International Journal of Instruction
e-ISSN: 1308-1470 ● www.e-iji.net

April 2025 ● Vol.18, No.2
p-ISSN: 1694-609X
pp. 593-614

Article submission code:
20241022025056

Received: 22/10/2024
Revision: 21/12/2024

Accepted: 01/01/2025
OnlineFirst: 07/01/2025

Unlocking Spatial Wisdom: A Polya-Inspired Approach to Strengthening Year 5 Students' Higher-Order Thinking Skills

Dantong Li
Universiti Teknologi Malaysia, Johor Bahru, Malaysia, *lidantong@graduate.utm.my*

Sharifah Osman
Corresponding author, Universiti Teknologi Malaysia, Johor Bahru, Malaysia, *sharifah.o@utm.my*

Najua Syuhada Ahmad Alhassora
Universiti Teknologi Malaysia, Johor Bahru, Malaysia, *najuasyuhada@utm.my*

The development of higher-order thinking skills (HOTS) is central to educational objectives. There are few instruments to cultivate students' HOTS, and an instrument based on the Polya problem-solving strategy was designed. This study assessed the HOTS among 23 students in a spatial field in Johor Bahru, Malaysia. Utilising the content from Chapter 6 of the current Malaysian school curriculum textbook for Year 5, which is based on the Kurikulum Standard Sekolah Menengah(KSSM), this research employs a detailed scoring rubric to evaluate the quality of student responses and their adherence to the Polya problem-solving strategy across a series of spatial tasks. The results indicate a varied performance level, with some students demonstrating excellent problem-solving abilities, while others showed room for improvement. A significant decline in student performance occurs as task complexity increases, particularly in the design and steps of problem-solving. Only two out of 23 students met the requirements of HOTS, emphasising the necessity of differentiated teaching strategies. The study also examined the impact of gender on these skills and found no substantial effect. When examining the effect of different ethnic groups on students' HOTS scores, it was found that the Malay-Chinese group had the highest mean ranking (14.00), followed by the Malay group (12.39), while the Indian group had the lowest mean ranking (10.50). Although there were some differences in rankings between different ethnic groups, such differences were not large enough to be statistically significant, and the sample should be expanded. This study recommends an instrument to strengthen HOTS and provides a replicable framework for future spatial reasoning research.

Keywords: polya problem-solving strategy, higher-order thinking skills (HOTS), educational assessment, spatial reasoning, differentiated instruction (DI)

Citation: Li, D., Osman, S., & Alhassora, N. S. A. (2025). Unlocking spatial wisdom: A polya-inspired approach to strengthening year 5 students' higher-order thinking skills. *International Journal of Instruction*, *18*(2), 593-614. https://doi.org/10.29333/iji.2025.18232a

INTRODUCTION

It is critical for preparing students to excel in a complex, dynamic world and cope with changes in the 21st century (Atit et al., 2020). The incorporation of higher-order thinking skills (HOTS) has become one of the reform measures used to cultivate students' critical and creative thinking (Shaheen, 2010; Qasrawi & BeniAbdelrahman, 2020). At present, the definition of HOTS can be divided into two categories. One is the concept proposed on the basis of the taxonomy of educational objectives, such as Barak (2008) proposed that higher-order thinking (HOT) refers to all intellectual activity tasks beyond information retrieval. Lewis and Smith (1993) believe that HOTS refers to the process of people associating new information with information stored in memory and reorganizing it to achieve a certain purpose or find possible answers in a complex situation. Compared with the two categories of concepts, the first category, which starts with the taxonomy of educational objectives, defines HOTS in a more specific, hierarchical, and operational way. This is mainly because the taxonomy of educational objectives itself is a theory used to guide the teaching and testing of thinking skills, whereas the second type of concept starts from the characteristics of HOTS itself, which can describe its characteristics more clearly and is closer to the essence. Some scholars, such as Dewey (1933, 1986), consider that if "HOT" is defined as a series of comprehensive skills produced at a high level of cognition, then it is "HOTS". In HOTS, the core elements all involve comprehensive analysis, problem solving, innovation, and creation.

This study has high requirements for the pertinence and operability of concepts, so the first type of HOTS concept is adopted. As emphasized in the revised Bloom's Taxonomy (2001), these different levels of cognitive thinking skills are divided into two levels: lower-order thinking skills (LOTS) and higher-order thinking skills (HOTS). To reflect the cognitive complexity of educational goals, skills are categorized from basic knowledge recall to the creation of new ideas or products (Anderson; Krathwohl, 2001). The development of HOTS includes applying, analyzing, evaluating, and creating; it involves understanding the content of the material; and the process of decomposing, integrating, or judging knowledge, that is, the reorganization of knowledge.The HOTS paradigm is advocated not only to improve academic outcomes but also to prepare students for the demands of modern citizenship in a complex, dynamic society. HOTS involve the ability to analyze and classify information, then connect it to other concepts in order to solve problems through specific reasoning and manipulation strategies (Ratna & Retnawati, 2019; Rohid et al., 2019).

Problem solving is an integral part of disciplinary learning, driving deep understanding and practical application of knowledge (Ling et al., 2019). Polya's Problem-Solving Theory offers a structured approach that can help demystify these complex cognitive processes.The steps outlined by Polya—understanding the problem, devising a plan, carrying out the plan, and reflecting upon the solution—establish a foundation for cultivating analytical and strategic thinking in students (Polya, 1945). Each step is a pedagogical tool, providing a scaffold for learners to build upon their existing knowledge and encouraging a methodical approach to mathematical challenges (Weiland & Poling, 2022). Dewey (1986) discussed the mechanism of thinking and suggested that "problems" are the greatest source of motivation for HOTS. This tradition has been continued by researchers such as Resnick (1987) and Halpern (1998), who further explored the characteristics and development of HOTS and argued that HOTS is complex and self-regulating, involving the application of various criteria to generate a range of problem-solving strategies. In addition to recent contributions such as those of Zhou et al. (2024), they not only express enthusiasm for HOTS research around the world but also contribute to the advancement of HOTS by developing and validating the HOTS construct for students.

Abosalem, Y. (2016) reported that performance-based assessments can help teachers gain insight into the student learning process. It provides us with information about the daily progress of students. Widana (2017) highlights the characteristics, steps, and benefits of developing HOTS assessments for teachers. Sofyan et al.'s (2024) research suggested that providing open-ended problems (OEP) or open-ended questions means that the problem can be solved in multiple ways (flexibility). Therefore, solving OEP problems is used to examine students' thinking and analytical abilities in mathematics classes, and a descriptive presentation method and case study method are used. In the field of education, although these articles are relatively rich in methods for assessing HOTS, there is a clear lack of methods for effectively cultivating these skills. These studies aim to cultivate teachers' ability to assess HOTS (Abosalem, 2016; Widana, 2017; Paul & Elder, 2016; Sofyan et al., 2024; Mulyoto et al., 2024), but ignore how to implement HOTS in practice to cultivate students' HOTS.

