

Development and Validation of a Multiple Representation-Based Concept Test on Algebra Topics for Form Four Mathematics

Zarina Maarof; Raja Lailatul Zuraida Raja Maamor Shah

Graduate Faculty of Sciences and Mathematics Sultan Idris Education University

DOI: <https://dx.doi.org/10.47772/IJRISS.2026.1026EDU0346>

Received: 17 May 2026; Accepted: 22 May 2026; Published: 17 June 2026

ABSTRACT

This study developed and validated the Algebra Concept Understanding Test (ACUT), a multiple representation-based concept for Form Four mathematics, specifically targeting the conceptual understanding of Variables, Algebraic Formulas, and Algebraic Expressions. Grounded in the Revised Bloom's Taxonomy (2001) and aligned with a Table of Test Specifications (TTS), the ACUT was designed to assess algebraic comprehension across six distinct modalities of representation. Utilizing a quantitative survey approach, a pilot study ($n = 32$) and a main study ($n = 135$) were conducted among Form Four students in Selangor, Malaysia selected via a two-stage stratified cluster random sampling method. Psychometric evaluation using Winsteps software version 3.64.2 revealed high item reliability indices, yielding 0.88 in the pilot study and 0.98 in the main study. Item fit and construct validity were confirmed through point-measure correlation (PT-MEA CORR), Infit/Outfit MNSQ, and Z-STD values. Additionally, content validity was verified by three mathematics education experts, achieving a Content Validity Index (CVI) of 0.92. These findings demonstrate that the ACUT is a highly reliable and valid instrument for diagnosing student comprehension and systematically identifying conceptual weaknesses in algebra. Ultimately, this diagnostic tool enables targeted instructional interventions and offers broader implications for developing multi-representational assessments in other mathematical and scientific domains.

Keywords: Algebra Comprehension, Form Four Mathematics, Educational Assessment, Test Validation, Multiple Representations.

INTRODUCTION

Although mathematical problem-solving approaches have proven effective in fostering procedural fluency, they often fall short when students conceptualize mathematics as a rigid set of rules rather than a coherent, meaningful system of understanding (Tularam & Hassan, 2025; Erbilgin & Gningue, 2023; Luki Luqmanul Hakim, 2021; Flores et al., 2015). This rule-based perception limits cognitive engagement and restricts the ability to transfer knowledge across diverse mathematical contexts. In their study, Tularam and Hassan (2025) observed that many learners merely memorize mathematical symbols and mechanical steps without grasping their underlying definitions, resulting in superficial comprehension and persistent conceptual gaps. Such deficiencies severely hinder long-term retention and compromise readiness for advanced mathematical tracks (Erbilgin & Gningue, 2023; Gitonga, 2016). To dismantle these algorithmic limitations, recent literature underscores the necessity of diagnostic frameworks—such as the Rasch Measurement Model—and pedagogical scaffolding through six modalities of the multiple representations (textual, numerical, arithmetic, symbolic, pictorial, and verbal forms) to cultivate a more balanced, multidimensional conceptual mastery.

This cognitive disconnect becomes particularly acute within school algebra, a domain that presents unique pedagogical hurdles due to its heavy reliance on abstract symbolic language. When introduced in isolation without sufficient contextual scaffolding, this abstract symbolism frequently induces frustration and cognitive disengagement among students. Consequently, learners struggle to interpret mathematical syntax or translate real-world situations into coherent algebraic expressions, a foundational competency required for advanced algebraic thinking and systemic problem-solving (Amir et al., 2016; Dahiana et al., 2023; Panasuk & Beyranevand, 2010). Crucially, these conceptual hurdles are not confined to low-performing students; even high-

ability learners frequently misinterpret algebraic structures due to fragile conceptual foundations (Yahaya & Elanggovan, 2010; Dahiana et al., 2023). Over time, such pervasive deficiencies erode academic interest and compound learning difficulties (Ali & Aziz, 2008; Baco, 2022). To mitigate these persistent misinterpretations, researchers emphasize that students must be equipped to transition fluidly across non-symbolic domains, thereby establishing the critical need for a multiple representation-based assessment capable of assessing conceptual mastery beyond mere symbolic manipulation.

