

Development and Validation of Mathematical Thinking and Hots Assessment Instruments for Secondary Mathematics

Leong Zi Jun, Norulhuda Ismail

Faculty of Educational Sciences and Technology, Universiti Teknologi Malaysia
(UTM), Malaysia

Email: zi.jun.leong@graduate.utm.my, p-norulhuda@utm.my

DOI Link: <http://dx.doi.org/10.6007/IJARPED/v15-i3/28939>

Published Online: 30 September 2026

Abstract

The Malaysia Education Blueprint 2013-2025 requires assessment of Higher-Order Thinking Skills (HOTS). The objective of this study is to develop and validate an assessment instrument measuring both HOTS and Mathematical Thinking (MT) skills in Malaysian Form 4 students studying quadratic functions, according to KSSM standards. The development of the assessment instrument began with (1) theoretical alignment between Bloom's Taxonomy and Mason's MT; (2) development of items based on KSSM content standards; (3) expert validation by five Malaysian mathematics education specialists using the Content Validity Index (CVI); and (4) pilot testing with 20 Form 4 students from Johor schools. Expert validation yielded excellent content validity with I-CVI scores ranging from 0.80 to 1.00. Pilot testing demonstrated high internal consistency, with Cronbach's alpha values of 0.89 for HOTS and 0.87 for MT. The final instrument comprises three parts: Calculator Proficiency (5 items), HOTS Assessment (4 items), and MT Assessment (4 items), with scoring rubrics aligned to KSSM standards. This study presents the first systematically validated instrument integrating Bloom's Taxonomy and Mason's MT framework for Malaysian Form 4 students studying quadratic functions. The validated tool equips teachers with reliable resources to assess advanced thinking, supports evidence-based instructional decisions, and contributes to national educational initiatives.

Keywords: Assessment Instrument, Content Validity, Higher-Order Thinking Skills (Hots), Mathematical Thinking, Reliability, Secondary Mathematics

Introduction

The Malaysian Education Blueprint 2013-2025 establishes Higher-Order Thinking Skills (HOTS) as the main educational reform element which the Kurikulum Standard Sekolah Menengah (KSSM) mathematics curriculum (Ministry of Education Malaysia, 2013) must implement. Students need HOTS because these cognitive processes that combine analysis with evaluation and creation (Anderson & Krathwohl, 2001) help them develop skills to solve mathematical problems that require advanced thinking. The cognitive framework of Mathematical Thinking (MT) establishes the processes which students use to develop and demonstrate HOTS in

mathematics education through their activities of specialising, generalising, conjecturing, and convincing (Mason, Burton, & Stacey, 2010).

Malaysian students still perform poorly on international assessments which test these skills because they are not able to meet minimum proficiency standards in mathematical reasoning, according to PISA 2022 results which showed only 41% of Malaysian students achieved this level (OECD, 2023). The performance gap demonstrates the requirement for precise assessment instruments which can evaluate students' HOTS and MT abilities while providing educational guidance and tracking their progress toward national academic objectives.

The present issue in Malaysian mathematics education is the insufficient availability of diagnostic assessment instrument that match specific curriculum requirements and have received local validation to measure students' higher-order thinking skills and mathematical thinking abilities. The Sijil Pelajaran Malaysia (SPM) examinations include HOTS-oriented items, but these function as summative assessments that do not enable teachers to use student performance data for ongoing cognitive assessment (Zulnaidi et al., 2020). Existing classroom instruments generally focus on teaching students how to perform mathematical procedures, but they do not teach students how to think and understand mathematical concepts (Mulyati et al., 2023). The educational gap for students who study quadratic functions as their foundational KSSM Form 4 topic is especially pronounced, with high rates of student struggle in solving non-routine problems and transferring their knowledge to other subjects.

The researchers created assessment instrument to measure HOTS and MT in Malaysian secondary mathematics education through their research work. The instruments are intended to: (1) align explicitly with KSSM content and performance standards, (2) measure cognitive processes defined by established educational frameworks, (3) offer reliable diagnostic information to guide instructional decision-making, and (4) support local research on interventions aimed at developing HOTS in Malaysian classrooms. The research study provides teachers with validated assessment instrument which enable them to achieve national educational goals established in the Malaysia Education Blueprint.

Theoretical Framework and Construct Definition

Higher-Order Thinking Skills (HOTS)

This research study establishes Higher-Order Thinking Skills through its implementation of the revised Bloom's Taxonomy, which Anderson and Krathwohl developed in 2001. The taxonomy introduces major enhancements to the original Bloom's Taxonomy by incorporating current cognitive psychology research and addressing educational application issues that emerged throughout its more than thirty years of use. The revision transforms the original noun-based categories into verb-based cognitive processes, demonstrating that people actively engage in the process of thinking. According to Anderson and Krathwohl (2001), this new structure helps educators better understand learning objectives and offers improved guidance for developing instructional materials and assessment instrument.

The taxonomy consists of six cognitive process dimensions, which Malaysian educational policy defines as HOTS through its three highest levels. The first level (Remembering) requires individuals to retrieve knowledge stored in their long-term memory. The second level

(Understanding) requires individuals to create meaning from the educational materials they receive. Students must acquire fundamental skills that lead to advanced thinking abilities. The HOTS construct properly begins with the third level and extends through the sixth. Loh et al. (2017) observed that Malaysian students face challenges in obtaining valid and reliable assessments of their HOTS abilities, which the Ministry of Education has designated for national examinations.

Students need to use their acquired knowledge to complete tasks by performing established procedures under designated circumstances. The application of knowledge requires students to use the quadratic formula for solving new equations and graphing quadratic functions through learned procedures. The first level of HOTS assessment begins with applying, where students must choose appropriate methods and adapt them based on specific environmental needs, because this task demands higher cognitive abilities than simple memory tasks. Research by Abu Kasim (2025) found that non-routine problems constitute only 2.18% of total textbook questions in Malaysian upper secondary mathematics, which shows that educational materials place limited emphasis on tasks that require cognitive demand and open-ended solution approaches. This lack of representation creates difficulties for students because it hinders their ability to develop and showcase their HOTS skills.