HOTS learning is often accompanied by different performances among students, which inevitably leads to differences in the results of the learning process. Differentiated instruction (DI) is a systematic teaching method that requires teachers to adjust teaching plans or tasks according to differences in students' learning levels. Research shows that DI can improve students' mathematical understanding, especially in primary school. Teachers can employ various strategies, such as flexible grouping and tiered assignments, to better address the diverse learning needs in a mathematics classroom. By providing students with tailored problems that match their ability levels, educators can facilitate a deeper understanding of mathematical concepts (Aladwan et al., 2023).Although DI has many advantages in theory, it may face some challenges in practice, including the level of teachers' professional ability, how to manage the classroom, and how to effectively meet the individual needs of different students (Bushie, 2015; Smale-Jacobse et al., 2019; Hu, 2024). Furthermore, Bondie et al. (2019) emphasized that future research should focus on how to customize differentiated teaching tasks in the classroom to meet the ability levels of different students.

The development of HOTS often interferes with several external factors, and this literature explores the impact on individual development. According to the research results of Liu et al. (2024), the classroom environment and students' psychological and intellectual characteristics have a direct effect on HOTS. Kheloui et al. (2023) reported that there are gender differences in human cognitive abilities, which arise from a combination of biological and social processes. Although sex differences in cognitive ability are frequently reported, the magnitude of these differences and whether they have practical significance for boys’ and girls’ educational outcomes remain highly controversial (Ying et al., 2020). Gender role attitudes and stereotypes may affect expectations and performance in education (Reilly, 2012; Jäncke, 2018).

Specifically, some studies have explored the impact of gender on mathematical achievement. Kustati Martin and Nana Sepriyanti (2022) explored the impact of 21st century learning methods on HOTS and the mathematical literacy of Indonesian science students. The study revealed that boys generally performed better than girls did in multiple cognitive dimensions. According to Rahayuningsih et al. (2019), in solving the mathematics of group theory, boys outperform girls in analyzing, while girls outperform boys in evaluating, and boys and girls score the same but zero in creating. However, in a study of critical competence in mathematics by Widyastuti, Erna, and Hella Jusra (2022), compared with boys, girls were able to describe and explain the information they obtained when solving HOTS problems in more detail and coherence. These studies suggest that whether gender differences affect students' HOTS performance depends on many factors, such as different math domains and different countries.

HOTS are also particularly relevant in the mathematical context of spatial reasoning, where students must apply their understanding in novel and abstract contexts. Spatial reasoning is fundamental for various academic and professional fields beyond traditional STEM (science, technology, engineering, and mathematics). It involves the ability to understand and manipulate spatial relationships and can be predictive of success in a wide array of disciplines. Spatial reasoning encompasses both two-dimensional (planar) and three-dimensional (solid) aspects. It includes the ability to understand and reason about the relationships within planar and solid spaces (Mulligan et al., 2018, Antolí, 2018, Petersen, 1985).

Wang et al. (2024) reported that although there is ample evidence to support the value of developing spatial skills in childhood, previous studies have focused mostly on secondary school students, with limited attention given to primary school students. However, spatial learning is essential for developing skills such as visualization, logical reasoning, and problem solving in primary school (Lu et al., 2022, Liu et al., 2024), which are applicable across a range of scientific disciplines. One of the barriers to spatial reasoning is the lack of deep engagement, and the abstract nature of spatial learning often requires students to go beyond memorization and embrace complex thinking. The chronic lack of spatial ability and spatial thinking may lead to socioeconomic and gender differences in education (Fujita et al., 2020; Pavlovičová et al., 2022; Preciado-Babb, 2020). Additional, Malaysia is a multi-ethnic and multicultural country, and its classrooms often reflect this diversity with students from various ethnic backgrounds, including Malays, Chinese, Indians, and indigenous groups. This diversity provides a unique educational environment that requires tailored teaching strategies to address the needs of different ethnic groups (Department of Statistics Malaysia, 2010). According to the analysis of the above literature, this study proposed four research questions to study students' HOTS in the spatial field:

RQ1. How can an instructional instrument based on Polya's problem-solving model be developed to enhance higher-order thinking skills (HOTS) among Year 5 students?

RQ2. What role do differentiated teaching strategies play in fostering higher-order thinking skills (HOTS) among Year 5 students, and how can these strategies be effectively implemented?

RQ3. What differences, if any, exist in the development of HOTS between male and female students, and how should teaching strategies be adapted to address these differences?

RQ4. Is there a significant difference in higher-order thinking skills (HOTS) performance among students from different ethnic groups?

METHOD

Research Design and Sampling

This study adapts a developmental research approach to create, test, and refine an instructional instrument based on Polya's problem-solving model. A quantitative approach was used to provide a comprehensive understanding of the effectiveness of the instrument and teaching strategies.

In Year 4, students learned more about the theme of "space" in the plane, including angles, areas, etc., and they only learned the volume formula of rectangular prisms and cubes. In Year 5, they learned complex shapes, which is more suitable for investigating students' HOTS ability. So a sample of 23 Year 5 students was selected from a primary school in Malaysia, with an emphasis on aligning with the Kurikulum Standard Sekolah Menengah (KSSM) framework. Purposive sampling was employed to ensure that students from different performance levels were included. The small sample size allowed for in-depth analysis and piloting of the instrument before broader implementation. The demographic structure of the study included 23 students; 12 were boys (52.2%) and 11 were girls (47.8%), with a nearly equal ratio of boys to girls. The ethnic composition of the class included 9 Malays (39.1%), 5 Malay-Chinese (21.7%), and 9 Indians (39.1%), reflecting rich cultural diversity.

Research Framework

To meet the challenges of teaching and learning spatial reasoning, this study adopts the Polya problem-solving strategy. This is a systematic approach to teaching students how to solve problems, emphasizing a step-by-step analytical process that includes understand, design, carry out, and look-back steps that foster an analytical ability to solve problems (Polya,1945). To develop HOTS in classroom teaching or individual learning, a series of measures should be designed in advance to intervene in and guide the learning process. Figure 1 shows the framework of the entire study, which provides a clearer and more intuitive understanding of the research content.

