Jurnal Pendidikan MIPA
e-ISSN: 2685-5488 | p-ISSN: 1411-2531
URL: https://jpmipa.fkip.unila.ac.id/index.php/jpmipa
Jounal's DOI: http://doi.org/10.23960/jpmipa

Vol. 27, No. 2, 2026
Article ID: 1372
Pages: 1129-1150

Mathematical Critical Thinking Ability of Dyscalculia Students: Analysis of HOTS Problem Solving

Nur Afifah, & Ayu Faradillah*
Department of Mathematics Education, Universitas Muhammadiyah Prof. DR. HAMKA, Indonesia

*Corresponding email: ayufaradillah@uhamka.ac.id

Received: 06 May 2026 *Accepted: 28 June 2026* *Published: 04 July 2026*

Abstract: Critical thinking (CT) skills in mathematics are essential for solving problems involving High Order Thinking Skills (HOTS), although students with dyscalculia often experience difficulties in this process. This study was conducted to analyze the mathematical CT skills of students with dyscalculia in solving HOTS problems by gender using a qualitative method. This study involved male and female students with dyscalculia. Data were collected through a dyscalculia identification instrument, a HOTS-based CT test, and interviews. The analysis of mathematical CT skills was aligned with four indicators: interpretation, analysis, evaluation, and inference. The findings show differences in students' mathematical CT skills based on gender. Female students demonstrated competence across all four indicators of CT. The results show V1 (dyscalculia female students) is capable of accurately interpreting problem statements, analyzing relevant information, evaluating appropriate strategies, and drawing correct conclusions. In contrast, V2 (dyscalculia male students) experiences difficulties, particularly in interpreting problems and determining appropriate solution strategies, which in turn affects their overall problem-solving ability. These results suggest that gender may play a role in how students with dyscalculia approach and solve mathematical problems. Therefore, mathematics instruction must consider the unique characteristics of students with dyscalculia as well as gender differences to better support the development of CT skills. The integration of HOTS-based learning activities tailored to students' needs is highly recommended to create a more inclusive and effective learning environment.

Keywords: mathematical critical thinking, dyscalculia, problem solving, higher-order thinking skills, gender.

Article's DOI: https://doi.org/10.23960/jpmipa.v27i2.pp1129-1150

INTRODUCTION

Strengthening mathematical critical thinking (CT) skills can be done by providing Higher Order Thinking Skills (HOTS) questions (Diana et al., 2025; Mitani, 2021). HOTS questions are designed to encourage students to engage in CT processes. Solving mathematical HOTS questions requires students to identify relevant information, analyze the relationships between mathematical concepts, and determine the appropriate solution strategy (May et al., 2020; Suseelan et al., 2023). This process can help students understand the thinking steps used in obtaining solutions.

In fact, students have difficulty understanding math problems that require CT skills, so they are unable to determine the appropriate solution strategy (Faradillah & Fadhilah, 2021; Khotimah et al., 2025; Martins & Martinho, 2021; Nurmaliza et al., 2022; Wahab A et al., 2024). Furthermore, students tend to focus only on the teacher-provided example problems, thereby limiting their ability to solve HOTS problems. Additionally, students with dyscalculia also face difficulties in mathematics learning (Salisa & Meiliasari, 2023). Students with dyscalculia generally experience difficulties in understanding number concepts, processing numerical information, and performing mathematical operations, which can affect their ability to analyze and solve mathematical problems (Cárdenas et al., 2021; Grant et al., 2020).

Low mathematical CT skills are caused by several factors, such as teaching methods that remain teacher-centered, leading students to become accustomed to passively receiving information; a lack of practice with problems that require higher-order reasoning; students tend to rely on procedural solutions; and students' habit of memorizing formulas without deeply understanding the underlying concepts, with a lack of confidence when facing unfamiliar problems (Dotan & Ginat, 2022; Simanjuntak et al., 2021; Hartawan et al., 2026). Consequently, specific learning difficulties such as dyscalculia make it hard for students to understand numbers and process numerical information (Klein & Knops, 2023). Given these factors, mathematical CT skills can be strengthened by providing HOTS problems (Diana et al., 2025; Mitani, 2021).