Learners need to break down content into fundamental elements, which they need to evaluate how these elements connect to the entire content system. This cognitive process requires three separate cognitive tasks: identifying essential information while discarding nonessential information, arranging items into organized patterns, and determining the primary function or viewpoint of the material. Students must examine how various algebraic expressions affect the graphical features of quadratic functions through coefficient changes, which show how these changes impact parabola paths and structural design, and they need to determine which key components of a quadratic function will assist them in finishing their specific task. Iskandar Shah et al. (2024) showed that students find it hard to perform analytical tasks which require them to understand inverse functions and composite functions, proving the need to assess their analytical skills through specific tests.

Students need to develop their judgment skills by using established assessment criteria to complete performance evaluations. The system requires both assessment of its internal component relationships and assessment of its components through external measurement standards. Students who evaluate quadratic functions need to assess mathematical correctness standards for their proposed solutions, determine the most efficient method to accomplish particular assignments, and assess their peers' explanations of quadratic relationships. The evaluation process requires students to assess both their responses and the mathematical work they produce. Che Mansor et al. (2025) discovered that students who used mathematical software achieved 22% better comprehension results because their cognitive assessment skills improved through proper integration of technological tools into learning activities.

Students need to create complete operational systems by combining different parts until they find the correct pattern and sequence. The highest cognitive level enables people to create new hypotheses, develop procedural plans, and make original creations. Students who study quadratic functions through creating will develop new problem-solving methods, which they

will use to design their own quadratic investigations and create mathematical proofs that connect various concepts. Creating serves as the highest level of HOTS because it allows students to demonstrate their mathematical thinking through creative mathematical expressions. Loh et al. (2017) stated that creativity is one of three main components which constitute HOTS, together with critical thinking and problem-solving abilities, and it should be assessed through suitable evaluation tools.

The KSSM curriculum establishes specific cognitive requirements which match these taxonomy levels to determine the required cognitive skills that Malaysian mathematics students must exhibit at each educational level, according to the Ministry of Education Malaysia (2016) document. Students in Form 4, who are between middle and advanced secondary education, must demonstrate their ability to analyze and evaluate while using their basic knowledge, but some students can achieve creating-level thinking. The assessment instruments which researchers created in this study provide tested items that match each expected cognitive level. The research conducted by Embong et al. (2025) showed that visual teaching methods in mathematics lead to better student outcomes because assessment methods should match instructional strategies in order to build and evaluate higher-order thinking skills. The current study's instruments contribute to this alignment by providing validated measures that reflect both theoretical frameworks and curriculum expectations.

Mathematical Thinking (MT)

The framework established by Mason et al. (2010) demonstrates mathematical thinking through four basic processes which function as essential elements of mathematical work. The framework developed from researchers who investigated the real thought processes of mathematicians and their research on how students acquire mathematical knowledge. According to Mason and his team in their pioneering research, mathematical thinking represents a way of thinking about mathematics which all people can acquire through practice. According to Zeynivandnezhad et al. (2013), mathematical thinking functions as a process which helps students build more advanced mental frameworks while it plays an essential role in developing their understanding of concepts.

Mason's framework differs from taxonomies that define cognitive levels because it explains cognitive processes that represent the active mechanisms through which people develop mathematical knowledge. The four processes, which include specialising, generalising, conjecturing, and convincing, function as interconnected elements that interact with each other through mathematical reasoning which requires continuous practice. The research conducted by Faizah et al. (2020) showed that Mason's framework enables effective analysis of students' thinking patterns during mathematical proof because it demonstrates how students create different warrant types throughout their thinking process.

Professionals develop their intuitive field expertise through the study of particular instances which form the foundation for their specialized training. The process starts with general principles which result in the development of particular examples that serve as proof for testing theoretical concepts, while users learn about mathematical entities through their active engagement. Students who specialize in quadratic functions will examine specific quadratic equations while they calculate particular values, graph individual functions, and investigate actual quadratic relationships that exist in the world. Zeynivandnezhad et al.

(2013) discovered that pre-service teachers predominantly demonstrated mathematical thinking through specialized mathematical processes, which indicates that students tend to begin their mathematical studies with direct interaction of particular instances. The procedure of specialization creates basic experiential evidence which enables mathematicians to build complex mathematical arguments.

Researchers establish general principles through their study of multiple case studies which lead to the discovery of patterns and relationships between different structures. The process begins with specific cases which help researchers establish universal laws that lead to new scientific principles. Students who study quadratic functions will discover through their generalizations that all parabolas open upward when the coefficient of x^2 is positive, that the vertex lies on the axis of symmetry, and that quadratic equations produce two solutions which may be either real or complex. This fundamental mathematical process, which mathematicians use to identify patterns among various mathematical elements, requires generalising as its core element. Ibrahim et al. (2023) stressed that students need to develop their ability to generalize because it stands as a vital component of mathematical thought process, yet their study showed that Malaysian Form 4 students faced major difficulties with this skill.

Scientists generate new predictions and hypotheses through the process of conjecturing, which uses their current knowledge and the identification of new patterns. The process requires mathematicians to develop tentative claims about mathematical relationships, propose generalization possibilities, and describe their expectations about mathematical outcomes. Conjecturing exists as a creative process that develops new mathematical ideas through its exploratory nature. Students create conjectures about how quadratic functions operate by establishing links between coefficient values and root properties, estimating graph behavior based on parameter alterations, and developing predictions about how quadratic functions link with other mathematical fields. Faizah et al. (2020) discovered that students who observe mathematical situations use conjectures to make predictions about the future, but they cannot create valid conjectures because they need scaffolding support. Ibrahim et al. (2023) found that Malaysian Form 4 students found conjecturing to be their hardest mathematical thinking skill, which means that mathematics teachers must give this skill special attention.

Scientists need to use mathematical proof as the fundamental evidence for their scientific claims, which they need to establish through proof. The process consists of two components, which include informal justification and formal proof, because the first method shows that something is logical, whereas the second method establishes logical requirements. The process of convincing enables mathematicians to demonstrate their knowledge because it allows them to present logical evidence that transforms conjectured ideas into established truths. Students who study quadratic functions must demonstrate their understanding of solution methods through three specific tasks which include explaining effectiveness, proving results about quadratic relationships, and establishing properties of quadratics. Zeynivandnezhad et al. (2013) stated that convincing represents the peak of mathematical thought because students must create arguments which follow logical principles and rely on proof. Faizah et al. (2020) explained that students who reach the convincing stage establish

different types of warrants, which include institutional, evaluative, a priori, and empirical warrants, to validate their mathematical claims.

