LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

DEVELOPMENT AND CONTENT VALIDATION OF A CRITICAL THINKING SKILLS ASSESSMENT INSTRUMENT FOR GRADE 8 MATHEMATICS: EVIDENCE FROM MULTIDISCIPLINARY EXPERT JUDGMENT

Saba Irshad[*1], Saba Hanif[2], Asif Iqbal[3], Umber Shkeikh[4]

[*1]*PhD Scholar of Education, University of Education Lahore*
[2]*Assistant Professor of Education, University of Education Lahore*
[3]*Associate Professor of Education, University of Education Lahore*
[4]*Associate Professor of Mathematics, University of Education Lahore*

[*1]saba26074@gmail.com

DOI: https://doi.org/10.5281/zenodo.22158147

Keywords
critical thinking assessment; content validity; content validity index; expert judgement; mathematics education; Grade 8; instrument development

Article History
Received: 11 May 2026
Accepted: 21 June 2026
Published: 30 June 2026

Copyright @Author

Corresponding Author: *
Saba Irshad

Abstract

Critical thinking is an important educational goal, but measuring it across younger grades can be more challenging because it is multifaceted and can be demonstrated through content, language, and context. The current study presents the development of the Grade 8 mathematics Critical Thinking Skills Assessment Instrument (CTSAI) and evidence of content validation. This instrument was developed iteratively through construct identification, curriculum mapping, item generation, inter- and multidisciplinary expert review, and quantitative and qualitative analyses of expert feedback. Thirty items were written in three domains of Grade 8 Mathematics: Number Sequences and Patterns, Expansion and Factorization, and Linear Equations and Inequalities. The items were developed to call up five critical thinking processes operationalized for mathematics: interpretation, analysis, inference, evaluation, and self-regulation. Eight independent experts from psychology, mathematics education, English education, educational assessment, education administration, and education assessed each item on a four-point scale for relevance, clarity, grade-level appropriateness, critical-thinking alignment, and curriculum alignment. Item-level and scale-level content validity indices were computed along with a qualitative review of expert comments. The overall S-CVI/Ave was 0.999; all criteria except clarity had an S-CVI/Ave of 1.000, while clarity was 0.9958. There was unanimous consensus on the relevance criterion used in retention decisions for twenty-nine items, and the one item (Item 9) had an I-CVI of 0.875 and was retained with minor word changes. The results offer preliminary test content evidence, but lack evidence of reliability, factor structure, item functioning, and criterion-related validity. Therefore, the CTSAI is a content-valid instrument that needs field testing with Grade 8 students before making claims about broader measurement validity.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

1.INTRODUCTION

1.1 The Importance of Critical Thinking in Elementary Education

The expectation that today's students can interpret, evaluate, and analyze information; evaluate options; differentiate and justify ideas; and control their thinking has made this a major goal of education. These processes are especially central to meaningful mathematical activity because when mathematics is meaningful, relationships are discovered, multiple solution strategies are chosen and compared, arguments are assessed for validity, and explanations are offered as to why a solution is justified. Therefore, doing a maths task successfully on paper, without demonstrating higher-order thinking, may only be partial evidence of a student's higher-order thinking. (Abrami et al., 2015).

According to Facione (1990), critical thinking can be conceptualised as the processes of interpretation, analysis, evaluation, inference, explanation, and self-regulation. While these are conceptually separate, they can be implemented by disciplinary tasks. For instance, in school mathematics, interpretation can refer to understanding the meaning of a mathematical representation, analysis can refer to considering relationships between quantities, inference can refer to making a warranted conclusion, evaluation can refer to making a judgment about a solution or argument, and self-regulation can refer to checking and revising one's reasoning.

1.2 Challenges in Assessing Critical Thinking Among Elementary Learners

Evaluating critical thinking in younger students presents at least three measurement challenges. First, critical thinking exists at many levels, and not all items measure only factual recall. Secondly, language, prior knowledge, and familiarity with the disciplinary context are important in assessment performance. Theoretically suitable tests can therefore be unsuitable for students due to their language/cognitive development levels. Third, generic critical-thinking tests may not capture the expression of reasoning in a specific school subject. (Liu, 2014; Mooij et al., 2016; Davies & Barnett, 2015). For these reasons, instrument development must establish a defensible relationship among the construct definition, curriculum content, instrument attributes, and the targeted student population. This is where subject-matter, assessment, language, and educational specialists can help, as they may see issues that are not obvious from the item writer's perspective.