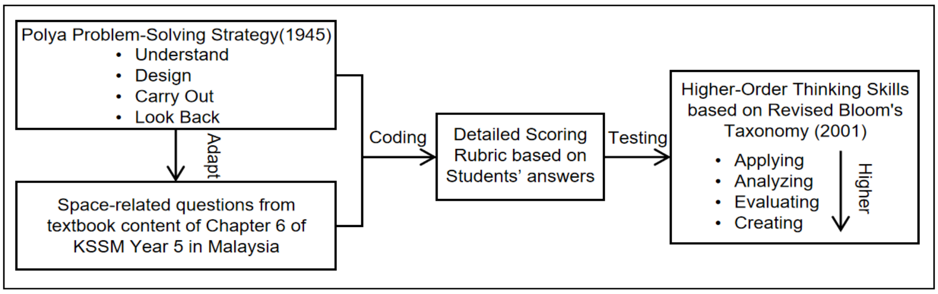


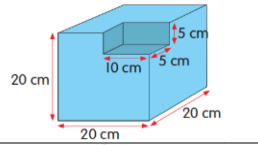
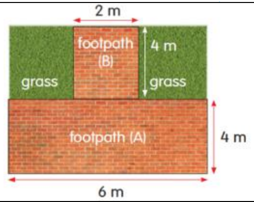
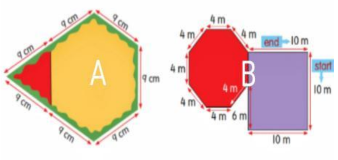
Figure 1
Research framework

Student Answer Sheet and Scoring Rubrics

The validity and reliability of the questions Q1–Q4 (Refer Table 1) and the scoring rubrics (refer Table 2) have been thoroughly evaluated by two experts: an experienced primary school mathematics teacher and a mathematics education researcher. Their expertise ensured that the questions were aligned with the research objectives and suitable for the target group. Additionally, the researchers have ensured that the items are clear, appropriate, and effective in capturing the intended constructs.

Table 1 shows the content of the student test paper and the measurement questions adapted from the Polya problem-solving strategy. The difficulty is in a gradient form to measure students' HOTS, with one question corresponding to one level of HOTS.

Table 1
Student answer sheet adapted from the Polya problem-solving strategy to measure students' high-order thinking skills

Questions		Subquestions
Q1 In a design project, a cuboid was cut and removed from a cube. Calculate the volume of the cuboid removed.		a. What are the key information points in the question? Please list them. b. What are your approaches to solving this problem? c. List the steps to solve the problem. d. Check again and write the final answer. (Each question will have 4 subquestions based on Polya Problem-solving Strategy)
Q2 Dad wants to build a brick footpath in the house compound as shown in the diagram. He has allocated RM4 000 for this. The cost of building 1 m^2 brick footpath is RM100. Is there enough money?		
Q3 Mickey said that shape A has a longer perimeter than shape B. Minnie said that shape B has a longer perimeter. Who is right?		
Q4 There is a rectangular piece of paper in Jaha's hand. Divide the rectangle into other equal parts in at least three ways and show the process.		

Detailed Scoring Rubrics for Data Coding

The purpose of constructing HOTS evaluation indicators is to solve the problem of how to measure the HOTS of primary school students. Scoring criteria were developed on the basis of students' answers, and students' responses were coded using points so that differences in students' scores could be more clearly seen. Table 2 provides detailed explanations.

Table 2
Detailed scoring rubrics based on students’ answers

Criteria / Question	0 Point	1 Point	2 Points	3 Points
a. Key Information Points	Failed to list any key information points.	The listed key information points are partially incorrect or incomplete.	Key information points are mostly correct but it lack detail.	Accurately and comprehensively listed all key information points with detailed explanations.
b. Approach to Solve the Problem	No solution is proposed, or the proposed solution is completely irrelevant.	Proposed solution is relevant but it lacks clarity and logical coherence.	The proposed solution is relevant and logical, but lacks innovation.	Proposed solution is not only relevant and logical but also innovative and practical.
c. Steps to Solve the Problem	Failed to list steps, or the listed steps are completely incorrect.	The steps listed are basically relevant but contain significant errors or omit important steps.	The steps are correct, clearly expressed, but may lack precision in execution.	The steps are completely correct, logically clear, and precisely executed.
d. Review the Answer	No review was conducted, or errors persist after review.	Review conducted but not thoroughly; some errors were missed.	Review is thorough, most errors were identified and corrected.	The review is very thorough, ensuring the answer is completely correct.

Legend:

(1) Each step is pointed out independently and is based on the student's specific performance in that step.

(2) Teachers should decide specific points on the basis of students’ actual operations and performance.

Table 3 provides a concise summary of the research methods employed to address four distinct research questions, detailing the data collection and analysis processes for each.

Table 3
Summary of Research Methods

Research Question	Data Collection	Data Analysis
RQ1	The researchers used a quantitative research method to select 23 fifth-grade students from a primary school in Malaysia as a sample to assess and improve their higher-order thinking skills (HOTS). The research team designed a teaching tool based on George Polya's problem-solving strategies, combined with the content of Chapter 6 of the fifth-grade textbook of the National Primary School Science Curriculum (KSSM), and constructed four task problems with difficulty gradient space domains based on Bloom's higher-order thinking skills, and each problem has four sub-problems containing Polya's problem-solving steps.	In order to comprehensively evaluate students' responses, the researchers developed a detailed rating scale with four categories: key information points, problem-solving methods, problem-solving steps, and answer review, each of which was scored from 0 to 3 points based on Polya's problem-solving model. Through this scoring system, the researchers were able to quantify students' ability to understand and apply Polya's problem-solving strategies. In addition, the researchers conducted descriptive analyses to evaluate students' performance on questions of different difficulty levels, thereby gaining insight into students' abilities in higher-order thinking skills.
RQ2	Researchers employed a detailed scoring rubric to evaluate students' responses and recorded each student's total score to assess their problem-solving skills and higher-order thinking skills (HOTS). Based on the scores, student performance was categorized into three levels: Lack of Problem Solving Skills, Average or Developing Problem Solving Skills, and Exhibiting Higher-Order Thinking Skills (HOTS).	Descriptive statistical analysis was conducted to calculate the minimum, maximum, mean, and standard deviation of the students' total scores, providing insight into the overall distribution of student performance. Students were then categorized into three levels based on their scores: those scoring between 0 and 12 were identified as lacking problem-solving skills, those with scores between 13 and 28 were considered to have average or developing problem-solving skills, and those scoring between 29 and 48 were recognized for exhibiting higher-order thinking skills (HOTS).
RQ3	The data collected were divided into 12 boys and 11 girls according to gender and then the performance scores of male and female students on HOTS questions was compared.	An independent samples t-test was used to compare the mean scores of male and female students to determine whether gender had a significant effect on performance.
RQ4	Collect HOTS scores of students of different ethnic groups including Malays, Malay-Chinese and Indians.	The Kruskal-Wallis test was used to compare the differences in HOTS total scores among the three ethnic groups to determine whether ethnic background had a significant effect on the measure.