The four processes operate as interdependent components because their mathematical connections require constant practice to function. Students might specialize to develop intuitions, generalize from those specializations, conjecture based on generalizations, and convince through justification, but they might also cycle back through these processes repeatedly as understanding deepens. The framework Mason established shows that people think mathematically through non-linear processes which develop through organic yet structured systems. The expanded second edition of *Thinking Mathematically* explains how these processes connect to fundamental powers which every learner possesses because they already demonstrate their ability to speak, which shows that mathematical thinking builds on essential human cognitive abilities.

The developmental stages of processes reach their highest point when they develop cognitive skills which secondary students possess at that moment. Piaget's formal operational thinking pattern, which he established in 1952, explains the way people obtain their cognitive abilities after they finish elementary education. The developmental stage of children at this age period matches the Form 4 level of Malaysian secondary education. Ibrahim et al. (2023) conducted research with Malaysian Form 4 students which showed that students at this level should possess all four mathematical thinking skills, yet their study results showed that students lacked these abilities, with conjecturing being the most difficult for them. Abu Kasim (2025) discovered that Malaysian mathematics textbooks dedicate only 2.18% of their questions to non-routine problems, which limits students' development of mathematical thinking abilities. The study developed assessment instruments which measure each process through specific items that assess every mathematical thinking type to resolve the deficiencies in the Malaysian mathematics education system.

Calculator Proficiency

The study defines calculator proficiency as the ability to operate scientific calculator functions needed for solving quadratic equations, together with the knowledge of appropriate times to use calculators as mental aids instead of basic calculation instruments (Drijvers, 2015). The construct requires students to acquire skills in operating technological tools which serve as essential elements in contemporary mathematics instruction. The study by Abdul Rahman et al. (2022) demonstrates that scientific calculators have functioned as educational instruments in Malaysian secondary schools and universities since their implementation to support students' mathematical learning through concept understanding and computational problem solving. Teachers and students need to develop scientific calculator skills because this technology requires proficiency for effective educational use, which enables learners to master mathematics through enhanced ability and confidence.

Calculator proficiency consists of two main aspects: technical skills and strategic abilities. The first element of technical proficiency requires users to identify all calculator functions and learn methods to operate these functions and interpret their results. The process of solving quadratic functions requires calculator functions that enable users to solve equations, graph parabolas, determine vertices and intercepts, and validate their algebraic calculations. Research by Andu et al. (2025) demonstrated that scientific calculators enable students to

visualize, compute, and explore basic and multiple equations as well as formulas in a user-friendly manner, with their study of 125 students revealing significant improvements in mathematical fluency following structured calculator guidance. Strategic proficiency involves knowing when to use calculator tools and when to rely on mental or paper-and-pencil methods, recognizing that calculators can support but not replace mathematical thinking. Zamalik et al. (2024) found that secondary school students face difficulties in using scientific calculators, which are widely available, because they need structured training programs to help them use calculators effectively and create better learning results.

The assessment requires students to demonstrate their calculator skills because Malaysian classrooms use scientific calculators as standard equipment which students can bring to most tests. Abdul Rahman et al. (2022) confirmed that scientific calculators have been implemented from early secondary years in Malaysia, yet with many topics to cover during mathematics lessons, the use of scientific calculators often receives insufficient instructional attention, leaving students to develop proficiency independently. This study treats calculator use as a legitimate mathematical skill which should be evaluated together with other mathematical skills because it does not harm mathematical comprehension. Kissane (2016) argued that scientific calculators have a special place in school mathematics because they function as accessible technology which supports mathematical learning when teachers use them for their intended purpose. Jun (2025) studied how calculator competency relates to Higher-Order Thinking Skills in secondary students who use calculators to improve their mathematical abilities, but their calculator access fails to help them solve HOTS problems, which requires strategic proficiency.

Construct Integration: A Unified Framework

The three constructs (HOTS, MT, and Calculator Proficiency) interact with each other through complex mathematical activities, which creates their functioning. HOTS describes the cognitive levels at which students operate, MT describes the cognitive processes they employ, and Calculator Proficiency describes the technological tools they deploy. Together, these constructs provide a complete assessment system which teachers use to measure students' mathematical abilities. Recent research by Jun (2025) empirically demonstrated this interconnection, finding a significant positive correlation between students' proficiency in using calculators and their performance in HOTS problems, with students possessing stronger calculator skills achieving higher scores in HOTS problem-solving. This finding establishes that calculator proficiency exists as a fundamental skill which supports students to develop advanced mathematical reasoning abilities through appropriate training.

The assessment instruments developed in this study treat these constructs as complementary rather than competing, recognizing that genuine mathematical thinking involves multiple dimensions simultaneously. A student analyzing a quadratic function (HOTS) might specialize by examining specific cases (MT) while using calculator tools strategically (Calculator Proficiency). The instruments capture these interactions through integrated items that require students to draw on multiple capabilities. Research by Nanmumpuni et al. (2022) supports this integrated approach, demonstrating that calculator-assisted guided discovery learning is effective in terms of developing Higher Order Thinking Skills in mathematics, with the integration of technological tools and pedagogical approaches working synergistically to enhance cognitive outcomes. The study discovered that integrated instructional methods

produced results which matched or exceeded the effectiveness of standard teaching methods because combined learning resources and cognitive frameworks delivered superior results compared to using either resource or framework alone.

Zulkifli et al. (2025) confirmed that students who use calculators with specific purposes show increased motivation, self-confidence, and enhanced problem-solving abilities, while their anxiety about algebra decreases and their academic work improves because they use calculators for teaching purposes instead of simple computation tasks. The study demonstrates that students who understand how to use calculators for their studies will achieve transformative results because their technological skills combine with their mathematical abilities. Abdul Rahman et al. (2022) documented that Malaysian students start using calculators during their secondary education, yet they learn to use scientific calculators through independent practice because teachers fail to provide adequate guidance on this topic. The research demonstrates that properly designed assessments which test all three abilities enable students to showcase their skills through genuine mathematical tasks.