External factors aside, internal factors also contribute to students' low mathematical CT skills. These internal factors include low motivation to learn mathematics, which leads students to be reluctant to delve deeper into the subject (Parsazadeh et al., 2021; Ye et al., 2022). Furthermore, a lack of self-efficacy in solving math problems causes students to give up easily when faced with non-routine CT problems (Muharani & Sucipto, 2023). Another internal factor is a tendency toward rigid thinking, where students struggle to change the approach they are accustomed to, even if it is ineffective for a specific problem (Kohar et al., 2024; Rahayuningsih et al., 2020). Anxiety mathematics also plays a significant role, as the fear and tension it causes can hinder students' analysis and evaluation processes, both of which are indicators of mathematical critical thinking skills (Gökçe & Güner, 2024; Kusmaharti et al., 2025).

In addition to general learning challenges, students with specific learning difficulties such as dyscalculia face greater obstacles in learning mathematics (Salisa & Meiliasari, 2023). Dyscalculia is characterized by difficulties in understanding number concepts, processing numerical information, and performing basic mathematical operations, all of which can hinder students' ability to analyze and solve mathematical problems (Cárdenas et al., 2021; Grant et al., 2020). These difficulties are closely linked to weak working memory, a core deficit in students with dyscalculia, which is below average compared to their peers (Mohammed et al., 2024). As a result, students struggle to recall procedural steps and retain intermediate calculation results, which prevents them from connecting information and drawing logical conclusions, especially when solving multi-step problems. These working memory limitations also affect students' critical thinking processes in constructing mental representations, connecting mathematical concepts, and selecting appropriate problem-solving strategies, all of which are key indicators of mathematical critical thinking ability (Saga et al., 2022). Consequently, previous research has consistently found that students with dyscalculia demonstrate lower problem-solving abilities compared to their peers (Fadilla et al., 2025; Salisa & Meiliasari, 2023)

Therefore, to illustrate this situation in a real-world context, preliminary empirical observations were conducted using HOTS-based problem-solving tasks. The observation results showed that students were able to create representations of the problems. However, these representations were still incomplete and lacked strong conceptual integration, even though arithmetic operations and the Pythagorean theorem had been applied. On the other hand, some students were unable to organize the solution process systematically, failed to identify relevant information, immediately moved on to computational operations, and did not provide a conclusion. These findings indicate varying levels of difficulty across various aspects of mathematical critical thinking, particularly regarding understanding, representation, analysis, and evaluation.

In terms of learning difficulties, individual student characteristics also influence how students solve math problems. Previous research has shown that factors such as thinking strategy tendencies, attention to detail, and psychological interest in mathematics can vary from one individual to another (Faradillah & Fadhilah, 2021; Stoet & Geary, 2020). Students have different characteristic thinking strategies in solving math problems (Schreiber & Ashkenazi, 2024; Stoet & Geary, 2020). These differences in potential affect how students analyze information, develop solution strategies, and determine the steps to take in solving mathematical problems (Onoshakpokaiye & Eyetan, 2026; Schreiber & Ashkenazi, 2024; Syaidi & Condro Murti, 2024).

Several previous studies have discussed CT skills in mathematics based on gender when solving HOTS questions (Diana et al., 2025; Harahap et al., 2025; May et al., 2020). In general, previous research has shown no difference in students' mathematical critical thinking skills, although there is a tendency toward differences in certain aspects. However, some studies have produced contradictory findings regarding gender differences in mathematical critical thinking skills. Male students demonstrate better critical thinking skills, particularly in terms of interpretation (Marni et al., 2020)

Numerous studies have examined mathematical critical thinking skills in relation to the ability to answer HOTS questions among the general student population. In addition to mathematical critical thinking skills, another relevant area of research concerns the cognitive abilities of students' dyscalculia. For example, students with dyscalculia struggle with mathematical problem-solving skills (Kohn et al., 2020). In addition, students with dyscalculia demonstrate below-average performance in numeracy skills (Peters et al., 2020; Reigosa-Crespo et al., 2020). Based on previous research, students with dyscalculia exhibit limitations in cognitive abilities.

However, research specifically examining how students with dyscalculia demonstrate their mathematical critical thinking skills and how they think when faced with HOTS questions remains limited. Therefore, the research question is: How do students with dyscalculia demonstrate their mathematical critical thinking skills when solving HOTS-based maths problems? Furthermore, how is a comparative case study conducted between male and female subjects when solving HOTS questions?