To mitigate these symbolic barriers and address underlying conceptual issues, contemporary mathematics education emphasizes the integration of six multiple representation modalities as a strategic pedagogical framework for deepening conceptual understanding (Arefaine et al., 2022). Within this framework, representations function not merely as tools for communication but as crucial cognitive scaffolds that help students bridge abstract mathematical ideas with concrete interpretations (Arefaine et al., 2022; Dahiana et al., 2023). By actively engaging with these diverse multi-representational formats, students are better equipped to visualize complex algebraic relationships, construct structural meaning, and develop cognitive flexibility in problem-solving (Arefaine et al., 2022; Yahaya & Elanggovan, 2010; Kusumaningsih et al., 2018). When systematically integrated into instruction and assessment, these multiple representations have been shown to support diverse learning preferences, reduce abstract frustrations, and significantly enhance students' overall mathematical reasoning skills.

Grounded in these instructional and assessment principles, the present study builds on this theoretical foundation by developing and validating the ACUT, an instrument explicitly designed to assess students' conceptual understanding of core algebraic domains, namely variables, algebraic expressions, and algebraic formulas. To ensure a comprehensive diagnostic capacity, the ACUT is constructed using the six modalities of the multiple representation framework and is structurally aligned with the Revised Bloom's Taxonomy (2001), ensuring a balanced coverage across various cognitive process levels. Furthermore, test items were systematically developed using a Table of Test Specifications (TTS) to align the content with expected learning outcomes and cognitive complexity.

To rigorously evaluate the quality of the items generated through this systematic development process, this study employs the Rasch Measurement Model, a probabilistic measurement approach situated within Item Response Theory (IRT). The Rasch model offers several distinct advantages over Classical Test Theory (CTT), most notably the capability to separate item difficulty from student ability, thereby providing sample-invariant measures (Avcu, 2025; Dahiana et al., 2023; Sumintono, 2017). Unlike CTT, which depends heavily on raw scores and is inherently sample-dependent, the Rasch model enables a more precise calibration of item functioning and student proficiency by examining specific fit statistics and reliability indices. This calibration makes it an ideal framework for validating diagnostic tools aimed at assessing latent traits, such as conceptual understanding in mathematics. Concurrently, the newly developed ACUT items were subjected to comprehensive expert reviews to establish qualitative face and content validity.

Ultimately, this research aims to provide a rigorously validated assessment tool capable of identifying specific student misconceptions in algebra, supporting data-driven instructional interventions, and contributing substantially to the broader literature on multi-representational assessment in mathematics education.

METHODOLOGY

Research Conceptual Framework

This study adopts the McIntire and Miller (2019) test construction model as a guiding framework to develop a valid and reliable Algebra Concept Understanding Test (ACUT). The methodology integrates document analysis and quantitative survey methods to ensure comprehensive content representation, construct validity, and item reliability. In the instrument development phase, constructs were identified and test items involving various algebraic representations were generated based on curriculum documents, textbooks, and past examination questions; furthermore, several critical elements were factored into the construction of the ACUT items, including an updated Table of Test Specifications (TTS) that incorporated the topics, cognitive levels, and

procedural knowledge from the Revised Bloom’s Taxonomy (2001). During the validation and refinement phase, a pilot study involving 32 students was conducted where the draft ACUT was administered, and item analysis was performed using the Rasch Measurement Model to determine item fit, reliability, and construct validity, alongside expert reviews to establish content and face validity.