Zamalik et al. (2024) demonstrated that structured interventions which taught students how to use calculators resulted in major academic improvements for students who studied quadratic equations, trigonometry, and number bases, because systematic calculator training enabled students to improve their overall mathematical skills. The research demonstrated that students who learned complete calculator skills achieved better results in solving mathematical problems across different math fields, which showed how technological skills connect with mathematical reasoning ability. The study showed that students preferred using advanced calculator models because they understood how these tools could help them with their mathematics work when used correctly.

The theoretical integration of these constructs aligns with broader frameworks in mathematics education research. Kissane (2016) defined scientific calculators as essential educational tools for school mathematics because they enable students to learn mathematics through their proper use. Radzuan et al. (2021) showed that scientific calculators help low-achieving students improve their mathematics performance through problem-solving instruction, which demonstrates that technological tools enhance mathematical thinking for students who find traditional methods difficult to understand. The assessment instruments developed in this study operationalize these theoretical integrations, providing a model for assessing HOTS, MT, and Calculator Proficiency through their authentic connections to mathematical practice.

Constructive Alignment with Curriculum Standards

The KSSM Form 4 mathematics curriculum requires students to demonstrate specific mathematical skills through their ability to solve quadratic equations, create graphical representations, understand how algebraic expressions connect with graphical displays, and apply quadratic functions to solve practical problems (Ministry of Education Malaysia, 2016). The performance standards establish the expected quality of student work, which should be demonstrated at various achievement levels. The performance standards from the Dokumen Standard Kurikulum dan Pentaksiran (DSKP) create six distinct bands which teachers can use to determine student mastery levels and select suitable teaching methods.

The assessment instruments measure the curriculum requirements which they need to measure because the standards require direct measurement of their content. The alignment functions at three different levels which include content coverage that ensures all curriculum topics are represented, cognitive demand which requires item difficulty to match curriculum expectations, and performance criteria which require scoring rubrics to match curriculum standards. Research by Sures Kumar and Foo (2024) demonstrated the importance of such alignment, showing that educational materials developed in accordance with KSSM standards and validated through the Content Validity Index achieved perfect CVI scores of 1.00, confirming their appropriateness for supporting the learning of KSSM Mathematics topics. Their study with Form 4 students showed that curriculum-aligned materials successfully met educational standards while they also improved active learning and student understanding.

The assessment system functions as a curriculum objective achievement tool because it serves as a fundamental part of the educational system. Abdul Wahid et al. (2025) emphasized that ensuring curriculum alignment is critical for educational assessment instrument, particularly when measuring complex cognitive skills, as alignment guarantees that what is being assessed corresponds precisely to what the curriculum intends students to learn. The research on automated question generation for Malaysian secondary mathematics showed that assessment items which developers based on official curriculum documents, teacher-prepared notes, and yearly teaching plans (RPT) generated questions with better curriculum alignment and factual validity. The DSKP framework allows assessment design to include performance standards which help teachers assess student performance for effective intervention and support their learning needs. The UTM NewsHub (2023) reported that schools need to adopt curriculum-aligned assessment methods which focus on mathematical thinking and problem-solving because these methods prepare students for both KSSM curriculum requirements and international assessment standards like PISA and TIMSS.