FINDINGS

The overall performance level of students on higher-order thinking skills questions

Q1-Q4 students' responses

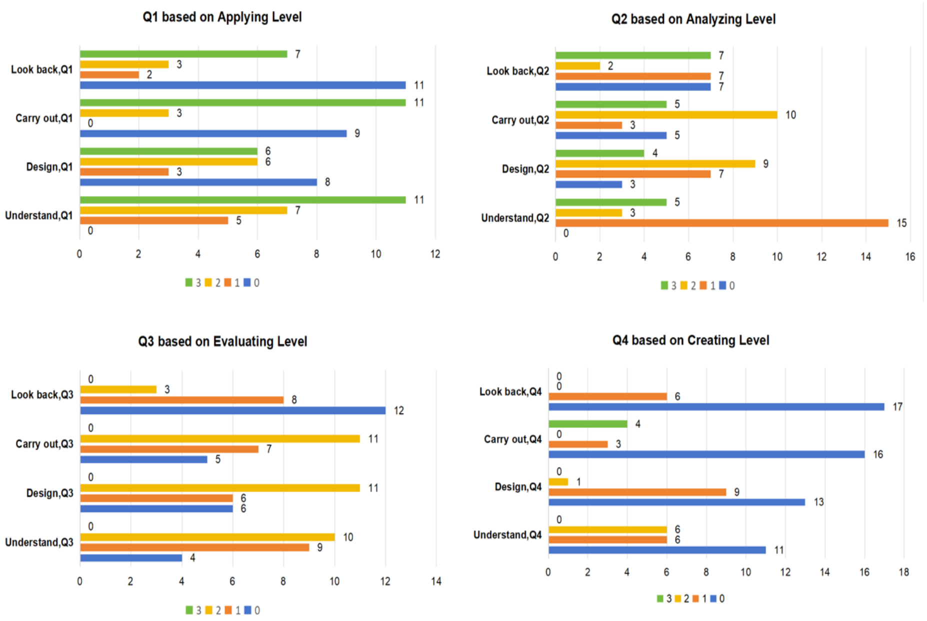


Figure 2
Q1-Q4 Points of each step

Figure 2 shows the students' responses in Q1–Q4. Each question has four small steps: understand, design, carry out, and look back. Each step is evaluated at 0–3 points. The number of students who received 0–3 points is shown in the picture, which allows us to see how the students responded and how many points they received for each step.

In Applying Level question (Q1), which focuses on applying known concepts, students performed consistently well in the understanding step, with many scoring 3 points. However, their performance showed greater variation in the design and carry out steps, where fewer students reached the highest score. The look back step saw a balanced range of scores, indicating moderate ability to reflect on their solutions.

As the difficulty increases in the analysis level (Q2), students perform better in the understanding step, with 15 students scoring 1, 3 students scoring 2, and 5 students scoring 3. However, in the more complex design and carry out steps, more students scored 2 than in the understand step, indicating that there are some challenges in translating understanding into action, but the difficulty can be overcome. The review step produced a bimodal score distribution, indicating that the level of reflective ability varies.

Q3 (Evaluation Level) requires more evaluation skills, so the performance is more uneven, but no student scored 3 points in the 4 steps. In the understanding stage, 10 students scored 2 points, 9 students scored 1 point, and 4 students scored 0 points. The design and implementation stage is very challenging, and most students scored 1-2 points. In the review stage, 3 students scored 2 points, 8 students only scored 1 point, and 12 students scored 0 points, indicating that reflection and self-evaluation are difficult at this stage.

The most complex question, Q4 (creative level), was meant for testing students' creative thinking. Six students scored 2 points in the understanding step, and the design step scores dropped sharply, with only one student scoring 2 points and nine students scoring only 1 point, but three students scored 3 points in the execution stage. However, the largest number of students (17) scored 0 points in the look-back step, indicating that they had more difficulty reflecting on and evaluating their solutions. The poor performance of students in the look-back step indicates that students have poor awareness and less learning or training in this area.

Overall, students had a relatively strong understanding of questions Q1 and Q2, with some students scoring 3 points, but no students scored 2 points in the more complex Q3 and Q4. In the questions Q1-Q4, as the difficulty increased, some students still scored 2 points in the first two steps of the understand and design step. No student scored 3 points in any step of Q3, which shows that students' evaluation ability needs to be strengthened, but 3 students scored 3 points in the carry out step of Q4, which shows that some students have the ability to discuss in categories or the knowledge points of the questions are familiar to students, so they can solve the difficulties. More than half of the students scored 0 points in the look-back step of Q3 and Q4, which shows that students are better at reflection and evaluation when facing evaluation and creative challenges.

Description of each step of the Polya problem-solving strategy

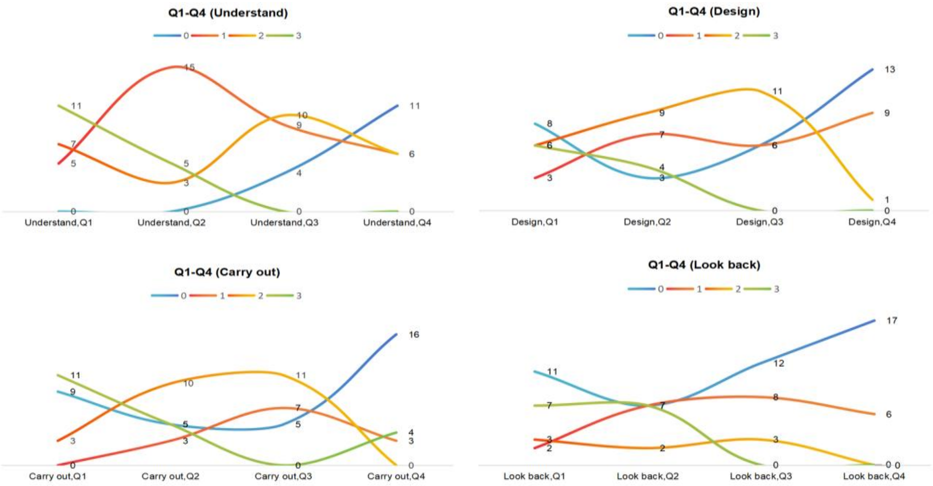


Figure 3
Q1-Q4 Descriptions of the understand, design, carry out and look back steps

Figure 3 compares the responses of each step in Q1–Q4. It shows the students' answers in the understanding, design, carry out and look back sections in Q1–Q4, with points of 0--3 and the number of students. The difficulty of questions in Q1–Q4 gradually increases, testing students' HOTS in applying, analyzing, evaluating, and creating.

In the understanding step (first graph), even as the difficulty increased, the high scores in this step indicate that students' understanding is not always hindered by the complexity of the task. The designing step (second graph) shows more variability as the tasks become more complex. For example, Q3 is more difficult than Q2, but more students scored 2 points in Q3 than in Q2. This suggests that when students are familiar with the analysis required, they are better able to design solutions, even if the overall task is challenging.