METHOD

Research Design

This study uses a qualitative approach to describe in detail how dyscalculic individuals think as they solve HOTS questions. According to Busetto et al. (2020), the qualitative method was chosen because it can reveal phenomena more deeply by interpreting data obtained from tests, questionnaires, and interviews with research subjects. Therefore, the phenomenological approach was used in this study. The phenomenological approach focuses on understanding individual experiences in dealing with a particular phenomenon (Alhazmi & Kaufmann, 2022). In this study, a phenomenological approach was used to explore the experiences of students with dyscalculia when solving HOTS questions in mathematics. Through this approach, researchers can observe how dyscalculia students think when understanding problems, analyzing information, and determining the solution strategies used (Alhazmi & Kaufmann, 2022).

A preliminary quantitative screening phase using the Rasch measurement model was integrated into the first stage of this sequential explanatory design (Deonovic et al., 2020;

Padgett & Morgan, 2020). This phase was not intended to test hypotheses or generalize findings to a broader population. Rather, it served as an objective psychometric mechanism to identify participants who consistently exhibited strong characteristics of mathematical dyscalculia before being included in the subsequent qualitative phase (Hope et al., 2025). The Rasch model was employed because it allows for direct comparisons between person ability and item difficulty on the same logit scale, thereby supporting more objective participant classification and enhancing measurement accuracy (Jacques et al., 2025; Papini et al., 2021; Putri et al., 2024). This approach also helps ensure that the selected participants truly represent the essential characteristics necessary for a valid phenomenological inquiry (Uto, 2021). Through this systematic procedure, students with high dyscalculia traits were objectively identified from a total population of 320 students across three schools. Once students with high levels of dyscalculia were objectively identified, this study proceeded to the second phase: a qualitative exploration to gain an in-depth understanding of how these students' CT skills operate when solving HOTS problems in mathematics learning.

Research Procedures

The research procedure was conducted in several stages. In the first stage, a dyscalculia screening instrument was distributed to 320 senior high school and vocational high school students. The collected data were analyzed using the Rasch model with Winsteps to identify participants who exhibited characteristics of dyscalculia across demographic categories, particularly gender and age. Prior to this stage, the researchers had obtained permission from the school to carry out the research. To protect privacy, the researcher anonymized all student identities by assigning unique codes, ensuring that personal data and school information remained confidential throughout the research process. The selection of these two subjects was based on several considerations. Subjects were categorized as having the highest level of dyscalculia based on the Rasch model analysis and the Wrightmaps results. In addition, the selected subjects have comparable abilities and communication skills.

Therefore, two participants were purposively selected to represent male and female students based on the dyscalculia screening results. The selected participants completed HOTS-based mathematical CT tasks and subsequently participated in semi-structured interviews to explore their cognitive processes in solving mathematical problems. Final stage: data from the written test and interviews were analyzed phenomenologically through triangulation and data collection techniques, using indicators of mathematical CT skills to obtain comprehensive and credible findings.

Participants

Students' dyscalculia was categorized according to their demographic data, namely gender and age, as follows.

Table 1. demographics dyscalculia subjects

School	Gender		Age	
	Male	Female	$14 \leq x < 16$	$16 \leq x \leq 18$
F1	44	55	-	99
F2	31	28	-	59
F3	89	73	57	105
Total	164	156	320	

Based on Table 1, participants in this study are from different schools. In addition, there are Male students account for 51.25% of the total number of subjects. Students aged 16 d" x d" 18 had the largest number of participants, with a difference of 57 compared to those aged 14 d" x d" 16.

Selection of two subjects was carried out using Wrightmaps (Sukarelawan et al., 2022). In classification using Wrightmaps, center point M (Mean), which is placed at a score of zero, while the separation distances are determined by S (Standard Deviation) at one unit to the left and right of M. It is at the positions (M - S) and (M + S) that T (Thresholds) is placed, namely two dividing lines that divide the map into three ability regions. The left T at (M - S) functions as the upper limit of the low category and the lower limit of the medium category. The right T at (M + S) functions as the upper limit of the medium category and the lower limit of the high category. Therefore, data were analyzed using Wrightmaps to assess the students' level of dyscalculia, with the results presented below.