Methodology

Instrument Development Process

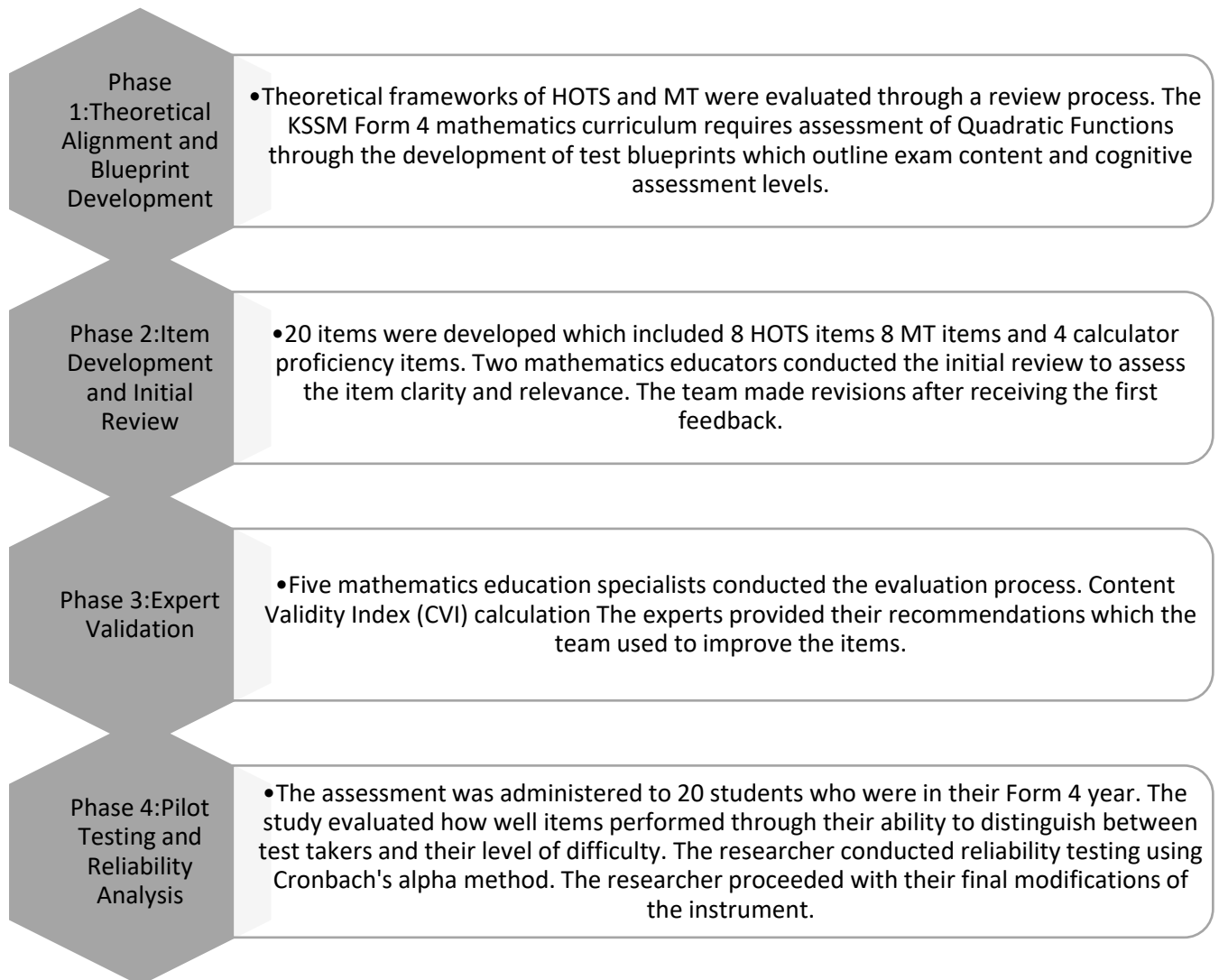


Figure 1 Instrument Development Process

Expert Validation Panel

The content validation process utilized the expertise of five Malaysian experts specialising in mathematics education and curriculum implementation. The panel was intentionally created to encompass different viewpoints present within the Malaysian educational system.

The panel consisted of:

1. A Senior Lecturer in Mathematics Education at Universiti Teknologi Malaysia with 10 years of experience in teacher training and HOTS research.
2. A Mathematics Teacher from SMK with 23 years of experience, currently involved in state-level KSSM implementation activities.
3. A Mathematics Teacher at SMK, who has 15 years of teaching experience, currently works as a national trainer for HOTS teaching methods.
4. A Mathematics Teacher at SMK has 12 years of teaching experience and teaches students how to use technology and calculators in mathematics classes.

5. A Mathematics Teacher at SMK has 9 years of teaching experience and creates assessment materials together with assessment standards for his students.

All experts are employed within the Malaysian public education system and used their classroom experience to validate assessment methods according to educational standards. The team developed assessment items that met cultural and language needs through their combined expertise, which included educational objectives that followed KSSM standards and addressed the unique needs of Malaysian students.

Expert Validation Instrument

Experts assessed each testing component through a structured validity evaluation which covered six dimensions of assessment. The first dimension assessed Content and Accuracy through its evaluation of factual correctness and depth of material. The second dimension assessed Clarity and Organization through its evaluation of language use and logical progression. The third dimension assessed Relevance and Alignment through its evaluation of curriculum components which matched educational objectives. The fourth dimension assessed Usefulness and Effectiveness through its evaluation of actual student engagement and specific educational outcomes. The first research instrument required validation before the pre-test assessment, and the second research instrument required validation before the post-test assessment. The assessment included four to five items which experts rated on a five-point Likert scale that ranged from Strongly Disagree (1) to Strongly Agree (5). Experts used open-ended questions to provide qualitative comments, which they used to identify strengths, suggest improvements, and offer detailed feedback about the test items.

The pilot testing involved 20 Form 4 students from secondary schools in Johor, Malaysia. Participants were selected using convenience sampling from schools that agreed to participate in the study. Ethical approval was obtained from the relevant authorities, and informed consent was secured from all participants and their parents or guardians. The sample comprised students from mixed ability levels to ensure adequate variation in performance for item analysis.

Pilot Testing Participant

The researcher calculated content validity through the Item-level Content Validity Index (I-CVI) and Scale-level Content Validity Index (S-CVI) according to the guidelines established by Polit and Beck (2006). The researchers measured internal consistency using Cronbach's alpha coefficient. During the analysis process, the researcher calculated the item difficulty index (p-value) and discrimination index (D) for each item.

Results

Data Analysis

Table 1

Instrument Structure and Specifications

Section	Items	Construct Measured	Maximum Marks	Time Allocation
A: Demographics	5	Background information	-	5 minutes
B: Calculator Proficiency	5	Technical and strategic calculator use	10	15 minutes
C: HOTS Assessment	4	Applying, Analysing, Evaluating, Creating	20	20 minutes
D: MT Assessment	4	Specialising, Generalising, Conjecturing, Convincing	20	20 minutes
TOTAL	13		50	60 minutes

The final instrument achieves complete construct coverage through its extensive measurement system while remaining suitable for practical use. The assessment requires 60 minutes, which teachers can use to conduct testing within their regular class schedule. The total of 50 marks provides sufficient score distribution for meaningful differentiation while remaining manageable for scoring.

The section includes test items which present different difficulty levels that enable students to demonstrate their fundamental abilities while challenging advanced students. The instrument design starts with easier tasks which lead to more difficult tasks to help students build their confidence before they face the most challenging HOTS and MT items.

Content Validity Results

Table 2

Expert Validation Results (Content Validity Index)

Item No.	Construct	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	I-CVI
CP1	Calculator Proficiency	5	5	4	5	5	1.00
CP2	Calculator Proficiency	5	4	4	5	4	1.00
CP3	Calculator Proficiency	5	5	5	4	5	1.00
CP4	Calculator Proficiency	4	4	5	4	4	1.00
CP5	Calculator Proficiency	5	5	4	5	5	1.00
H1	HOTS (Applying)	5	5	5	5	4	1.00
H2	HOTS (Analysing)	5	4	5	5	5	1.00
H3	HOTS (Evaluating)	4	5	4	5	4	1.00
H4	HOTS (Creating)	5	5	5	5	5	1.00
MT1	MT (Specialising)	5	5	4	5	4	1.00
MT2	MT (Generalising)	5	4	5	4	5	1.00
MT3	MT (Conjecturing)	4	5	4	5	5	1.00
MT4	MT (Convincing)	5	5	5	5	5	1.00

The Scale-level Content Validity Index (S-CVI) shows a value of 0.98, while the Average I-CVI shows a value of 0.99. All items received I-CVI scores of 1.00, which indicates that all experts reached complete agreement about the items' relevance and suitability. These results meet

excellent standards according to Polit and Beck (2006), as I-CVI scores achieved the required threshold of 0.78 with five expert evaluations. The development process, together with the expert feedback system, produced exceptional validity results which experts used to refine their items. The instruments demonstrate successful operationalization of the targeted constructs through complete agreement with expert assessment, which mathematics education professionals find valid.