In the carry out step (third graph), the performance again fluctuates, but there are some notable strong responses even in the more difficult questions. No students scored 3 points in Q3, but four students scored 3 points in Q4, which again verifies the design step conclusion, suggesting that executing solutions can vary depending on both the task and the student's familiarity with the required processes.The look back step (fourth graph), Q2 is more difficult than Q1, but the number of 0 points in Q2 is less than that in Q1. Attention should be given to students' responses in Q1 and their grasp of key concepts, indicating that reflection and evaluation skills may vary by task. Additionally, no student achieved a score of 3 in Q3 and Q4, highlighting the need to strengthen their ability to provide effective feedback on more challenging questions.

In conclusion, Figure 3 reveals that while the difficulty of questions increases, students' performance does not always decline across all steps. Teachers and educators should take into account both the complexity of the task and the students' prior knowledge and experience when assessing their problem-solving abilities.

Analyzing the quality of students' responses and their congruence with the polya problem-solving strategy via a detailed scoring rubric

Table 4
Overall performance of student points

N	Minimum	Maximum	Mean	Std.Deviation
23	*9*	*31*	*19.22*	*6.789*

Table 5
Dividing students into different levels on the basis of the point criteria

Lack of Problem Solving Skills	• Indicates that the student performs poorly on problem solving and scores low. • The score range can be set between 0% and 25% of the total score. • For a total score of 48, this would be *0 to 12.*
Average or Developing Problem Solving Skills	• Indicates that the student performs averagely and is developing their higher-order thinking skills. • The score range can be set between 26% and 59% of the total score. • For a total score of 48, this would be *13 to 28.*
Exhibiting Higher-Order Thinking Skills (HOTS)	• Usually, means that the student performs well on most or all problems. • The score range can be set between 60% and 100% of the total score. • For a total score of 48, this would be *29 to 48.*

Table 4 shows the overall performance of the students. On the basis of the criteria in Table 5, 8 students (s2, s4, s7, s9, s11, s14, s15, s22) scored between 0 and 12, indicating that they lacked problem-solving skills. Thirteen students (s1, s3, s5, s6, s8, s10, s12, s16, s17, s18, s20, s21, and s23) scored between 13 and 28, indicating that their problem-solving skills were average or developing. Two students (s13, s19) scored above 29, indicating that they demonstrated higher-order thinking skills.

Differences between males and females in terms of performance on spatial topics

An independent sample t test is performed to compare whether there is a significant difference in the average scores of male and female students on the HOTS questions in Table 6 . If there is a significant difference, it indicates that gender may have an impact on students' performance.

Table 6
Group statistics and Independent sample t test for gender influence

G	N	Mean	Std. Deviation	Std. Error Mean	Levene's Test (F)	Sig	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	95% CI Lower	95% CI Upper
F	11	20.18	7.264	2.19	.43	.52	.64	21	.53	1.85	2.87	-4.12	7.82
M	12	18.33	6.513	1.88	.64	.52	.64	20.198	.53	1.85	2.88	-4.17	7.87

Legend:G:Group;F:Female;M:Male

The mean test score of the male group in Group Statistics is 20.18, with a standard deviation of 7.264. The mean test score of the female group is 18.33, with a standard deviation of 6.513.

Levene's test for the equality of variances, with a p value of .644, indicates that there is insufficient evidence to reject the null hypothesis that the variances of the two gender groups are equal. T test for the Equality of Means, with a p value of 527 when the variances are assumed to be equal and a p value of 0.529 when the variances are assumed to be unequal. Both p values are much higher than the commonly used significance level of 0.05, indicating that there are no significant gender differences. The 95% confidence interval of the difference, with a confidence interval of -4.125--7.822, includes 0, which means that the difference between the two groups is not statistically significant.

In summary, the results of this analysis revealed that there was no significant difference in the scores of the male and female students on this test and that gender was not a significant factor affecting the scores.

Differences ethnic composition students in the performance of higher-order thinking skills (HOTS)

Table 7 showed the results of the Kruskal-Wallis test were used to compare the differences in the total scores in HOTS among the three ethnic groups (Malays, Malay-Chinese, Indians).

Table 7
Kruskal-Wallis test among the three ethnic groups

Group	N	Mean Rank	Kruskal-Wallis H	df	Asymp. Sig.
Malay	9	12.39	.909	2	.635
Malay Chinese	5	14.00			
Indian	9	10.50			

The ranking mean data show that the average ranking of the Malay-Chinese group is the highest (14.00), followed by the Malay group (12.39), and the Indian group is the lowest (10.50). The statistical test results are Kruskal-Wallis H value of 0.909, degree of freedom 2, and significance level (Asymp. Sig.) of 0.635. Since the significance value is greater than 0.05, this indicates that the ranking differences between different ethnic groups are not statistically significant.

DISCUSSION

Intention of designing Questions

This study selected the knowledge points of Chapter 6 and adapted the questions. Table 8 explains why these four questions were selected to study HOTS and provides the corresponding explanations.

Table 8
Aims and Levels of Questions Q1-Q4

Questions	Aims	Levels
Q1	This question is designed to test students' understanding and calculation ability of geometric bodies. Students need to understand the concepts of cubes and cuboids and be able to apply the corresponding formulas to calculate their volumes.	Applying
Q2	This question tests students' mathematical abilities, specifically their ability to budget and solve economic problems in real-world situations. Students are required to use basic arithmetic and unit conversions to determine whether the funds set aside by their father are sufficient to build a specified area of brick sidewalks. This question also involves integrating mathematical concepts with practical situations and cultivating students' mathematical modelling abilities.	Analyzing
Q3	This question tests students' understanding of the concept of perimeter and their ability to calculate the perimeter of irregular shapes. Students need to understand that perimeter is the sum of the lengths of the boundaries of closed figures and be able to apply perimeter calculations to compare the lengths of two irregular figures. Through this question, students can also develop logical reasoning skills, analyse and judge which shape has a longer circumference, and give a reasonable explanation.	Evaluating
Q4	This question tests students' understanding of the concepts of segmentation of geometric shapes and fractions, as well as their problem-solving abilities. Students are required to use geometry knowledge to divide a rectangular piece of paper into at least three equal parts and demonstrate this process. Through this problem, students develop geometric thinking and creative problem-solving skills while deepening their understanding of fractions and geometric shapes.	Creating

Q1-Q4 students' step responses based on the Polya Problem-Solving Strategy

The four questions (Q1-Q4) are meticulously crafted to assess students' abilities in HOTS, encompassing application, analysis, evaluation, and creation. Q1 evaluates students' understanding and calculation abilities regarding geometric shapes, requiring them to apply formulas to calculate volumes. Q2 delves into students' mathematical capabilities to budget and solve real-world economic problems, integrating basic arithmetic and unit conversions to determine budget sufficiency for building brick sidewalks. Q3 challenges students to understand the concept of perimeter and calculate it for irregular shapes, fostering logical reasoning to compare lengths and provide explanations. Q4 tests students' grasp of geometric shape segmentation and fractions, as well as their problem-solving skills, by asking them to divide a rectangular paper into equal parts, enhancing their geometric thinking and creative problem-solving abilities.