1.3 Need for a Curriculum-Aligned Instrument

Many existing critical-thinking assessments were created for adolescents, university students, or general reasoning contexts, but they have played important roles in the field of educational measurement. A Grade 8 mathematics instrument needs to balance mathematical content and comprehensibility; questions should include enough mathematical content to elicit reasoning. (Fabio & Patafi, 2026; Fabio et al., 2025; Piyaporn & Tuksino, 2025). The present study thus concentrates on creating and content-validating a Critical Thinking Skills Assessment Instrument (CTSAI) for grade 8 mathematics. The study is part of a larger doctoral thesis on cooperative learning, students' critical thinking, and academic achievement at elementary level. This is not a test of the effectiveness of cooperative learning, but it lays the foundation for the measurement instrument to be used in future field-based research.

1.4 Research Gap and Objectives

The study was stimulated by three gaps. First, there is still a need for measures that assess critical thinking in school-age children in a developmentally appropriate manner. Second, the disciplinary context is important: mathematical critical thinking involves reasoning with mathematical representations and mathematical concepts, not decontextualized mathematical problems about reasoning. Third, intervention research must provide defensible outcome measures so changes in student performance can be attributed to instructional strategies.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

Accordingly, the study addressed the following objectives:

- To develop a curriculum-aligned critical-thinking assessment instrument for Grade 8 mathematics.
- To obtain multidisciplinary expert evidence concerning the content validity of the instrument.
- To quantify expert agreement regarding item relevance, clarity, grade-level appropriateness, critical-thinking alignment, and curriculum alignment.

2. LITERATURE REVIEW

2.1 Conceptualizing Critical Thinking

Critical thinking is not one specific skill but a collection of related cognitive and metacognitive skills. The cognitive skills identified in Facione's (1990) Delphi consensus process are interpretation, analysis, evaluation, inference, explanation, and self-regulation. Similarly, Ennis (1987, 2018) stresses logical and thoughtful thinking focused on choosing what to believe or do. All these views focus on meaningful judgment, evidence, reasoning and reflection. (Ennis, 1987, 2018).

The multidimensionality of critical thinking poses an important design issue for assessment – an instrument should indicate which cognitive processes its items are designed to elicit for assessment. The main operational dimensions of the present CTSAI are, therefore, the 5 processes: interpretation, analysis, inference, evaluation, and self-regulation. This is because the current instrument was specifically designed around the mathematical reasoning processes represented in the study blueprint; therefore, explanation is not recorded as a separate scoring dimension.

2.2 Critical Thinking in Mathematics

Mathematical problem-solving tasks offer a rich environment for assessing students' mathematical thinking skills because they involve interpreting representations, identifying relationships, making and testing conjectures, comparing strategies, justifying procedures, and evaluating conclusions. The mathematical skills identified in research also focus on reasoning, problem solving, communicating, and justifying mathematical thinking (Niss & Højgaard, 2019).

Thus, a mathematics-specific measure can give evidence of critical thinking, while also being grounded in curricular content. (Monteleone et al., 2023; Su et al., 2016). This disciplinary orientation does not mean that critical thinking can only be measured in terms of mathematics achievement. Instead, it starts from the fact that students use the knowledge and representations within a given domain to demonstrate reasoning. The present study thus differentiates between mathematical content areas and critical thinking processes, and explicitly represents both in the instrument blueprint.

2.3 Instrument Development and Validity Evidence

A modern view of validity sees validity as an argument about the interpretations and uses of assessment scores, not a property of the test itself (AERA, APA, & NCME, 2014; Messick, 1995). Evidence based on test content is consequently an important early component of an overall validity argument. It asks if the content domain is well covered and if individual items are relevant, clear, and appropriate for the intended population. The instrument development process generally includes construct definition, theoretical and empirical review, instrument development, expert evaluation, pre-testing, and psychometric analysis (Boateng et al., 2018; DeVellis, 2021). This study focuses on the expert-review stage. It is therefore understood that this is preliminary validation work, not a full psychometric validation. (Boateng et al., 2018).

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

2.4 Content Validity and Expert Judgment

Structured evidence of the correctness of assessment items can be obtained from expert judgment regarding the construct they represent and whether they are appropriate to the intended context. One quantitative method used to summarize the expert agreement is the Content Validity Index (CVI). The I-CVI is the proportion of experts who rate an item as relevant at or above a specified score, and the S-CVI, or instrument content validity, is a summary of the content validity across the instrument.