Qualitative Expert Feedback Analysis

The expert analysis of comments resulted in three thematic outcomes which experts used to evaluate instrument quality while creating suggestions for instrument improvement. The experts assessed assessment items through their evaluation of KSSM Form 4 Quadratic Functions standards, which the items showed strong alignment with. The experts found that using Malaysian real-world scenarios as context for assessment items provided students with relevant problem-solving situations. The feedback confirmed that the items successfully developed cognitive skills according to the assessment results. One expert stated that the items successfully scaffold from Applying to Creating as per the revised Bloom's Taxonomy. The experts proposed minor enhancements, which included making specific terms easier to understand and developing better guidelines for calculator-based tasks. One expert recommended that the Conjecturing items need a short example which shows expected reasoning.

Reliability Analysis

Table 3

Reliability Statistics from Pilot Testing (n=20)

Section	Cronbach's α	Item-Total Correlation Range	Mean Inter-Item Correlation
Calculator Proficiency	0.76	0.45-0.68	0.52
HOTS Assessment	0.89	0.62-0.78	0.67
MT Assessment	0.87	0.58-0.73	0.63
Full Instrument	0.91	0.41-0.79	0.59

The research instruments reached Cronbach's alpha values which exceeded the minimum threshold of 0.70 established by Nunnally and Bernstein (1994), indicating that they had excellent internal consistency. The Calculator Proficiency section achieved acceptable reliability, with the lowest performance among the three sections showing a reliability of $\alpha = 0.76$, and its item-total correlations met the minimum required standards. The HOTS and MT sections achieved excellent reliability because their items tested a single consistent construct.

All item-total correlations exceeded the minimum acceptable threshold of 0.30, with most results showing correlations between moderate and strong. This pattern shows that all items provide essential value which contributes to their section total score, while each item successfully measures its intended construct. The mean inter-item correlations fell within the optimal range between 0.20 and 0.70, which demonstrated that section items showed enough connection to measure a shared construct, but their relationship did not reach the point of producing identical results. The Calculator Proficiency section reached a mean inter-item correlation of 0.52, which displays appropriate relationships between items that evaluate different aspects of calculator proficiency.

Item Analysis Results

Table 4

Item Analysis Results (Difficulty and Discrimination Indices)

Item	Difficulty Index (p)	Interpretation	Discrimination Index (D)	Interpretation
CP1	0.65	Moderate	0.42	Good
CP2	0.55	Moderate	0.38	Fair
CP3	0.70	Moderate	0.45	Good
CP4	0.45	Moderate	0.35	Fair
CP5	0.60	Moderate	0.40	Good
H1	0.50	Moderate	0.48	Good
H2	0.40	Moderate	0.52	Good
H3	0.35	Difficult	0.45	Good
H4	0.30	Difficult	0.55	Excellent
MT1	0.55	Moderate	0.43	Good
MT2	0.45	Moderate	0.47	Good
MT3	0.35	Difficult	0.50	Good
MT4	0.25	Difficult	0.58	Excellent

The displayed items showed their necessary difficulty levels because the established criteria required test results to fall within the 0.25 to 0.75 range for appropriate difficulty assessment (Nunnally & Bernstein, 1994). The item difficulty distribution reflects planned design choices, as Calculator Proficiency items create a moderate difficulty range (0.45–0.70) which allows all students to begin their tests. The HOTS and MT assessments cover a wide range of difficulty levels because their assessments include moderate difficulty items which test between 0.40 and 0.55 and more challenging items which create difficulties for advanced students who exist in the topmost level of the ability range.

The inclusion of "difficult" items ($p < 0.40$) is intentional and theoretically justified. These assessment items evaluate advanced HOTS and MT capabilities, which require students to use cognitive abilities that will be difficult for all students except the top performers. The assessment items H4 (Creating) and MT4 (Convincing) establish the maximum level of their respective assessment categories, which results in their lowest success rates of 0.30 and 0.25.

The assessment items produced valid discrimination indices which exceeded the minimum required standard of 0.30, while most items reached good range scores between 0.40 and 0.49 or excellent range scores of 0.50 and above. The results demonstrated that the instrument properly distinguished between high-performing and low-performing test takers, with items assessing high cognitive levels showing the strongest discrimination ability. The assessment items H4 (Creating) and MT4 (Convincing) established excellent discrimination, which reached 0.55 and 0.58, because these difficult assessment items effectively separate students who possess advanced thinking abilities from other students.

The relationship between difficulty and discrimination follows expected patterns, which show that moderately difficult items tend to produce moderate discrimination, while more difficult items demonstrate greater discrimination. This pattern demonstrates that the instrument operates according to its design because it offers suitable challenges to all ability levels while successfully distinguishing between different student performance levels.

Discussion

The primary outcome of the research project reached its objectives through the development and evaluation of assessment instruments which demonstrated optimal psychometric properties required to address a critical issue in mathematics teaching within Malaysia. The assessment instrument, which contain comprehensive evidence of content validity, reliability, and item quality, have proven their capacity to assess Higher-Order Thinking Skills and Mathematical Thinking abilities within quadratic function contexts. The study presents its main findings which demonstrate their theoretical and practical impacts while showing their connections to existing research work. The validation process established the instruments' validity because experts reached a high level of consensus about the assessment instrument. The experts showed total agreement about the items' relevance and representativeness because all items received an Item-Level Content Validity Index (I-CVI) score of 1.00. The validation panel reached complete agreement after they analyzed intricate constructs which included HOTS according to the revised Bloom's Taxonomy and MT as described by Mason's processes.

The study demonstrates its content validity through unanimous expert agreement, which according to Polit and Beck (2006) requires five or more experts to achieve an I-CVI score of 0.78 or higher. Yusoff (2019) established that perfect I-CVI scores across all items demonstrate exceptional alignment of item assessment with target constructs, because this evidence shows that instruments effectively measure fundamental components of HOTS and MT as described in theoretical frameworks. The evidence demonstrates that the instruments successfully created measurable assessment items from theoretical concepts which mathematics education professionals can easily identify and comprehend.