As the difficulty of the questions increases, student performance in the steps of understanding, designing, executing, and reviewing declines, with Q4 showing the most significant drop, indicating a greater challenge to students' comprehension and execution. The design and execution steps are pivotal in the problem-solving process and are where students are most likely to encounter difficulties with complex problems. Educators should focus on enhancing student performance in these steps and developing their abilities to review and self-evaluate, which are crucial for learning and growth from each problem-solving experience.

Fifth-grade students are at a critical stage of cognitive development, becoming capable of handling more complex thinking tasks. In the cultivation of HOTS, fifth-grade students can develop specific cognitive abilities such as the ability to apply geometric concepts and formulas to solve practical problems, like calculating volumes (applying level). They can analyze real-world economic issues and use basic arithmetic and unit conversions to address budgeting problems (analyzing level). They can evaluate the perimeter of irregular shapes, logically deduce which has a longer circumference, and provide rational explanations (evaluating level). Lastly, they can creatively divide a rectangular piece of paper into at least three equal parts, deepening their understanding of fractions and geometric shapes (creating level). Through these activities, students not only improve their problem-solving skills but also develop critical thinking and innovation abilities in the process of understanding and creating new knowledge. Educators can facilitate the development of these cognitive abilities by designing challenging problems and providing appropriate guidance.

Discussion students’ HOTS performance and Differentiated Instruction Strategies

The construction of evaluation indicators for HOTS is actually about solving the problem of how to measure primary school students' HOTS. According to the data analysis in Tables 3 and 4, students' performance in the HOTS test was clearly hierarchical. Eight students scored between 0 and 12 points, indicating that they had obvious deficiencies in problem solving. Thirteen students scored between 13 and 28 points, indicating that their problem-solving skills were at an average level or were developing. Two students scored more than 29 points, indicating excellent HOTS. Students who score in the higher range (29-48 points) demonstrate a better understanding and application of the Polya problem-solving strategy, which includes understanding the problem, devising a plan, carrying out the plan, and looking back on the solution process. This suggests that their cognitive skills are more developed in terms of analytical, evaluative, and creative thinking. This distribution emphasizes the differences in cognitive skill development among students, suggesting that differentiated teaching strategies should be adopted for students of different levels to promote the development of HOTS. The factor that makes a difference in cognitive skill development among students, as indicated by the hierarchical performance, is the level of engagement and proficiency in applying problem-solving strategies. By tailoring instruction to meet the specific needs of students at each level, educators can more effectively promote the development of HOTS across the student body.The research provides evidence that tailored instruction can significantly impact student performance and cognitive growth.

Discussion of difference between Male and Female to solving Higher-Order Thinking Skills questions in overall performance

Table 5 and 6 display the results of the independent sample t-test, focusing on HOTS, a high-level assessment of students' abilities. Gender was examined as a potential factor influencing performance, but the results indicated no significant differences based on gender. This finding suggests that gender does not play a substantial role in shaping educational achievement, even for challenging HOTS assessments. Additionally, it implies that teaching strategies need not differ by gender, as its influence diminishes as students progress to higher grades. Gender has often been explored as a variable in the development of Higher-Order Thinking Skills (HOTS). In contrast, teaching strategies have a profound impact on the development of HOTS. A meta-analysis by Hattie (2009) showed that effective teaching strategies, such as feedback, interactive teaching, and cooperative learning, have a high effect size on student learning outcomes, including HOTS development. Hattie emphasizes that these instructional approaches account for a larger variance in student achievement compared to demographic factors like gender.

Discussion of differences ethnic composition students in the performance of higher-order thinking skills (HOTS)

From the statistical results, although the mean ranking of the Malay-Chinese group is slightly higher, the effect of ethnic background on the measured indicators is not statistically significant. This may be related to the small sample size (especially the Malay-Chinese group with only 5 people), which limits the sensitivity of the statistical test. In addition, cultural differences or uneven distribution of educational resources may show significant effects in larger samples or more refined indicators. Therefore, future research can expand the sample size and further refine the variables, such as exploring the performance of students of different races in specific learning modules. In 2024, the total population of Malaysia is estimated to be 34.1 million, mainly composed of Malays, Chinese, Indians, and several indigenous ethnic groups (Department of Statistics Malaysia, 2024). According to the 2010 census data, among the citizens of the country, indigenous people (Malays and indigenous peoples) accounted for 67.4%, Chinese accounted for 24.6%, Indians accounted for 7.3%, and other ethnic groups accounted for 0.7% (Department of Statistics Malaysia, 2010; Lim, 2013).

However, in the sample of this study, the proportion of Malay and Indian students was equal, which may be due to the fact that the mathematics teacher of the selected class was Indian. Despite the larger proportion of Indian students, the statistical results show that the average ranking of Indian students is still lower than that of other groups. This may reflect that factors other than ethnic background (such as differences in teacher teaching style, family support, language, or cultural background) have a more significant impact on student performance. Future research needs to further explore these potential variables and reveal the relationship between complex educational backgrounds and student performance through diversified research designs.

Discussion of Implementation Expansion

The senior grade mathematics curriculum in Malaysia's KSSM gradually increases the difficulty in space, and it is necessary to help students build the ability to solve spatial problems in primary school. In addition, secondary school students will gradually shift their studies from Euclidean geometry to solid geometry. The van Hiele theory, proposed by Dutch mathematics educator Hans van Hiele, focuses on the developmental stages of geometric thinking (Crowley, 1987). The theory holds that students go through different levels of thought, following instruction divided into five learning stages, from intuitive perception to more abstract reasoning, which is closely related to applying, analyzing, evaluating, and creating in HOTS. In educational practice, teachers can use van Hiele theory to assess students' current stage in mathematical thinking.

In addition to selecting a specific chapter from the textbook, the assessment of spatial skills can also be carefully constructed around dimensions. Building on Clements' (2003) foundational questions for geometric concept development, spatial skills are delineated into four dimensions: recognition of graphics, assessing the ability to identify and classify shapes; movement of graphics, examining spatial transformations such as translation and rotation; orientation of graphics, evaluating the understanding of direction and position in space; and measurement of graphics, focusing on quantifying attributes such as length and area. These interconnected dimensions form a comprehensive framework for assessing spatial reasoning and geometric understanding, enabling educators to systematically enhance curricula and evaluation tools. Incorporating differentiated instructional (DI) applications into these dimensions can further enhance students’ learning experiences by offering interactive tools to practice spatial transformations and measurements, thus fostering a deeper understanding of geometric concepts.