Population and Sample

The target population for this investigation comprised 470 Form Four students from randomly selected secondary schools in Selangor, Malaysia. These students had completed foundational algebraic topics at the lower secondary level under the Standard Curriculum for Secondary Schools (KSSM). Selangor was strategically selected as the research locale due to its dense student demographic and diverse structural distribution of secondary schools, which provided a representative sampling frame for capturing varied algebraic performance. Based on the total population (N = 470), a sample size of 135 students was selected. This sample constitutes approximately 28.7% of the population, thereby significantly exceeding the minimum 10% sampling threshold recommended by Yount (1999) for populations ranging between 101 and 1,000 subjects (see Table 1). To ensure a robust and unbiased distribution, a two-stage stratified cluster random sampling method was deployed. In the first stage, cluster sampling was applied to identify the classroom pools representing three distinct academic streams: pure science, technical science, and social science. In the second stage, stratified random selection was performed within these student pools to guarantee proportional representation across three explicit mathematical achievement levels - high (Grades A and B), moderate (Grade C), and low (Grades D and E) - benchmarked against the students’ lower secondary school mathematics assessment results. This dual-layered sampling design minimized selection bias and controlled for confounding variables arising from academic streaming. Crucially, within the paradigm of IRT, a sample size of 135 participants is highly sufficient and optimal for execution under the Rasch Measurement Model. According to Linacre (1994), the minimum sample size required to maintain item calibration stability within ± 0.5 logits at a 95% confidence interval ranges between 30 and 100 respondents. This threshold applies regardless of whether the assessment targets high- or low-ability cohorts. By employing 135 respondents, this study surpasses the conventional Rasch requirement, ensuring that the calibrated item difficulty parameters remain highly stable, sample-invariant, and free from anomalous skewing. Consequently, the sample size provides adequate statistical power to execute reliable item-person calibrations, construct validity verifications, and fit statistical analyses (Infit/Outfit MNSQ) without risking overfitting or model instability.

Table 1: Sample Size Determination Table

Population Size	Sample Size
1 – 100	100 %
101 – 1 000	10 %
1 001 – 5 000	5%
5 001 – 10 000	3 %
10 001 – ∞	1%

(Source: Yount, 1999: p. 100)

Research Instrument

The research instrument utilized in this study is the ACUT, developed to assess students’ understanding of algebraic concepts, specifically within the three core algebraic domains. The instrument was constructed based on the KSSM Mathematics Curriculum Specifications to ensure curriculum relevance. The ACUT comprises six subjective, open-ended problem-solving items, with two items allocated to each targeted topic. With a testing

duration of one hour and forty-five minutes, the instrument provides a focused and structured means of evaluating conceptual mastery among Form Four students. Subjective items were specifically selected because they allow students to demonstrate the depth of their algebraic understanding (Sarimah Baco & Mohd Zaki Ishak, 2021). Through this test, students exhibit their conceptual understanding across six modalities of representation. Students are required to apply all six forms of representation within their answers, and the successful integration of these elements reflects comprehensive algebraic understanding.

Content validity was evaluated by three expert mathematics teachers (Creswell & Guetterman, 2024). This expert review process was essential to ensure that the developed items possessed high content validity and were capable of measuring their intended constructs (Shamsina, 2013). Consequently, refinements were made based on the experts' feedback, specifically regarding sentence structure, content clarity, the appropriateness of the marking scheme and representations, and the suitability of the test duration.

Research Procedure

This study followed a systematic, eight-stage procedure comprising instrument construction, expert validation, administrative approval, pilot testing, instructional delivery, main study implementation, data collection and scoring, and data analysis.

Stage 1: Instrument Construction

The ACUT was developed based on the TTS, incorporating cognitive levels, the three core algebraic domains, and six modalities of multiple representations. Items were adapted and modified from the KSSM textbooks, reference books, and past examination questions to ensure curriculum alignment and the integration of all six representations within each item.

Stage 2: Expert Validation:

The ACUT underwent rigorous expert review to establish content validity. Three mathematics experts with over 10 years of teaching experience evaluated the items using the Content Validity Index (CVI). Based on their feedback, refinements were made to language clarity, content accuracy, marking scheme alignment, representation integration, and time allocation.

Stage 3: Approval and Ethical Clearance

Prior to data collection, formal administrative approval and ethical clearance were obtained from the Ministry of Education Malaysia (MOE), the Selangor State Education Department (JPNS), and the administrators of the participating schools.

Stage 4: Pilot Study

A pilot study was conducted with 32 Form Four students from a selected secondary school in Selangor who had undergone classroom sessions on the three core algebraic domains. The primary objective of this phase was to evaluate item clarity, optimal test duration, and the operational feasibility of the multiple representation design. Student responses were scored using a structured rubric, and the preliminary dataset was prepared for subsequent psychometric calibration.

Stage 5: Instructional Delivery

For both the pilot and main studies, PdP (teaching and learning) sessions were conducted by the researcher over a total of six hours (two hours per topic). During these sessions, students were explicitly introduced to the six representations and instructed on how to apply them systematically in mathematical problem-solving.

