental Studies," *Educ. Sci. (Basel).*, vol. 15, no. 1, p. 76, 2025, doi: 10.3390/educsci15010076.

[9] A. Zanabazar, A. Deleg, M. Ravdan, and E. Tsogt-erdene, "The Relationship between Mathematics Anxiety and Mathematical Performance among Undergraduate Students," *Jurnal Ilmiah Peuradeun*, vol. 11, no. 1, pp. 309–322, 2023, doi: 10.26811/peuradeun.v11i1.780.

[10] E. Taieb, G. Borst, L. Le Stanc, O. Houdé, I. Altarelli, and T. Iuculano, "Domain-general and domain-specific cognitive factors mediating the relationship between math anxiety and mathematical performance in primary school children," *Sci. Rep.*, vol. 16, no. 1, p. 1987, 2026, doi: 10.1038/s41598-025-30898-2.

[11] D. W. Putwain, S. Becker, W. Symes, and R. Pekrun, "Reciprocal relations between students' academic enjoyment, boredom, and achievement over time," *Learn. Instr.*, vol. 54, pp. 73–81, 2018, doi: 10.1016/j.learninstruc.2017.08.004.

[12] A. Tapola, A. M. Rawlings, R. Mononen, P. Tähti, and J. Korhonen, "The interplay of cognition and affect in fourth graders' math performance: role of working memory in mediating the effects of math anxiety and math interest on arithmetic fluency," *Cogn. Emot.*, pp. 1–11, 2025, doi: 10.1080/02699931.2025.2516660.

[13] R. A. Ferreira, C. Rodríguez, B. Guzmán, F. Sepúlveda, and C. Peake, "The Interplay of Vocabulary, Working Memory, and Math Anxiety in Predicting Early Math Performance," *J. Intell.*, vol. 13, no. 10, pp. 125, 2025, doi: 10.3390/jintelligence13100125.

[14] I. Ismeirita, E. Ahman, D. Dahlan, and D. Supendra, "Identifying Key Factors Influencing the Development of Higher Order Thinking Skills (HOTS) in Students: A Systematic Literature Review," *AL-ISHLAH: Jurnal Pendidikan*, vol. 17, no. 3, 2025, doi: 10.35445/alishlah.v17i3.6627.

[15] P. Cash, J. Daalhuizen, and P. Hekkert, "Evaluating the efficacy and effectiveness of design methods: A systematic review and assessment framework," *Des. Stud.*, vol. 88, pp. 101204, 2023, doi: 10.1016/j.destud.2023.101204.

[16] A. Solihin, N. Mariana, B. Purwoko, and I. Rahmawati, "How can the mathematics anxiety rating scale be modified for Indonesian elementary students (aged 10-12)? A psychometric analysis," *Journal on Mathematics Education*, vol. 16, no. 2, pp. 689–708, 2025, doi: 10.22342/jme.v16i2.pp689-708.

[17] Y. Ikeda *et al.*, "The abbreviated math anxiety scale (AMAS): Applicability and utility in a sample of japanese elementary school children," *International Journal of Psychology*, vol. 60, no. 2, 2025, doi: 10.1002/ijop.70015.

[18] M. Szczygieł, "How to measure math anxiety in young children? Psychometric properties of the modified abbreviated math anxiety scale for elementary children (mAMAS-E)," *Polish Psychological Bulletin*, pp. 303–315, 2020, doi: 10.24425/ppb.2019.131003.