In the field of education, although many methods exist for assessing geometric ability, methods for effectively developing these skills are lacking. Research, such as the work of Abosalem (2016) and Sofyan et al. (2024), aims to cultivate teachers' ability to assess HOTS but ignores how to apply these methods in practice to cultivate students' HOTS. The proposed instrument aims to meet interdisciplinary needs, supporting both individual exploration and group collaboration. While grounded in mathematics, the tool's flexible design enables application across different disciplines, facilitating research beyond specific subjects and offering dedisciplinary insights into problem-solving dynamics and student thought processes.

CONCLUSION

The development of higher-order thinking skills (HOTS) is a cornerstone of educational goals, but there is little research on how to effectively develop these skills in students. Our study fills this gap by using the Polya problem-solving strategy as an instrument to assess the HOTS of 23 students in Malaysia. It not only demonstrates operations through mathematical content but also expresses how the instrument can be used across disciplines. It can transcend disciplinary boundaries and be used as a multifunctional tool in various research fields. Next, students' responses were carefully reviewed to spatial tasks in the textbook via detailed scoring criteria, which also provides a replicable framework for future spatial reasoning research. The results painted a picture of varying skills. The students needed more support in the design and execution of problem solving, which was a clear challenge. Moreover, the analysis revealed that only two students fully accepted the HOTS criteria, which highlights the urgency of the differentiated instruction (DI) strategy.Finally, the researchers used independent sample T tests and found that gender had no significant effect on HOTS scores in terms of spatial quality, and more emphasis should be placed on improving cognitive and problem-solving skills overall. Several limitations should be acknowledged. The sample size of this study was relatively small, focusing on a group of 23 fifth-grade students from different ethnic backgrounds. The limited sample size and lack of significant differences may limit the generalizability of the findings to the wider population. As reflected in the study, the Malay-Chinese group had the highest average ranking (14.00), followed by the Malay group (12.39) and the Indian group (10.50), and it is possible that the performance differences between different ethnic groups were not statistically significant. Additionally, the study explored short-term outcomes of Polya’s model without examining the long-term effects on student performance. The gender analysis was limited in scope, relying on observed patterns rather than extensive statistical testing. Future studies should consider larger, more diverse samples across different school settings to validate the findings. Longitudinal research is also recommended to explore the long-term impact of Polya’s model and differentiated teaching strategies on HOTS development. Further investigations could examine the effectiveness of instructional tools in other mathematical domains or subjects. Additionally, more in-depth gender analysis is needed to develop tailored strategies that address the unique learning needs of male and female students. This study offers valuable insights for educators seeking to enhance HOTS in primary mathematics education. The findings highlight the importance of flexibility and support in teaching practices, encouraging teachers to adopt strategies to DI and ensure that students are appropriately challenged regardless of their abilities or learning preferences. Schools and policymakers should provide professional development opportunities for teachers to effectively implement the strategy and foster HOTS in their classrooms.

REFERENCES

Abosalem, Y. (2016). Assessment techniques and students’ higher-order thinking skills. *International Journal of Secondary Education*, *4*(1), 1-11. https://doi.org/10.11648/j.ijsedu.20160401.11

Aladwan, S. Q. A., Alfayez, M. Q. E., & Shaheen, H. R. A. (2023). The level of conceptual and procedural knowledge in mathematics and its relationship to creative thinking among gifted students. *International Journal of Instruction, 16*(4), 121–134. https://doi.org/10.29333/iji.2023.1648a

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. https://archive.org/details/taxonomyforlearn00unse

Atit, K., et al. (2020). Examining the role of spatial skills and mathematics motivation on middle school mathematics achievement. *International Journal of STEM Education*, 7. https://doi.org/10.1186/s40594-020-00234-3

Barak, M., & Shakhman, L. (2008). Fostering higher-order thinking in science: Teachers' reflections. *Teachers and Teaching: Theory and Practice, 14*(3), 191-208. https://doi.org/10.1080/13540600802006079

Bondie, R.S., Dahnke, C., & Zusho, A. (2019). How does changing “one-size-fits-all” to differentiated instruction affect teaching? *Review of Research in Education*, *43*(1), 336-362. https://doi.org/10.3102/0091732X18821130

Bushie, C. (2015). Literature review: Differentiation in education. BU Journal of Graduate Studies in Education, 7(2), 35–42. https://eric.ed.gov/?id=EJ1230703

Clements, D. H. (2003). Teaching and learning geometry. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), A research companion to principles and standards for school mathematics (pp. 151–178). *National Council of Teachers of Mathematics.* https://www.academia.edu/12978334/Teaching_and_learning_geometry

Crowley, M. L. (1987). The van Hiele model of the development of geometric thought. In M. M. Lindquist (Ed.), Learning and teaching geometry, K-12 (pp. 1–16). *National Council of Teachers of Mathematics.* https://eric.ed.gov/?id=ED303448

Department of Statistics Malaysia (2010). Population and Housing Census of Malaysia 2010 (Updated Report). https://www.dosm.gov.my

Department of Statistics Malaysia (2024). Current Population Estimates, Malaysia, 2024. https://www.dosm.gov.my

Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process. D.C. Heath and Company.

Fujita, T., et al. (2020). Spatial reasoning skills for 2D representations of 3D geometrical shapes in grades 4 to 9. *Mathematics Education Research Journal*, 32, 235-255. https://doi.org/10.1007/s13394-020-00335-w

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449-455. https://doi.org/10.1037/0003-066X.53.4.449

Hattie, J. (2009). Visible Learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge. https://doi.org/10.1007/s11159-011-9198-8 https://archive.org/details/howwethinkrestat00dewe

https://searchworks.stanford.edu/view/11428

Hu, L. (2024). Utilization of differentiated instruction in K-12 classrooms: a systematic literature review (2000–2022). *Asia Pacific Education Review*, 1-19. https://doi.org/10.1007/s12564-024-09931-y

Jäncke,L.(2018). Sex/gender differences in cognition, neurophysiology, and neuroanatomy. *F1000Research*, 7. https://doi.org/10.12688/f1000research.13917

Julià, C., & Antolí, J.O. (2018). Enhancing spatial ability and mechanical reasoning through a STEM course. *International Journal of Technology and Design Education*, 28, 957-983. https://doi.org/10.1007/s10798-017-9428-x

Kheloui, S., Jacmin-Park, S., Larocque, O., Kerr, P., Rossi, M., Cartier, L., & Juster, R. P. (2023). Sex/gender differences in cognitive abilities. *Neuroscience & Biobehavioral Reviews*, 105333. https://doi.org/10.1016/j.neubiorev.2023.10533