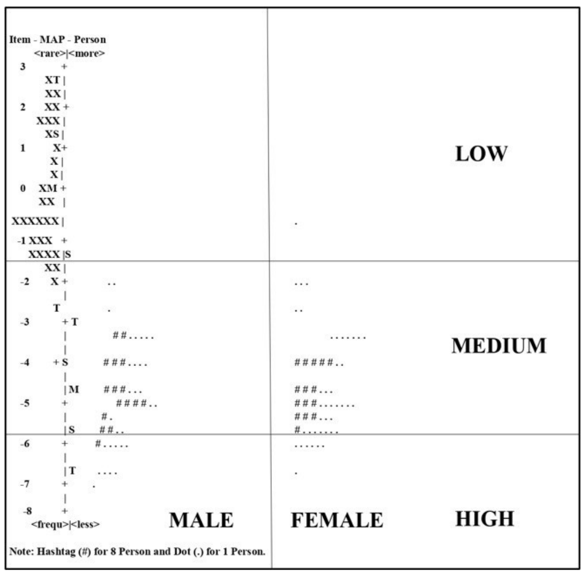


Figure 1. Wrightmaps of dyscalculia levels

Results in Figure 1 and Figure 2 show the distribution of dyscalculia levels among the 320 participants. The results indicate that the largest number of students with dyscalculia are at the moderate level. At the intermediate level, there are more female students than male students, by 14. However, at the high level, male students outnumber female students.

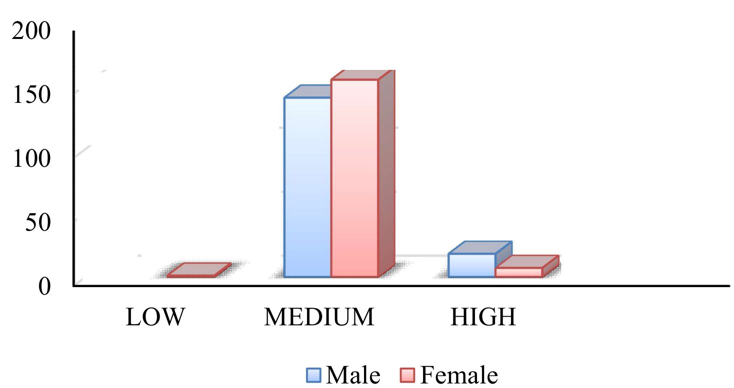


Figure 2. Bar chart visualization of students' ability levels in dyscalculia

Based on the Wrightmaps results, one male student and one female student were selected. Selection of these extreme cases aligns with a phenomenological approach that prioritizes depth of exploration over the number of participants. Focusing on a small number of information-rich participants allows researchers to explore cognitive experiences in a more in-depth, detailed, and nuanced manner without compromising the intensity of the analysis (Bartholomew et al., 2021; Hennink & Kaiser, 2022). In this context, the goal is not statistical representativeness or generalization to a broader population of students with dyscalculia, but rather an in-depth exploration of the lived experiences of students who consistently exhibit the strongest characteristics of dyscalculia (Cheetham, 2024). Furthermore, according to the Qualitative Framework for Operationalizing Respondent Sampling (Q-FORS), sample size determination in qualitative research should be based on data adequacy to answer the research questions, not merely on the number of participants (Braun & Clarke, 2021; Hennink & Kaiser, 2022). The number of participants in phenomenological research generally ranges from 3 to 15 people, but what is most important is that all participants genuinely experience the phenomenon under study (Tomaszewski et al., 2020).

Based on teachers' recommendations and the initial screening process, both subjects demonstrated strong communicative openness, including high verbal expression skills and a willingness to engage in deep thinking when solving problems, which supported the phenomenological interview process. Although there was an age difference, both subjects had covered equivalent mathematics material on the topics under study and were interviewed using a semi-structured interview with core questions that were equivalent for both participants. Therefore, variations in responses were attributed more to differences in individual thinking patterns than to age or socioeconomic background.

Data collection

Data were collected using three instruments: a dyscalculia identification instrument, HOTS-based mathematical CT tasks, and semi-structured interviews. The mathematics CT instrument was adapted from Afifah et al. (2026), while the HOTS questions were developed based on four CT indicators: interpretation, analysis, evaluation, and inference (Karim, 2015). To ensure the validity of the instruments in this study, several validation procedures were implemented.