Stage 6: Actual Study Implementation

Following the validation and refinement of the instrument, the main study was implemented with a separate sample of 135 Form Four students from randomly selected schools in Selangor. Adhering to the same structural protocol as the pilot phase, students attended the designated PdP sessions and subsequently completed the refined ACUT within one hour and forty-five minutes.

Stage 7: Data Collection and Scoring

Student test scripts were collected and assessed using a structured scoring rubric. Each representational component was evaluated separately to ensure consistency, objectivity, and high inter-rater reliability across scorers.

Stage 8: Data Analysis

Following the main study, all student scripts were systematically evaluated and scored. The definitive data set was then subjected to comprehensive psychometric validation to test the calibration and diagnostic capacity of the ACUT framework. The mathematical probability of the data functioning was modeled based on the fundamental Rasch formulation:

$$\ln\left(\frac{P_{ni}}{1 - P_{ni}}\right) = \beta_n - \delta_i$$

Note: Detailed parameters used to interpret this model are further elaborated in the Data Analysis section below.

$$P_{ni}(\theta = 1) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$

Where:

$P_{ni}(\theta=1)$ is the probability that student n answers item i correctly.

e is the base of natural logarithms (Euler's number ≈ 2.71828).

β_n is the algebraic ability of student n

δ_i is the difficulty level of item i

Data analysis

To evaluate the psychometric properties of the ACUT, quantitative data obtained from both the pilot and main studies were subjected to Rasch Measurement Analysis using Winsteps software version 3.64.2. Unlike CTT, the Rasch model applies a probabilistic approach to separate item difficulty from student ability, providing sample-invariant measures crucial for diagnostic instruments (Avcu, 2025; Sumintono, 2017). The analytical framework was executed in two primary phases: instrument calibration and cognitive diagnostics. First, instrument validity and reliability were established by evaluating multiple Rasch parameters. Internal consistency was benchmarked using Cronbach's alpha (α), while the reproducibility of the scores was determined through item and person reliability and separation indices (Bond & Fox, 2015). Construct validity and item unidimensionality were verified through Point-Measure Correlation (PT-MEA CORR) values, ensuring all items measured the single latent trait of algebraic conceptual understanding. Additionally, item fit was monitored using Infit and Outfit Mean Square (MNSQ) and standardized Z-scores (Z-STD) to detect any

anomalous or misfitting items that deviated from model expectations instrument (Adnan et al., 2019; Nur Afifah Athirah & Asbulah, 2022). Second, to diagnose the depth of students' conceptual understanding across the six representation modalities, a Guttman scalogram analysis was employed. The Guttman scalogram systematically ordered the 36 ACUT items along a hierarchy of difficulty (from easiest on the left to most difficult on the right) against student ability levels (Azrillah et al., 2017). This scalogram distribution was utilized to identify non-procedural response patterns and cognitive vulnerabilities. Specifically, the analysis distinguished between conceptual gaps and performance anomalies, where an incorrect response to an easy item by a high-ability student indicated carelessness, whereas a correct response to a high-difficulty item by a low-ability student implied random guessing. This dual-layered statistical approach ensured that the ACUT functioned effectively as a highly valid, curriculum-aligned diagnostic tool.

RESULTS AND DISCUSSION

The content validity of the ACUT was designed to measure Form Four students' conceptual understanding across six modalities of representation and was rigorously evaluated by subject matter experts. The overall CVI achieved a score of 0.92, significantly exceeding the minimum acceptable threshold of 0.80 (Davis, 1992). This high index indicates that the instrument's items are clear, highly relevant, and structurally sound in measuring the intended algebraic constructs.

To complement this qualitative expert endorsement, further empirical validation was conducted using the Rasch Measurement Model. The instrument demonstrated excellent internal consistency, yielding a Cronbach's alpha (α) value of 0.92. According to Bond and Fox (2015), a reliability coefficient within this range indicates a highly dependable instrument. As interpreted through the established criteria outlined in Table 2, an alpha score between 0.9 and 1.0 represents a very good and effective level of measurement consistency. Consequently, these robust psychometric indices confirm that the ACUT is a highly reliable and effectively calibrated tool, making it well-suited for subsequent empirical deployment in diagnosing algebraic comprehension.