Kustati, M., & Sepriyanti, N. (2022). The effect of 21st-century learning on higher-order thingking skills (hots) and numerical literacy of science students in Indonesa based on gender. *Jurnal Pendidikan IPA Indonesia*, *11*(2). https://doi.org/10.15294/jpii.v11i2.36384

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice*, 32(3), 131-137. https://doi.org/10.1080/00405849309543588

Lim, L. (2013). Education and ethnic relations in Malaysia: Policy and practice. *Race Ethnicity and Education, 16*(3), 329–352. DOI: 10.1080/13613324.2011.618832

Ling, C. Y., Osman, S., Daud, M. F., & Hussin, W. N. W. (2019). Application of Vee Diagram as a Problem-Solving Strategy in Developing Students' Conceptual and Procedural Knowledge. In *International Journal of Innovative Technology and Exploring Engineering*, *8*(10), 2796–2800). https://doi.org/10.35940/ijitee.j9591.0881019

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6), 1479-1498. https://doi.org/10.2307/1130467

Liu, J., Liu, Z., Wang, C., Li, X., & Xu, Y. (2024). Key factors and mechanisms affecting higher-order thinking skills of primary and secondary school students in the smart classroom environment. *Current Psychology*, *43*(11), 9651-9664. https://doi.org/10.1007/s12144-023-05136-5

Lu, J., Wu, S., Wang, Y., & Zhang, Y. (2022). Visualizing the commognitive processes of collaborative problem solving in mathematics classrooms. *Asia-Pacific Education Research, 32*(4), 615 – 628. https://doi.org/10.1007/s40299-022-00681-2

Mulligan, J., et al. (2018). Connecting mathematics learning through spatial reasoning. *Mathematics Education Research Journal*, *30*, 77-87. https://doi.org/10.1007/s13394-017-0210-x

Mulyoto, Rugaiyah, & Susanto, T. T. J. (2024). Improving teacher creativity in teaching through career development. *Anatolian Journal of Education*, *9*(2), 1–16. https://doi.org/10.29333/aje.2024.921a

Paul, R., & Elder, L. (2016). Innovative approaches to teaching higher-order thinking skills. *Journal of Educational Strategies, 43*(3), 5–12. https://doi.org/10.1207/s15430421tip4303_5

Pavlovičová, G., et al. (2022). Spatial ability and geometric thinking of the students in teacher training for primary education. *TEM Journal*. https://doi.org/10.18421/TEM111-49

Polya, G. (1945). How to Solve It: A New Aspect of Mathematical Method. Princeton University Press. https://doi.org/10.1515/9781400828678

Qasrawi, R., & BeniAbdelrahman, A. (2020). The Higher and Lower-Order Thinking Skills (HOTS and LOTS) in Unlock English Textbooks (1st and 2nd Editions) Based on Bloom's Taxonomy: An Analysis Study. *International Online Journal of Education and Teaching, 7*(3), 744-758. https://iojet.org/index.php/IOJET/article/view/866

Rahayuningsih, S., & Jayanti, R. (2019). High order thinking skills (HOTS) students in solving group problem based gender. *Al-Jabar: Jurnal Pendidikan Matematika*, *10*(2), 243-250. https://doi.org/10.24042/ajpm.v10i2.4872

Ratna, A. M., & Retnawati, H. (2019). Content analysis of higher order thinking skills (HOTS) in mathematics national exam. Journal of Physics: Conference Series, 1320(1), 012067. https://doi.org/10.1088/1742-6596/1320/1/012067

Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PloS one*, *7*(7), e39904. https://doi.org/10.1371/journal.pone.0039904

Resnick, L.B. (1987). Education and learning to think. Washington, DC: *National Academy Press*. https://doi.org/10.17226/1032

Rohid, N., Suryaman, & Rusmawati, R. D. (2019). Students' Mathematical Communication Skills (MCS) in Solving Mathematics Problems: A Case in Indonesian Context. *Anatolian Journal of Education,* *4*(2), 19-30. https://doi.org/10.29333/aje.2019.423a

Smale-Jacobse, A.E., Meijer, A., Helms-Lorenz, M., & Maulana, R. (2019). Differentiated instruction in secondary education: A systematic review of research evidence. *Frontiers in psychology*, *10,* 472176. https://doi.org/10.3389/fpsyg.2019.02366

Smith, J. (2020). "The Role of Differentiated Instruction in Developing Higher-Order Thinking Skills." *Educational Psychology Review, 32*(2), 345-362. https://doi.org/10.1007/s10648-019-09459-9

Sofyan, F.A., Sartono, E., Badaruddin, K., Fauzi, M., Syarnubi, S., Oviyanti, F., ... & Sukirman, S. (2024, April). Analysis of Higher-Order Thinking Skill (HOTS) of Madrasah Ibtidaiyah students in solving open-ended mathematics problems. *In AIP Conference Proceedings* (Vol. 3058, No. 1). AIP Publishing. https://doi.org/10.1063/5.0201104

Tay, M.K., & Preciado-Babb, A.P. (2020). Fostering student teachers’ spatial reasoning: the robotics Mars challenge. *In Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. https://doi.org/10.51272/pmena.42.2020-289

Wang, L., Zhang, Q., & Sun, D. (2024). Exploring the Impact of An Augmented Reality-Integrated Mathematics Curriculum on Students’ Spatial Skills in Elementary School. *International Journal of Science and Mathematics Education*, 1-28. https://doi.org/10.1007/s10763-024-10473-3

Weiland, T., & Poling, L. L. (2022). Taking a spatial turn in mathematics teacher education. *Mathematics Teacher Educator, 11*(1), – 67. https://doi.org/10.5951/MTE.2021.0047

Widana, I. W. (2017). Higher order thinking skills assessment (HOTS). *JISAE: Journal of Indonesian Student Assessment and Evaluation*, *3*(1), 32-44. https://doi.org/10.21009/JISAE.031

Widyastuti, E., & Jusra, H. (2022). Mathematical critical thinking ability in solving HOTS problems based on cognitive style and gender. *Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, *10*(3), 535-545. https://doi.org/10.33394/j-ps.v10i3.5217

Ying, C.L., Osman, S., Kurniati, D., Masykuri, E.S., Kumar, J.A., Hanri, C. (2020). Difficulties that Students Face when Learning Algebraic Problem-Solving. *Universal Journal of Educational Research*, *8*(11), 5405 - 5413. DOI: 10.13189/ujer.2020.081143

Zhou, Y., Ning, Y., Chen, J., Zhang, W., & Wijaya, T.T. (2024). Development and validation of Mathematical Higher-Order Thinking Scale for high school students. *Psychology in the Schools*. https://doi.org/10.1002/pits.23213