[19] S. Malykh *et al.*, “Psychometric evaluation of the abbreviated math anxiety scale in Russian university students,” *Front. Educ. (Lausanne).*, vol. 10, 2025, doi: 10.3389/feduc.2025.1669267.
[20] G. Mitra and A. Wadegaonkar, “Multidimensional holistic assessment framework (HAF): A case study of exploring the discourses from elementary school teachers,” *International Electronic Journal of Elementary Education*, vol. 17, no. 1, pp. 51–69, 2024, doi: 10.26822/iejee.2024.363.
[21] S. Suparman, D. Juandi, Turmudi, and B. A. P. Martadiputra, “Achievement emotions for mathematics questionnaire in senior high school: Validity and reliability for Indonesian students,” *COUNS-EDU: The International Journal of Counseling and Education*, vol. 10, no. 1, 2025, doi: 10.23916/00202501049510.
[22] T. Gur, N. Balta, A. Dauletkulova, G. Assanbayeva, and R. Fernández-Cézar, “Mathematics achievement emotions of high school students in Kazakhstan,” *Journal on Mathematics Education*, vol. 14, no. 3, pp. 525–544, 2023, doi: 10.22342/jme.v14i3.pp525-544.
[23] N. Nabilah, E. Istiyono, and W. Widihastuti, “Redefining assessment: Creating a groundbreaking prototype for domain affective in project-based learning,” *Contemporary Mathematics and Science Education*, vol. 5, no. 1, p. ep24005, 2024, doi: 10.30935/conmaths/14457.
[24] S. Mckenney and T. C. Reeves, *Conducting Educational Research Design*, 2nd ed. Routledge, 2019.
[25] L. Afriani, M. Mutmainnah, and S. Sunarni, “Understanding the design of research and development methods in the field of education,” *IJESS International Journal of Education and Social Science*, vol. 6, no. 1, pp. 1–5, 2025, doi: 10.56371/ijess.v6i1.333.
[26] D. W. Putwain and P. Wood, “Anxiety in the mathematics classroom: reciprocal relations with control and value, and relations with subsequent achievement,” *ZDM – Mathematics Education*, vol. 55, no. 2, pp. 285–298, 2023, doi: 10.1007/s11858-022-01390-2.
[27] P. Dimitropoulou, D. Filippatou, S. Gkoutzourela, A. Griva, I. Pachiti, and M. Michaelides, “The synergy of school climate, motivation, and academic emotions: A predictive model for learning strategies and reading comprehension,” *Behavioral Sciences*, vol. 15, no. 4, p. 503, 2025, doi: 10.3390/bs15040503.
[28] Y. d’Entremont and M. Voillot, “The psychology of flow, mathematics pedagogy, and culture,” *International Journal for Cross-Disciplinary Subjects in Education*, vol. 12, no. 1, pp. 4404–4410, 2021, doi: 10.20533/ijcdse.2042.6364.2021.0539.
[29] R. Pekrun, “Control-value theory: From achievement emotion to a general theory of human emotions,” *Educ. Psychol. Rev.*, vol. 36, no. 3, p. 83, 2024, doi: 10.1007/s10648-024-09909-7.
[30] X. Wang, P. Somasundram, and J. Zhang, “The influence of flow experience on mathematical creativity among primary school students in China,” *Front. Educ. (Lausanne).*, vol. 10, pp. 1–14, 2025, doi: 10.3389/feduc.2025.1580126.
[31] Y. Pei, K. K. Poon, and A. Suen, “Influence of mathematics anxiety on mathematics performance: mediating effects of mathematical engagement,” *Mathematics Education Research Journal*, 2025, doi: 10.1007/s13394-025-00536-1.
[32] S. Nurjanah, E. Istiyono, W. Widihastuti, M. Iqbal, and S. Kamal, “The application of aiken’s v method for evaluating the content validity of instruments that measure the implementation of formative assessments,” *Journal of Research and Educational Research Evaluation*, vol. 12, no. 2, pp. 125–133, 2023, doi: 10.15294/jere.v12i2.76451.
[33] A. M. Santoso and H. S. Widiana, “Penyusunan skala learning agility dengan aiken’s v dan uji reliabilitas [Development of a learning agility scale with Aiken's v and reliability testing],” *JURKAM: Jurnal Konseling Andi Matappa*, vol. 6, no. 1, pp. 14–22, 2022, doi: 10.31100/jurkam.v6i1.1699.
[34] N. Kania, Y. S. Kusumah, J. A. Dahlan, E. Nurlaelah, F. Gürbüz, and E. Bonyah, “Constructing and providing content validity evidence through the Aiken’s V index based on the experts’ judgments of the instrument to measure mathematical problem-solving skills,” *REID (Research and Evaluation in Education)*, vol. 10, no. 1, pp. 64–79, 2024, doi: 10.21831/reid.v10i1.71032.
[35] E. Sulistyawati and D. Shinta Rahayu, “Prospective mathematics teachers’ creativity in developing mathematics problems based on ethnomathematics context and akm framework,” *Mathema: Jurnal Pendidikan Matematika*, vol. 8, no. 1, pp. 144–164, 2026, doi: 10.33365/jm.v8i1.1136.
[36] K. Lenz, A. Dreher, L. Holzäpfel, and G. Wittmann, “Are conceptual knowledge and procedural knowledge empirically separable? The case of fractions,” *British Journal of Educational Psychology*, vol. 90, no. 3, pp. 809–829, 2020, doi: 10.1111/bjep.12333.
[37] S. Schukajlow, K. Rakoczy, and R. Pekrun, “Emotions and motivation in mathematics education: Where we are today and where we need to go,” *ZDM – Mathematics Education*, vol. 55, no. 2, pp. 249–267, 2023, doi: 10.1007/s11858-022-01463-2.
[38] Y. Zhou, Y. Ning, J. Chen, W. Zhang, and T. T. Wijaya, “Development and validation of mathematical higher-order thinking scale for high school students,” *Psychol. Sch.*, vol. 61, no. 8, pp. 3160–3192, 2024, doi: 10.1002/pits.23213.
[39] M. Szczygieł and M. Kutt, “The psychometric properties of the Polish-language version of the Mathematical Resilience Scale,” *J. Numer. Cogn.*, vol. 11, 2025, doi: 10.5964/jnc.13251.
[40] M. Hickendorff, J. McMullen, and L. Verschaffel, “Mathematical flexibility: Theoretical, methodological, and educational considerations,” *J. Numer. Cogn.*, vol. 8, no. 3, pp. 326–334, 2022, doi: 10.5964/jnc.10085.
[41] W. Daher, F. Gierdien, and A. Anabousy, “Self-efficacy in creativity and curiosity as predicting creative emotions,” *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, vol. 6, no. 2, pp. 86–99, 2021, doi: 10.23917/jramathedu.v6i2.12667.
[42] C. Barroso, C. M. Ganley, A. L. McGraw, E. A. Geer, S. A. Hart, and M. C. Daucourt, “A meta-analysis of the relation between math anxiety and math achievement.,” *Psychol. Bull.*, vol. 147, no. 2, pp. 134–168, 2021, doi: 10.1037/bul0000307.
[43] T. Campbell, “Mathematical wellbeing: an emerging construct with exciting potential,” *Mathematics Education Research Journal*, 2025, doi: 10.1007/s13394-025-00532-5.