This study employed triangulation of data collection techniques using three data sources, namely (1) dyscalculia screening using the Rasch model; (2) HOTS-based mathematical critical thinking written tests; and (3) semi-structured interviews. In accordance with the phenomenological approach, data analysis followed the Interpretative Phenomenological Analysis (IPA) procedure, which includes repeated reading, theme mapping, identifying patterns of relationships between themes, and writing an integrative narrative (Love et al., 2020). Triangulation aims to enhance validity, provide a deeper understanding, and explore interpretations of the problem (Campbell et al., 2018; Farquhar et al., 2020). To ensure the reliability of the data used in this triangulation, the dyscalculia instrument was first validated through content and construct validity procedures.

The dyscalculia identification instrument consists of 12 items aligned with the study's indicators and was administered to 320 high school and vocational high school students. Prior to implementation, the instrument underwent content validity and construct validity procedures. Content validity was evaluated by mathematics education experts to ensure consistency between the instrument items and the measured indicators (Mokkink et al., 2025). Content validation was reviewed by two mathematics education lecturers and a high school mathematics teacher. Results indicate that revisions are needed, namely simplifying questions such as "The result of

$\sqrt[4]{845} - \sqrt[3]{180} + \sqrt[2]{500}$ is…" to ". The result of $\sqrt[4]{18} - \sqrt[3]{8} + \sqrt[2]{32}$ is…" Second, revise the multiple-choice answer options so that each option aligns with the numbers in the question, particularly when determining the smallest value, to avoid confusing students. Next, construct validity was evaluated empirically through respondent testing to assess the appropriateness of the instrument items (Indu et al., 2025). The test data were analyzed using Winsteps based on the Rasch model.

Data on dyscalculia identification were analyzed using the Rasch measurement model by Winsteps software (Bilhaki & Faradillah, 2025). Rasch analysis was used to evaluate respondent consistency and item quality through fit statistics, reliability, discrimination indices, and Wrightmaps interpretations. Additionally, the Rasch model facilitates comparisons between individual ability and item difficulty levels on the same measurement scale, thereby helping researchers classify participants more objectively (Papini et al., 2021; Putri et al., 2024; Uto, 2021)

Winsteps is used to process Rasch-based data and to conduct an in-depth evaluation of instrument quality using summary statistics (Faradillah & Septiana, 2022; Suherman & Vidákovich, 2022). Validity analysis refers to the Outfit Mean Square (MNSQ), Outfit ZSTD, and Point Measure Correlation (PT Measure Corr). The results of the analysis of non-conforming items are presented in Table 2.

Table 2. Misfit order items dyscalculia instrument

Item	Outfit MNSQ (0.5 > x >1.5)	Outfit ZSTD (-2.0 > x > + 2.0)	PT. MEASURE CORR (0.0 > x > 0.4)
5	0.9	0.11	0.26
9	0.2	1.25	0.30
12	1.0	1.22	0.30
1	1.3	1.19	0.31
10	0.7	1.18	0.30
3	1.5	1.18	0.33
7	0.3	1.18	0.28
4	0.0	1.17	0.28
11	0.14	1.14	0.30

Based on the validity test results shown in Table 2, several items had Outfit MNSQ values exceeding the recommended criteria, including one item with an MNSQ of 0.0. Although standard Rasch practice is to remove misfitting items to maintain unidimensionality, these items were retained for strong substantive reasons (Cris, 2020; Köhler et al., 2020). In Rasch analysis, a very low MNSQ value indicates overfit, implying that the response pattern is overly predictable and often caused by redundancy or item dependency (Dabaghi et al., 2020). However, items with overfit are not always removed if their difficulty level is not measured by other items in the same domain (Kim & Oh, 2020). Retaining non-conforming items does introduce a degree of measurement error. To mitigate this, we followed the recommendations of Fischer et al. (2021) for scenarios with a limited item pool, by carefully examining the extent of non-conformity and ensuring that its impact on the individual's overall measurement is not detrimental. In addition to a validity test, a reliability test is also needed to help researchers assess data accurately accountably (Flake & Fried, 2020; Mokkink et al., 2020). The reliability test results are shown in Table 3.