Multiple experts typically agree that I-CVI scores of about .78 or greater are acceptable, while scores above .90 are considered strong evidence (Yusoff, 2019). CVI, however, should not be used as proof of all aspects of validity – qualitative comments and later empirical testing should not be overlooked. (Almanasreh et al., 2019; Zamanzadeh et al., 2015). (Polit & Beck, 2006; Polit et al., 2007; Yusoff, 2019).

Chance agreement is another consideration. Substantive agreement and chance agreement are not completely separate in percentage agreement and CVI. When appropriate, therefore, modified kappa can be reported as a complementary measure of performance (Polit et al., 2007). The one item not agreed upon by all was analyzed using modified kappa as an additional test of the durability of expert consensus.

2.5 Research Gap and Contribution

Based on the literature, key considerations for critical-thinking measures for younger learners include being developmentally appropriate, discipline-sensitive, and having multiple sources of validity evidence. The contribution of the present study is both procedural and substantive: mapping a Grade 8 mathematics instrument to critical-thinking processes, a multidisciplinary panel review, revisions based on expert feedback, and quantitative evaluation using CVI indices. The evidence produced is designed to assist, but not substitute for, the psychometric stages that are to be conducted in the field for a full validation argument. (Guo et al., 2024; Tian et al., 2025).

3. METHOD

3.1 Research Design

The study used an instrument development and content validation design. The methodological focus was the development and expert evaluation of assessment items for an instrument to measure critical thinking in Mathematics at grade level 8. The process was sequential and followed a logical sequence of identifying the procedure, mapping the curriculum, developing items, having experts review the items, analyzing the quantitative ratings, and refining the items qualitatively. This study is part of a larger doctoral research project focused on developing measures of cooperative learning. In the present paper, we did not analyze student outcome data, so claims are limited to evidence based on expert judgment and item content.

3.2 Instrument Development Process

Table:1

Stage	Activity	Operational description
Stage 1	Construct identification and theoretical review	Critical thinking was defined using established frameworks, particularly Facione's (1990) cognitive-skill framework.
Stage 2	Curriculum and content mapping	Three Grade 8 mathematics domains were selected: Number Sequences and Patterns; Expansion and Factorization; and Linear Equations and Inequalities.
Stage 3	Item generation	Thirty items were drafted to elicit interpretation, analysis, inference, evaluation, and self-regulation while maintaining grade-level language and mathematical relevance.
Stage 4	Expert review	Eight multidisciplinary experts independently evaluated all items using five criteria and a four-point rating scale.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

Stage 5	Quantitative analysis	I-CVI and S-CVI/Ave were calculated from ratings of 3 or 4.
Stage 6	Qualitative refinement	Expert comments were reviewed to identify wording, content, and alignment issues; Item 9 was refined for clarity.

3.3 Instrument Blueprint

The initial instrument had 30 items, evenly spread across three mathematics domains. Each domain had 10 items. The blueprint was designed to make the assessment more than just a test of procedural recall by having students think about mathematical information and reasoning.

Table:2

Mathematics domain	Items	Primary critical-thinking processes
Number Sequences and Patterns	1–10	Interpretation, analysis, inference, evaluation, self-regulation
Expansion and Factorization	11–20	Interpretation, analysis, inference, evaluation, self-regulation
Linear Equations and Inequalities	21–30	Interpretation, analysis, inference, evaluation, self-regulation

3.4 Expert Panel

Eight experts were purposively selected to offer complementary expertise in the construct, mathematics content, language, assessment, and education context. All panelists were PhD's. They had expertise in educational assessment, educational administration, educational psychology, English education, mathematics education, and education.

Table:3

Code	Qualification	Designation	Expertise	Validation role
E01	PhD	Assistant Professor	Psychology / Educational Psychology	Critical-thinking construct
E02	PhD	Associate Professor	Mathematics Education	Mathematical content
E03	PhD	Assistant Professor	Mathematics Education	Mathematical reasoning
E04	PhD	Assistant Professor	English Education	Language clarity
E05	PhD	Associate Professor	Educational Assessment	Assessment quality
E06	PhD	Associate Professor	Educational Administration	Educational appropriateness
E07	PhD	Assistant Professor	Education	Pedagogical alignment
E08	PhD	Assistant Professor	Education	Curriculum relevance

3.5 Expert-Rating Procedure

The 30 assessment items were distributed to all experts with the instrument framework, assessment criteria, and rating instructions. Experts rated each item independently on five criteria: relevance, clarity, grade-level appropriateness, critical thinking alignment, and curriculum alignment. Ratings were done on

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

a four-point scale: 1 = not appropriate, 2 = needs major revision, 3 = appropriate, and 4 = highly appropriate. Ratings of 3 or 4 were considered agreement for calculating CVI.