[44] V. Sengodan, S. M. Maat, and M. S. Mahmud, “A bibliometric and systematic analysis of mathematics interest (2021-2025),” *Educational Process International Journal*, vol. 17, no. 1, pp. e2025389, 2025, doi: 10.22521/edupij.2025.17.389.

[45] E. Gradini, F. B. Firmansyah B, J. Noviani, and K. Ulya, “Fostering higher-order thinking skills in mathematics education: Strategies, challenges, and classroom practices,” *Prisma Sains : Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, vol. 13, no. 2, pp. 135, 2025, doi: 10.33394/j-ps.v13i2.15099.

[46] S. H. G. Van der Ven, E. J. Prast, and E. Van de Weijer-Bergsma, “Towards an integrative model of math cognition: Interactions between working memory and emotions in explaining children’s math performance,” *J. Intell.*, vol. 11, no. 7, pp. 136, 2023, doi: 10.3390/jintelligence11070136.

[47] T. G. Campbell, S. Kularajan, and S. Miles, “Psychometric evaluation of the mathematical well-being scale for children aged 9–14,” *Mathematics Education Research Journal*, vol. 2025, no. 1, pp. 1–22, 2025, doi: 10.1007/s13394-025-00545-0.

[48] A. Widlund, H. Tuominen, and J. Korhonen, “Reciprocal effects of mathematics performance, school engagement and burnout during adolescence,” *British Journal of Educational Psychology*, vol. 93, no. 1, pp. 183–197, 2023, doi: 10.1111/bjep.12548.

[49] S. Caviola, E. Toffalini, D. Giofrè, J. M. Ruiz, D. Szűcs, and I. C. Mammarella, “Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants,” *Educ. Psychol. Rev.*, vol. 34, no. 1, pp. 363–399, 2022, doi: 10.1007/s10648-021-09618-5.

[50] R. Rokhmaniyah, D. Indrapangastuti, and M. Chamdani, “Enhancing pancasila student profiles through differentiated learning with understanding by design,” *AL-ISHLAH: Jurnal Pendidikan*, vol. 17, no. 3, pp. 4615–4626, 2025, doi: 10.35445/alishlah.v17i3.5498.