Table 2: Interpretation of Cronbach Alpha-Score

Alpha-Cronbach Score	Reliability
0.9 – 1.0	Very good and effective level of consistency
0.7 – 0.8	Good and acceptable
0.6 – 0.7	Acceptable
<0.6	Item need to be repaired
<0.5	Items need to be dropped

The Rasch analysis revealed an outstanding item reliability of 0.98 and an item separation index of 6.86, indicating that the test possesses a robust range of item difficulties. Similarly, the person reliability was high at 0.91, coupled with a person separation index of 3.26, which proves the instrument's strong capability to accurately distinguish between different levels of student proficiency.

In terms of construct validity, the PT-MEA CORR for the main study items yielded entirely positive values, confirming that all items measure in the same direction and align perfectly with the target construct, namely the conceptual understanding of algebra. Table 3 below presents the ACUT item ordering based on PT-MEA CORR values.

Table 3: ACUT Item Ordering Based on PT-MEA CORR Values

Entry Number

	PT-MEA Corr.	Item	Entry Number	PT-MEA Corr.	Item
1	0.17	S1A	6	0.49	NU1F
8	0.34	A2B	31	0.52	G6A
2	0.37	A1B	34	0.53	N6D
12	0.38	L2F	18	0.54	L3F
5	0.39	L1E	25	0.55	G5A
7	0.41	S2A	14	0.56	S3B
13	0.41	G3A	36	0.57	L6F
26	0.42	L5B	15	0.57	A3C
17	0.43	N3E	30	0.58	NU5F
4	0.45	N1D	9	0.58	G2C
11	0.45	NU2E	33	0.58	S6C
21	0.45	G4C	23	0.58	L4E
20	0.46	A4B	32	0.60	A6B
10	0.47	N2D	28	0.60	S5D
16	0.47	NU3D	27	0.61	A5C
19	0.48	S4A	22	0.64	A4D
24	0.48	NU4F	35	0.67	NU6E
3	0.49	G1C	29	0.72	N5E

As shown in the statistical summary in Table 4, the mean infit MNSQ value of 0.98 falls within the acceptable range of 0.81 to 1.17. This infit range indicates an excellent item reliability fit, as it lies well within the standard parameter of 0.5 to 1.51, which is deemed highly productive and effective for psychometric assessment (Azrillah et al., 2017). These statistics confirm that the ACUT items fit the model's expectations psychometrically. Consequently, this demonstrates that the ACUT is fully capable of measuring the conceptual understanding of algebra through the systemic integration of all six modalities of multiple representations in students' responses (Azrillah et al., 2017). Beyond mere statistical compliance, the optimal Infit and Outfit MNSQ values presented in Table 4 carry profound diagnostic significance regarding the students' internal cognitive processes. Because these fit indices sit tightly within the acceptable thresholds, they empirically demonstrate that the students' response patterns were driven by actual latent algebraic abilities rather than external noise, systematic guessing, or item ambiguity. This high psychometrically alignment validates the authenticity of the cognitive data, confirming that the measurement of students' conceptual understanding is both stable and legitimate. This empirical fit justifies a deeper transition from static statistical ordering to a qualitative examination of the students' active cognitive structures.

Table 4: Statistical summary for the study

Statistic	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
Mean	0.98	0.0	0.98	0.1
S.D.	0.18	1.5	0.42	1.1
Max	1.41	4.1	2.38	2.1

Statistic	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
Min	0.66	-2.8	0.25	-2.8

The utilization of an open-ended response format allowed for an in-depth evaluation of students' cognitive processes, particularly their ability to explain, elaborate, and connect algebraic concepts. The analysis of student responses via the Guttman scalogram revealed that high-ability students successfully and consistently integrated multiple representations, smoothly transitioning between symbols, textual sentences, and pictorial diagrams to solve complex problems. This integration reflects a profound and balanced conceptual understanding of algebra.

Conversely, when students failed to connect these representations or utilized them in isolation, it exposed fundamental conceptual weaknesses rather than mere procedural errors. The data showed that students who traditionally relied solely on symbolic representations demonstrated significantly improved comprehension when provided the opportunity to interpret problems through alternative formats, such as diagrams, numerical representations, or oral explanations.