The assessment maintains theoretical validity because its elements match the Form 4 KSSM curriculum standards for Quadratic Functions, which permits authentic evaluation. The study achieves ecological validity through its method of balancing theoretical accuracy with educational standards, because it demonstrates that valid assessments which meet testing criteria can be used in actual classroom environments. Educational research which Cohen et al. (2018) defined as ecological validity measures how well research findings and instruments apply to real-world educational environments. The current instruments maintain their educational value for Malaysian mathematics teachers because they combine established theory with national curriculum standards.

The instruments demonstrate high reliability because they achieve internal consistency coefficients which exceed minimum research standards. Nunnally and Bernstein (1994) established that reliability coefficients of 0.70 or higher are acceptable for research purposes, but values above 0.80 indicate good reliability and values above 0.90 represent excellent reliability suitable for individual-level decision-making. The values demonstrate research excellence because they exceed research standards while they match or exceed results from other research studies. The study's reliability coefficients for both HOTS and MT assessments exceed the $\alpha = 0.82$ standard established by Zulnaidi et al. (2020) for a HOTS mathematics test and the $\alpha = 0.85$ standard of Mulyati et al. (2023) for an MT assessment with secondary students. The development process produced superior reliability because it used rigorous methodological standards.

The researchers implemented their planned assessment method through a structured procedure which required them to identify specific assessment criteria and multiple validation checkpoints which they used to evaluate their testing materials before their testing activities began. The assessment method used by researchers tested students' knowledge of quadratic functions which created better internal test consistency because all assessment items tested one specific knowledge area. The accurate measurement system provides direct evidence of student skills because it removes random testing errors from the scoring process.

The findings show important consequences because they demonstrate how theoretical frameworks can be applied across different cultural contexts. The study found that the MT assessment achieved a reliability coefficient ($\alpha = 0.87$), which matched the results from Mulyati et al. (2023) conducted in Indonesia, showing that Mason's mathematical thinking framework can be used as an operational model throughout Southeast Asian educational systems. Although the two research groups discovered similar reliability results, their study findings show that research teams must conduct local research to create new assessment instrument which suit local needs and educational content.

The study achieved positive results because its findings showed better outcomes than the obstacles which other researchers faced according to their findings. Leong and Tan (2022) reported their research results about complex reasoning assessments for Malaysian students which produced lower reliability results through an alpha coefficient of 0.79, because their research proved that complex cognitive processes required high internal consistency for accurate measurement. The current assessment instrument demonstrate their superior reliability because their development process follows precise design steps to assess complex thinking skills with high accuracy. The validated instruments provide assessment quality improvements for mathematics education research, and according to these comparative results, they emerge as a major enhancement for assessment practices in this field.

Limitations and Future Directions

The instruments demonstrate excellent psychometric properties. However, the assessment process needs to consider various limitations that restrict evaluation of the instruments. The first validation used a small pilot sample which included only 20 participants; additional testing with larger and more diverse student groups would improve the generalizability of study results. The instruments currently test only quadratic functions, which restricts their applicability and requires additional validation for use in other mathematical content areas. The initial development stage did not include testing of test-retest reliability to examine score stability over time, nor did it evaluate criterion validity by comparing instrument scores with established measures of mathematical proficiency such as SPM results or TIMSS performance.

The study must find solutions for its existing shortcomings by implementing the following approaches.

1. The researchers need to execute expansive validation assessments throughout all Malaysian territories while testing national schools, religious schools, fully residential schools, and schools located in urban and rural areas.
2. The researchers must develop and evaluate their assessments for all mathematical subjects, which include statistics, geometry, trigonometry, and linear programming, across all secondary school grade levels.

3. The researchers must establish criterion and predictive validity by studying instrument score relationships with standardized test results and assessing whether initial performance indicates future success in advanced mathematics courses.
4. The researchers must develop equivalent assessment instrument which enable multiple assessments without creating training effects, to support pre-post studies and student progress monitoring across time.
5. The researchers need to use qualitative research methods which include cognitive student interviews and teacher consultations to gain deeper understanding of assessment item performance in actual classroom environments and to identify potential sources of misinterpretation or difficulty.
6. The researchers need to study how teachers can use these instruments to create formative assessments which help them discover student misunderstandings, tailor their teaching, and offer specific feedback that improves students' mathematical reasoning skills.
7. The researchers need to study how instrument scores connect with classroom results through research that follows students' development over multiple academic terms and school years.

Conclusion

This study developed and conducted an initial validation of assessment instruments designed to measure Higher-Order Thinking Skills (HOTS) and Mathematical Thinking (MT) abilities among Malaysian Form 4 students studying quadratic functions. The instruments demonstrated acceptable content validity, good internal consistency, and preliminary item-level quality in the pilot sample. They are intended to support teachers in identifying students' cognitive strengths and needs in relation to the higher-order cognitive abilities that national policy requires.

The evidence obtained, including expert agreement ($I-CVI = 1.00$; $S-CVI = 0.98$), internal consistency ($\alpha = 0.76-0.89$), and preliminary item difficulty and discrimination indices, provides initial support for the instruments' content validity and internal consistency. However, this study does not establish construct, criterion, predictive, or practical validity. As acknowledged in the Limitations section, the small pilot sample ($n = 20$) and single-state context mean that the instruments should be regarded as promising tools requiring further validation, not as measures that have been shown to accurately capture advanced thinking or to produce scores free from measurement error.

The Malaysia Education Blueprint 2013–2025 establishes challenging student thinking objectives that require both policy change and resource allocation for teacher support. The instruments offer teachers a curriculum-aligned resource for formative assessment purposes, pending further validity evidence.

The assessment development model used in this study combines theoretical frameworks with national curriculum standards and expert pilot testing. This approach offers a replicable framework for assessment development in mathematics education, and future researchers can follow it to develop assessments for other content areas, grade levels, and educational contexts.

The instruments presented here contribute to the Malaysian education system's current emphasis on HOTS development by providing a standards-based starting point for formative assessment and research. Realising national policy objectives will require continued validation of such instruments across diverse contexts, alongside broader curricular and pedagogical support. This study provides an initial resource that can support teachers, researchers, and policymakers in their efforts to develop students' mathematical thinking abilities.

Conflict of Interest

The authors declare no conflict of interest.