3.6 Content Validity Analysis

For item i, the I-CVI was calculated as the number of experts assigning a rating of 3 or 4 divided by the total number of experts:

I-CVI = n_a / N

N is the total number of experts, and n_a is the number of experts who rated the level as acceptable. If eight experts agree, the I-CVI is 1.00; if seven out of eight experts agree, the I-CVI is 0.875. The mean I-CVI across items was used to determine the S-CVI/Ave. Since the study compared five criteria separately, criterion-specific S-CVI/Ave values were calculated and then averaged to get the overall instrument value. Furthermore, the percentage of unanimous agreement for item-by-criterion evaluations was taken as a supplementary descriptive indicator. In the data, a single non-unanimous rating pattern was reported, so modified kappa was also computed following a standard procedure for computing the probability of chance agreement. This value was used descriptively, in addition to the CVI, and not as a replacement for the overall validity argument.

4. RESULTS

4.1 Expert Panel Characteristics

All eight experts completed the content validation evaluation. Because of the panel's multidisciplinary nature, the evaluation covered mathematical content, critical-thinking theory, language clarity, assessment quality, pedagogical appropriateness, and curriculum relevance. This breadth was significant because the CTSAI was intended to do both: to serve as a mathematics assessment tool and as an indicator of critical-thinking processes.

4.2 Curriculum Alignment

Items were constructed based on the Grade 8 mathematics curriculum documents that were identified in the larger research project. Each of the three selected domains included 10 items: Number Sequences and Patterns, Expansion and Factorization, and Linear Equations and Inequalities. This balanced distribution was not designed to focus on any one mathematical content area, but to ensure coverage across the chosen content areas.

4.3 Scale-Level Content Validity

Agreement among the expert ratings was extremely high for each of the five criteria assessed for this product. The S-CVI/Ave values for Relevance, grade-level appropriateness, critical-thinking alignment, and curriculum alignment were each 1.000. The S-CVI/Ave for clarity was 0.9958 because one expert rated one item below the agreement threshold. The mean of the five criterion level values was used to determine an overall S-CVI/Ave: 0.9992 (to three decimal places).

Table:4

Validation criterion	S-CVI/Ave	Interpretation
Relevance	1.000	Unanimous agreement
Clarity	0.9958	Very high agreement; one item required wording refinement
Grade-level appropriateness	1.000	Unanimous agreement
Critical-thinking alignment	1.000	Unanimous agreement
Curriculum alignment	1.000	Unanimous agreement

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

Overall instrument	0.9992 (~0.999)	Excellent initial content-validity evidence

4.4 Item-Level Content Validity

An I-CVI of 0.875 was achieved for the relevance criterion on twenty-nine items, and for the clarity criterion on Item 9. This is higher than the typical .78 value for a panel of this size and was left as is. Expert feedback suggested the issue was not the content of the mathematics problem, but the language used, or its lack of clarity. This resulted in a refinement of Item 9, but no changes to its intended construct.

Table:5

Item(s)	Issue	I-CVI	Decision
Items 1–8	No substantive concern	1.000	Retained
Item 9	Minor wording/clarity concern	0.875	Retained after wording refinement
Items 10–30	No substantive concern	1.000	Retained

4.5 Agreement Beyond CVI

The observed I-CVI for Item 9 was 0.875, with seven experts rating it as appropriate or highly appropriate. With the modified-kappa approach, the chance-corrected agreement was about 0.871, which is considered to be excellent. Similarly, items with full agreement resulted in a modified kappa of 1.00. The results confirm the conclusion that the observed agreement was not due to chance.

Table:6

Statistic	Value	Interpretation
Overall S-CVI/Ave	0.9992	Excellent initial content-validity evidence
Minimum I-CVI	0.875	Above commonly used acceptability threshold
Modified kappa for Item 9	0.871	Excellent chance-corrected agreement
Unanimous item-level ratings	29/30 for the reported criterion	Very high consensus

4.6 Final Instrument Framework

Experts then reviewed it, and items were slightly modified in wording; the final preliminary CTSAI consisted of 30 items. The instrument is designed to measure three domains of content in mathematics and five domains of critical thinking. The present study provides evidence based on test content and is not a final scoring model or psychometric structure.