These findings strongly align with constructivist principles and Bruner's theory of representation. The integration of multiple representations was shown to reduce misconceptions and clarify complex concepts, specifically within the domains of the three core algebraic domains, while broadening students' problem-solving strategies. Ultimately, this study establishes that the ACUT is not only a valid and reliable assessment instrument but also a highly effective diagnostic tool for identifying students' conceptual strengths and vulnerabilities in mathematics.

CONCLUSION

In conclusion, this study successfully demonstrated that the ACUT meets its primary objectives: developing test items that incorporate multiple representations to assess Form Four students' understanding of algebra; evaluating the validity and reliability of the items; assessing item fit; and measuring students' algebraic comprehension using the Rasch Measurement Model. Learning becomes significantly more effective when students actively engage in the classroom through multiple representations, which fosters a deeper, more balanced mastery of the concepts being taught. Furthermore, the systematic integration of multiple representations has been shown to positively influence students' confidence in grasping abstract concepts.

The empirical findings yielded high reliability indices, indicating excellent internal consistency and stability over time. Because the ACUT thoroughly fulfilled the criteria of the Rasch Measurement Model, the instrument can be robustly administered to different student samples sharing similar demographic and academic characteristics. The ACUT also demonstrated strong content validity, as verified by mathematics experts, alongside robust construct validity, evidenced by entirely positive PT-MEA CORR values. This confirms that the ACUT is unidimensional and aligns perfectly with the target construct of measuring students' conceptual understanding of algebra. Overall, the ACUT exhibits exemplary reliability and validity, establishing that students' algebraic comprehension can be effectively assessed through six distinct modalities of representation.

To expand the research scope and enhance assessment methodologies in algebraic education, several recommendations are put forward for future research. First, future studies should involve larger, demographically diverse student populations to improve generalizability, while incorporating a mixed-methods design to enrich quantitative psychometric data with deep qualitative insights into student thinking. Second, researchers should utilize diversified diagnostic instruments, such as structured interviews and targeted diagnostic tests, to precisely map the specific cognitive processes, hurdles, and root causes of students' difficulties in learning algebra. Additionally, future research should evaluate the efficacy of innovative, technology-driven instructional modules designed to sustainably foster student interest, cognitive potential, and academic performance.

Moreover, comparative institutional studies should be conducted across different academic contexts—such as Residential Schools (Sekolah Berasrama Penuh) and MARA Science Colleges (MRSM)—to assess whether

high-ability students exhibit a correspondingly superior mastery of multiple representations. Furthermore, future investigations should explore teachers' baseline pedagogical content knowledge and readiness regarding multiple representations, examining their implementation challenges and proficiency in leveraging digital tools to construct high-quality, multi-representational assessment items.

Finally, studies should examine the cross-disciplinary impact of embedding multiple representation modalities in other fields, such as science education, to determine its broader efficacy in improving overall conceptual understanding.