Table 3. Reliability result of dyscalculia instrument

Statistic	Person	Item
(KR-20)	0.72	-
Reliability	0.75	0.99
Separation	1.71	9.47

The reliability test in Table 3 using KR-20 yielded a value of 0.72, indicating that the instrument is reliable because it exceeds the threshold of 0.70 (Lena & Nikolov, 2025). However, the Person Separation value of 1.71 falls below the ideal threshold of 2.0 generally recommended in Rasch analysis. Nevertheless, this instrument is still considered suitable for use in this study based on several methodological considerations (Fridberg et al., 2020). This study is descriptive in nature and aims to describe the general distribution of dyscalculia levels, not to perform precise classification or diagnosis of individuals (Park, 2019; Uto, 2021). Furthermore, the study sample comes from a population that is relatively homogeneous in terms of educational level and age, which naturally reduces the variance in respondents' abilities and results in a low separation index (Fridberg et al., 2020; Kim, 2020). Thus, although the separation index does not reach the ideal threshold, the instrument remains valid and methodologically sound for addressing the research questions in this descriptive study. Meanwhile, the critical thinking instrument used in this study was adapted from an article on the same topic, and its validity and reliability have already been tested.

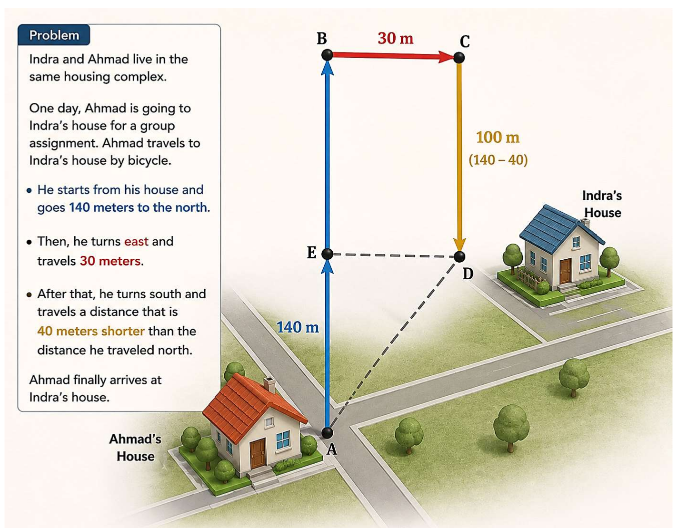


Figure 3. HOTS questions mathematical critical thinking skills

Figure 3 shows the HOTS questions given to the subjects. The questions cover indicators of mathematical CT skills consisting of (1) interpretation; (2) analysis; (3) evaluation; and (4) inference.

Data Analysis

This study employed qualitative data analysis procedures to examine the mathematical critical-thinking skills of students with dyscalculia when solving HOTS problems. The analysis focused on data obtained from written tests assessing mathematical critical-thinking skills aligned with HOTS and from semi-structured interviews. Furthermore, the analysis procedure consists of three components: (1) dyscalculia screening using the Rasch model; (2) HOTS-based mathematical critical thinking written tests; and (3) semi-structured interviews.

First, the researchers transcribed the interview data verbatim and repeatedly read the transcripts alongside the participants' responses on the HOTS assessment to understand the data. Second, key statements were identified and annotated with descriptive, linguistic, and conceptual comments. Third, these statements were analyzed to identify common patterns in the subjects' experiences. Finally, the findings were interpreted to capture the essence of the participants' life experiences, with triangulation of interview and HOTS assessment data.

RESULT AND DISCUSSION

Results of students with dyscalculia on the mathematical critical thinking (CT) skills assessment, which consists of four, namely (1) interpretation; (2) analysis; (3) evaluation; and (4) inference, which were adapted from the article (Karim, 2015). The subject codes selected are shown below.

Table 4 shows the codes of the selected students with dyscalculia and their scores. The subjects were selected to facilitate observation and gain an in-depth understanding of individual experiences; however, the results cannot be generalized to the entire population, as the description of the phenomenon focuses on the subjects' experiences. After both subjects took the HOTS assessment, which measures four indicators of critical thinking, an analysis of their written responses and interviews revealed two distinct cognitive patterns among students with math dyslexia. The results showed that V1 had the highest score, 4, on nearly all indicators across the two HOTS questions assessing mathematical critical thinking skills. Meanwhile, V2 scored 0 on the Inference indicator in the second question and achieved a maximum score of 2.