5. DISCUSSION

5.1 Interpretation of the Content-Validity Evidence

The study's findings were preliminary and showed high relevance, understandability, developmental appropriateness, and critical-thinking processes and linkages to the selected mathematics content for grade 8. The overall S-CVI/Ave is 0.9992, suggesting a high level of consensus across the five criteria. Importantly, this result should be taken as a test of content-based evidence and not as evidence of the instrument being completely valid for all intended uses. (Schoenfeld, 2016). Partly, this high amount of agreement is probably due to the structured development process.

This instrument was not meant to be a general reasoning test because the items were developed simultaneously and aligned with a specific framework for critical thinking and selected mathematics domains. A panel of multiple disciplines then considered the same issues through the lens of each profession. The sole concern about clarity in Item 9 is methodological, and when quantitative indices are extremely high, expert comments can identify wording issues.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

5.2 Relationship to Previous Research

The results are congruent with methodological literature which highlights expert review as a critical preliminary phase of the assessment development process (Haynes et al., 1995; Polit et al., 2007; Yusoff, 2019). They also align with a broader principle of instrument development: content evidence should precede, rather than supplant, empirical evidence of reliability and internal structure (Boateng et al., 2018; DeVellis, 2021).

This mathematics-specific orientation also aligns with scholarship that focuses on mathematical competencies and reasoning (Niss & Højgaard, 2019). In this way, the CTSAI expands the scope of critical thinking assessment to situations where students need to reason with mathematics rather than show reasoning without mathematics.

5.3 Theoretical and Methodological Contribution

The study's major contribution is to make the construct of critical thinking, the mathematics curriculum, and the demands at the item level explicit. This alignment clarifies the intended interpretation of the instrument and provides a foundation for further psychometric testing. A second contribution is the multidisciplinary validation panel.

Experts in mathematics provided content and reasoning evidence; experts in psychology and education provided evidence of the construct and pedagogy; and an English education expert provided evidence of language clarity. This triangulation is especially useful in the context of school-based assessment, where a construct representation may be confounded by language and curriculum knowledge.

5.4 Educational Implications

The CTSAI serves as an instructional model for teachers to use in mathematics assessment to help students develop evidence of reasoning, comparison, justification, evaluation, and reflection, rather than just getting the right answer. The study recommends that higher-order thinking is not isolated from a discipline's assessment process but embedded in the discipline itself. The instrument serves as a foundational, structured prompt for researchers to explore instructional interventions such as cooperative learning after it has been validated with student data. (Johnson & Johnson, 2009).

6. STRENGTHS, LIMITATIONS, AND FUTURE VALIDATION

6.1 Strengths of the Instrument

The instrument was specifically targeted to a specific critical-thinking framework and Grade 8 mathematics content. The expert panel was diverse and included representatives from mathematics, psychology, education, language, and assessment. The validation process used both quantitative CVI indices and expert feedback. The study also specifies the exact decision rule used to categorize ratings as agreement and records the only item that needed refinement.

6.2 Limitations

The results are subject to some limitations. First, the study is based on test content only and therefore does not provide evidence of internal consistency or item difficulty, discrimination, factor structure, measurement invariance, or relations with external variables. Second, the eight-member expert panel limits the range of professional perspectives.

Third, because the present manuscript does not present student cognitive interviews or pilot data, it makes no claims about how grade 8 students actually interpret the items. Fourth, the instrument is mathematics-specific, so the scores from this instrument cannot be used as a general measure of critical thinking without further evidence.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

6.3 Future Validation

Future steps should include administration of the instrument to a sufficient number of Grade 8 students. The first step in a staged validation program should involve cognitive interviewing or a small-scale pilot test to uncover misunderstandings, response-process problems, and completion-time issues. Follow this with item-level analysis of difficulty, discrimination, distractor functioning (where applicable), and response distributions.

Evaluate reliability using an adequate set of reliability coefficients, not just one. If the structure is not well defined beforehand, perform exploratory factor analysis, then confirmatory factor analysis in an independent sample. The sample size should also allow testing measurement invariance across other substantively relevant groups. Lastly, correlations with external variables, such as mathematics outcomes and theoretically relevant mathematics results, should be explored to reinforce the general validity argument (El-Hamamsy et al., 2022).