REFERENCES

1. Amir, M., Ahmad, S., Kebangsaan, S., Baharu, P., Kebangsaan, S., & Geruntum, U. (2016). Analisis Kesalahan Newman Dalam Penyelesaian Masalah Kurikulum Untuk Mata Pelajaran Matematik Berubah Seiring Dengan Perkembangan Kurikulum Yang Berlaku. *Lampiran Peraturan Menteri Pendidikan Nasional*, 6(2), 69–84.
2. Arefaine, N., Michael, K., & Assefa, S. (2022). Effect of Multiple Representations on Students' Performance on Interpretations and Techniques of Representation in Calculus. *Mathematics Education Journal*, 16(3), 351–372. <https://doi.org/10.22342/jpm.16.3.18291.351-372>
3. Avcu, A. (2025). The Contribution of Rasch Modeling on Measuring Attitudes for Better Classroom Assessment. *International Journal of Psychology and Educational Studies*, 12(3), 169–179. <https://doi.org/10.52380/ijpes.2025.12.3.1432>
4. Azizi Yahaya, & Elanggovan AL M Savarimuthu. (2010). *Kepentingan Kefahaman Konsep Dalam Matematik*. 1989.
5. Azrillah Abdul Aziz, Mohd Saidfudin Masodi, & Azami Zaharim, ... (2017). *Asas Model Pengukuran Rasch* (Third edit). Penerbit UKM.
6. Bond Trevor, & Christine, M. F. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Third Edit). Routledge Taylor & Francis Group.
7. Creswell, J. W., & Guetterman, T. C. (2024). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (7th ed.). Pearson Education, Inc.
8. Dahiana, W. O., Herman, T., Nurlaelah, E., & Pereira, J. (2023). Student Semiotic Representation Skills in Solving Mathematics Problems. *Jurnal Didaktik Matematika*, 10(1), 34–47. <https://doi.org/10.24815/jdm.v10i1.30770>
9. Erbilgin, E., & Gningue, S. M. (2023). Using The Onto-Semiotic Approach to Analyze Novice Algebra Learners' Meaning-Making Processes with Different Representations. *Educational Studies in Mathematics*, 114(2), 337–357. <https://doi.org/10.1007/s10649-023-10247-8>
10. Flores, R., Koontz, E., Inan, F. A., & Alagic, M. (2015). Multiple Representation Instruction First Versus Traditional Algorithmic Instruction First: Impact in Middle School Mathematics Classrooms. *Educational Studies in Mathematics*, 89(2), 267–281. <https://doi.org/10.1007/s10649-015-9597-z>
11. Gitonga, I. (2016). Utilizing Semiotic Perspective to Investigate Algebra II Students' Exposure to and Use of Multiple Representations In Understanding Algebraic Concepts. Georgia State University.
12. Kusumaningsih, W., Darhim, Herman, T., & Turmudi. (2018). Improvement algebraic thinking ability using multiple representation strategy on realistic mathematics education. *Journal on Mathematics Education*, 9(2), 281–290. <https://doi.org/10.22342/jme.9.2.5404.281-290>
13. Linacre, J. M. (1994). Sample Size and Item Calibration Stability. *Rasch Measurement Transactions*, 7(4), 328.
14. Luki Luqmanul Hakim, B. Y. (2021). Conceptual and Procedural Knowledge in Mathematics Education. *Design Engineering*, 9/pages:1271-1280, 1–10. <https://doi.org/10.4324/9780203063538>
15. Marlina Ali, & Aziz, N. H. A. (2008). Tahap Kefahaman Dalam Tajuk Ungkapan Algebra Pelajar Tingkatan Dua.
16. Miller, L.A., Lovler, R. L., & McIntire, S. A. (2019). *Foundations of Psychological Testing: A Practical Approach* (6th edition). In *Personnel Psychology* (Sixth edit, Vol. 60, Issue 3). Thousand Oak: SAGE Publications Inc. https://doi.org/10.1111/j.1744-6570.2007.00091_11.x
17. Nur Afiqah Athirah Mohd Rushdi, & Asbulah, L. H. (2022). Kesahan Dan Kebolehpercayaan Instrumen

- Sikap Pelajar Terhadap Pembelajaran Bahasa Arab (I-SPPBA) Menggunakan Model Pengukuran Rasch. 98–110.
18. Panasuk, R. M., & Beyranvand, M. L. (2010). Algebra Students' Ability to Recognize Multiple Representations and Achievement. *International Journal for Mathematics Teaching and Learning*.
 19. Sarimah Baco. (2022). Domain Kognitif dan Pencapaian Ungkapan Algebra dalam kalangan Pelajar Tingkatan Dua. *Borneo International Journal of Education (BIJE)*, 2, 55–68. <https://doi.org/10.51200/bije.v2i.4112>
 20. Sarimah Baco, & Mohd Zaki Ishak. (2021). Kesahan Dan Kebolehpercayaan Instrumen Penilaian Kendiri Pembelajaran Ungkapan Algebra Tingkatan Dua. [The Validity and Reliability of The Form Two Algebraic Expression Self- Assessment Learning Instrument]. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 6(11), 127–137.
 21. Shamsina Shamsuddin. (2013). Pengubalan Item Ujian (First edit). Freemind Horizons Sdn. Bhd.
 22. Sumintono, B. (2017). Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (assessment for learning). 1–19.
 23. Sumintono Bambang. (2017). Rasch Model Measurement as Tools in Assessment for Learnig. *International Conference on Education Innovation*.
 24. Tularam, G. A., & Hassan, O. M. (2025). Persistent Misconceptions in Algebra: A Critical Analysis of Errors with Implications for Teaching and Further Research. *Journal of Social Sciences*, 21(1), 38–50. <https://doi.org/10.3844/jssp.2025.38.50>