Table 4. Subjects codes and score

Number	Category and Demography	Code	HOTS scores of the mathematical critical thinking indicator			
			Interpretation	Analysis	Evaluation	Inference
1	Dyscalculia Female Seventeen years old	V1	Q1: 4 Q2: 4	Q1: 3 Q2: 4	Q1: 3 Q2: 4	Q1: 4 Q2: 4
2	Dyscalculia Male Fifteen years old	V2	Q1: 2 Q2: 2	Q1: 1 Q2: 2	Q1: 1 Q2: 2	Q1: 2 Q2: 0

Subjects' answers were then analyzed for their mathematical critical-thinking skills. Overall, the subjects demonstrated proficiency in the interpretation indicator, particularly in writing down the information given in the problem. Interpretation is a crucial first step in solving math problems, as at this stage, students are required to understand the information presented in the problem (Agusman et al., 2025). Interpretation skills serve as the foundation for progressing to the next indicators of critical thinking.

Next, regarding the analysis indicator subject demonstrated difficulty in writing mathematical models. Analysis indicators require the subject to identify relationships between concepts, break down the structure of the problem, and formulate these relationships into appropriate mathematical representations (Putra et al., 2025). Incorrect answers at this stage of the analysis indicate that the subject is unable to identify relevant concepts and to plan a systematic solution strategy. Consequently, on the evaluation indicators, the subject provides an inappropriate strategy. Evaluation indicators include the ability to evaluate the logical strength of inferential relationships and assess whether the chosen strategy is appropriate for solving the problem (Comstock & Grühn, 2026; Rodríguez-Rojas et al., 2024). Furthermore, such inadequate strategies can lead to errors in drawing conclusions. Inference indicators are often lacking due to weaknesses in previous stages, particularly in strategy formulation, which has a cascading effect on the quality of the resulting conclusions (Agusman et al., 2025; Shreeves et al., 2020). In addition, students exhibit different ability patterns by gender when solving HOTS questions.

V1 demonstrated systematic patterns across all four indicators. During the interpretation stage, V1 consistently noted both the given information and the question asked. During the analysis stage, he broke down the problem systematically. During the evaluation stage, he assessed the appropriateness of the chosen strategy. During the inference stage, he drew correct conclusions. In contrast, V2 exhibits a fragmented pattern. The most severe difficulties occur during the interpretation stage, when he writes down only partial information and fails to identify what the problem is asking for. Initial difficulties in interpretation result in an incomplete mathematical model during the analysis stage, an inability to evaluate strategies during the evaluation stage, and ultimately, no conclusion or an incorrect conclusion during the inference stage. These findings indicate that the most vulnerable indicator for students with dyscalculia in this study is interpretation; when it fails, the subsequent critical-thinking process collapses. Furthermore, to address the research question regarding how male and female students think when answering HOTS questions, this study analyzed the written responses of two participants. Both were students with a tendency toward dyscalculia, as identified through an initial assessment. To illustrate the differences in cognitive patterns between the two subjects more clearly, the following is the problem-solving process for HOTS V1 and V2 questions based on the four indicators of critical thinking.