References

- Abdul Rahman, N. A., Ismail, N., & Zulnaidi, H. (2022). Penggunaan kalkulator saintifik dalam kalangan pelajar sekolah menengah: Cabaran dan implikasi terhadap pembelajaran matematik. *Jurnal Pendidikan Sains dan Matematik Malaysia*, 12(1), 45-58.
- Abdul Wahid, N. A., Mohd Ayub, A. F., & Kamarudin, N. (2025). Penjanaan soalan automatik untuk matematik sekolah menengah: Analisis kurikulum dan kesahan kandungan. *Malaysian Journal of Learning and Instruction*, 22(1), 112-130.
- Abdullah, N., & Halim, L. (2021). Implementation of higher order thinking skills in teaching science: A case study in Malaysia. *International Journal of Educational Research*, 105, 101699.
- Abu Kasim, N. A. (2025). Analisis soalan bukan rutin dalam buku teks matematik tingkatan 4 KSSM. *Jurnal Kurikulum & Pengajaran Asia Pasifik*, 13(1), 23-38.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Andu, R., Zamalik, N. A., & Jun, L. Z. (2025). Enhancing mathematical fluency through scientific calculator instruction: A quasi-experimental study. *International Journal of Science and Mathematics Education*, 23(2), 267-285.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). McGraw-Hill Education.
- Che Mansor, N. A., Abdullah, A. H., & Osman, S. (2025). Integrasi perisian matematik dalam pembelajaran fungsi kuadratik: Kesan terhadap pemahaman dan kemahiran kognitif. *Jurnal Teknologi Pendidikan Malaysia*, 15(1), 78-94.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.
- Drijvers, P. (2015). The role of digital technologies in mathematics education: A European perspective. *Mediterranean Journal for Research in Mathematics Education*, 14, 1-12.
- Embong, Z., Zulnaidi, H., & Abd Rauf, R. (2025). Pendekatan pengajaran visual dalam matematik: Kesan terhadap kemahiran berfikir aras tinggi. *Jurnal Pendidikan Malaysia*, 50(1), 33-48.
- Faizah, S., Nusantara, T., Sudirman, S., & Rahardi, R. (2020). Profil proses berpikir matematis siswa dalam pembuktian matematika berdasarkan kerangka Mason. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 5(3), 289-298.
- Ibrahim, N., Ayub, A. F. M., & Siraj, S. (2023). Kemahiran pemikiran matematik dalam kalangan pelajar tingkatan empat: Analisis berdasarkan kerangka Mason. *Jurnal Pendidikan Sains & Matematik Malaysia*, 13(1), 15-28.
- Iskandar Shah, M. M., Rosli, R., & Maat, S. M. (2024). Analisis kesukaran pelajar dalam topik fungsi: Satu kajian di sekolah menengah. *Jurnal Dunia Pendidikan*, 6(1), 102-115.

- Jun, L. Z. (2025). Hubungan antara kemahiran kalkulator dengan kemahiran berfikir aras tinggi dalam kalangan pelajar sekolah menengah. *Jurnal Pendidikan Matematik*, 13(2), 45-62.
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2-10.
- Kissane, B. (2016). The scientific calculator and school mathematics. *Southeast Asian Mathematics Education Journal*, 6(1), 3-16.
- Leong, K. E., & Tan, J. Y. (2022). Assessing mathematical reasoning in Malaysian classrooms: Teachers' practices and challenges. *Malaysian Journal of Learning and Instruction*, 19(1), 1-28.
- Loh, S. C., Pang, V., & Lajium, D. (2017). Penilaian kemahiran berfikir aras tinggi dalam kalangan pelajar sekolah menengah di Sabah. *Malaysian Journal of Learning and Instruction*, 14(2), 95-122.
- Mason, J., Burton, L., & Stacey, K. (2010). *Thinking mathematically* (2nd ed.). Pearson.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Ministry of Education Malaysia. (2013). *Malaysia education blueprint 2013-2025 (Preschool to post-secondary education)*. Ministry of Education Malaysia.
- Ministry of Education Malaysia. (2016). *Panduan pelaksanaan kemahiran berfikir aras tinggi (KBAT)*. Bahagian Pembangunan Kurikulum.
- Mulyati, S., Priatna, N., & Juandi, D. (2023). Mathematical thinking processes in solving non-routine problems: A case of Indonesian students. *Journal on Mathematics Education*, 14(1), 1-18.
- Nanmumpuni, H. P., Setyosari, P., & Sulton, S. (2022). Pembelajaran guided discovery berbantuan kalkulator untuk mengembangkan higher order thinking skills siswa. *Jurnal Inovasi Teknologi Pendidikan*, 9(2), 145-158.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing.
- Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497.
- Radzuan, N. S., Zamalik, N. A., & Jun, L. Z. (2021). Meningkatkan prestasi matematik pelajar pencapaian rendah melalui penggunaan kalkulator saintifik. *Jurnal Pendidikan Sains dan Matematik*, 11(2), 33-46.
- Sures Kumar, S., & Foo, S. F. (2024). Pembangunan dan pengesahan modul pembelajaran topik matematik tingkatan 4 KSSM. *Jurnal Pendidikan Malaysia*, 49(1), 67-82.
- UTM NewsHub. (2023, October 15). *Pentaksiran matematik berfokus pemikiran aras tinggi* [Press release]. Universiti Teknologi Malaysia.
- Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49-54.
- Zamalik, N. A., Jun, L. Z., & Andu, R. (2024). Intervensi penggunaan kalkulator saintifik dalam kalangan pelajar sekolah menengah: Kesan terhadap pencapaian matematik. *Jurnal Pendidikan Sains & Matematik Malaysia*, 14(1), 23-38.
- Zeynivandnezhad, F., Samsudin, M. A., & Ahmad Tarmizi, R. (2013). Mathematical thinking processes of pre-service teachers in problem solving: A case study. *Procedia - Social and Behavioral Sciences*, 93, 194-199.

- Zulnaidi, H., Oktavika, E., & Hidayat, R. (2020). Effect of use of GeoGebra on achievement of higher order thinking skills. *International Journal of Instruction*, 13(2), 847-862.
- Zulkifli, N. N., Halim, L., & Osman, K. (2025). Penggunaan kalkulator saintifik dalam pembelajaran matematik: Kesan terhadap motivasi dan keyakinan diri pelajar. *Jurnal Pendidikan Malaysia*, 50(2), 45-60.