7. CONCLUSION

This study created and tested preliminary content-validity evidence for a 30-item Critical Thinking Skills Assessment Instrument for use with Grade 8 mathematics students. The instrument was systematically mapped to three mathematics domains and five operational critical-thinking processes and independently evaluated by eight multidisciplinary experts. Initial evidence of relevance, clarity, developmental appropriateness, construct alignment, and curriculum alignment is good, with an S-CVI/Ave of 0.9992 and very high agreement on the items.

A minor wording change was made to item 9, which was retained, and its level of chance-corrected agreement also was good. The study does not equate high content-validity indices with total instrument validation; however, this was intentional. The CTSAI should be viewed as a content-validated assessment that needs to be investigated for psychometric properties before it is used to make strong claims about students' critical-thinking ability or the effects of instructional interventions. This conservative interpretation further strengthens the study's contribution by situating the results within a modern, evidence-based validity framework.

REFERENCES

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A. R. (2015). Design and implementation of a content validity study. BMC Nursing, 14, 1–6.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. Education in Medicine Journal, 11(2), 49–54.

Tian, L., Gai, L., Huo, Y., Wang, J., Dong, H., & Zhang, L. (2025). Theoretical framework and assessment questionnaire of critical thinking for Chinese middle school students in science classrooms. Science & Education, 34, 4373–4404. https://doi.org/10.1007/s11191-025-00623-6

Su, H. F., Ricci, F. A., & Mnatsakanian, M. (2016). Mathematical teaching strategies: Pathways to critical thinking and metacognition. International Journal of Research in Education and Science, 2(1), 190–200.

Shen, J., et al. (2024). K-12 students' higher-order thinking skills: Conceptualization, components, and evaluation indicators. Thinking Skills and Creativity, 52, 101551. https://doi.org/10.1016/j.tsc.2024.101551

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. Journal of Education, 196(2), 1–38.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing & Health, 30(4), 459–467.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Research in Nursing & Health, 29(5), 489–497.

Piyaporn, P., & Tuksino, P. (2025). The development of a critical thinking skills measurement tool in science and technology for Grade 6 students. Journal of Educational Measurement Mahasarakham University.

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. Educational Studies in Mathematics, 102, 9–28.

Mooij, T., F. J. J. P. & R. Designing and implementing a test for measuring critical thinking in primary school. Thinking Skills and Creativity, 20, 40–49.

Messick, S. (1995). Validity of psychological assessment. American Psychologist, 50(9), 741–749.

Liu, O. L. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. ETS Research Report Series, 2014(1), 1–22. https://doi.org/10.1002/ets2.12009

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. Educational Researcher, 38(5), 365–379.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment. Psychological Assessment, 7(3), 238–247.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment, 7(3), 238–247. https://doi.org/10.1037/1040-3590.7.3.238

Guo, S., Shi, L., & Zhai, X. (2024). Validating an instrument for teachers' acceptance of artificial intelligence in education. Education and Information Technologies.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. California Academic Press.

Fabio, R. A., Antonietti, A., Iannello, P., & Suriano, R. (2025). Development and psychometric properties of the critical thinking attitude scale in Italian college students. Frontiers in Psychology, 16, Article 1599920.

Fabio, R. A., & Patafi, M. (2026). Assessing critical thinking in young children: Development and validation of the Critical Thinking Assessment for Children. Europe's Journal of Psychology, 22(2).

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. Topoi, 37, 165–184.

El-Hamamsy, L., Zapata-Cáceres, M., Martín Barroso, E., Mondada, F., Dehler Zufferey, J., & Bruno, B. (2022). The competent Computational Thinking test (cCTt): Development and validation of an unplugged Computational Thinking test for upper primary school. Computers & Education, 193, 104662.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. Thinking Skills and Creativity, 12, 43–52.

Davies, M., & Barnett, R. (Eds.). (2015). The Palgrave handbook of critical thinking in higher education. Palgrave Macmillan. https://doi.org/10.1057/9781137378057

DeVellis, R. F. (2021). Scale development: Theory and applications (5th ed.). Sage Publications.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), Teaching thinking skills: Theory and practice (pp. 9–26). W. H. Freeman.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research. Frontiers in Public Health, 6, 149.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research. Frontiers in Public Health, 6, 149.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

LIBERAL JOURNAL OF LANGUAGE AND LITERATURE REVIEW
An International Double-Blind Peer-Reviewed Open Access Journal
ISSN (Online): 3006-3175 ISSN (Print): 3006-3167
Volume 5, Issue 4 June, 2026

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. Research in Social and Administrative Pharmacy, 15(2), 214–221.

AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. Review of Educational Research, 85(2), 275–314.