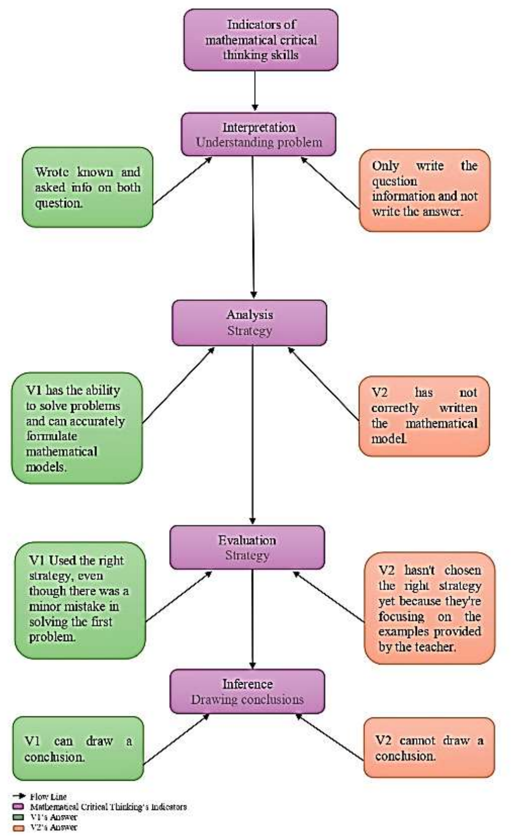
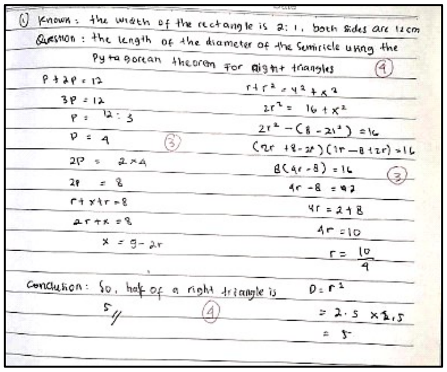
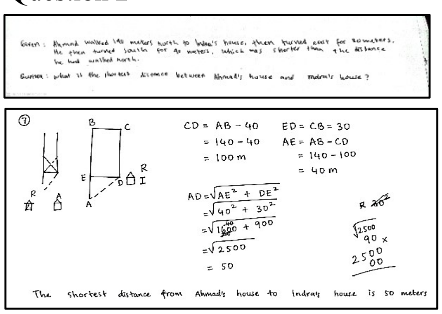
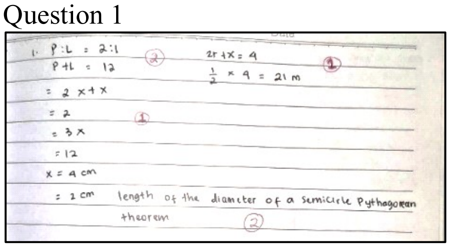


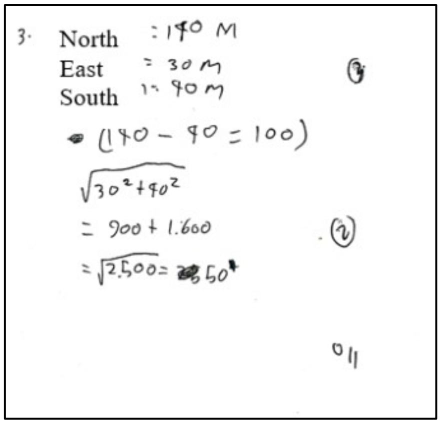
Figure 4. Flowchart of critical thinking dyscalculia

The flowchart in Figure 4 illustrates how both V1 and V2 exhibit starkly contrasting patterns of thinking when solving HOTS problems. V1 demonstrates a more systematic and coherent pattern on all indicators. In contrast, V2 exhibits a disjointed pattern from the very beginning on all indicators. These individual differences serve as a starting point for a more in-depth analysis of how each participant performs on each indicator of mathematical critical thinking skills. Therefore, to provide a more detailed understanding of differences in mathematical critical-thinking performance among participants, Table 6 presents the written answers and interview responses of each subject for further analysis.

The results in Table 5 indicate that the mathematical CT skills of dyscalculic students in solving HOTS questions still vary across

Table 5. Subjects answers

Subjects	Subjects Answer	Subject's interview responses (P: Code of reseacher, V1 and V2: code of subjects)
V1	Question 1 	V1 was able to write down the information provided in the problem and correctly apply the Pythagorean Theorem. However, there were still occasional doubts about which strategy to use because V1 stated. P: Do you find it difficult to analyze the information in the questions provided? V1: I had no trouble at all understanding the information in the question. However, I still sometimes get confused when deciding on a strategy for solving the problem. P: Do you know the material on the Pythagorean theorem? V1: I know the material on the Pythagorean Theorem, but I'm afraid of making a mistake because I've forgotten some of it.
V1	Question 2 	V1's success was supported by confidence and careful verification of answers, as evidenced during the interview. P: Do you feel confident when solving math problems? V1: I'm confident when doing math because, to me, math is like a puzzle. Moreover, during exams, I like to check my answers again. P: When you feel confident that you can answer a question, do you have trouble concluding? V1: No, because once I'm confident in my answer, it's not hard to draw conclusions.
V2	Question 1 	V2's incorrect strategy reflected difficulties in recalling and applying the required concepts, consistent with the interview findings. P: How did you feel while working on the HOTS math problem? V2: I felt scared. Q: In case, did you have any difficulty solving the problem, for example, analyzing the information provided in the problem or struggling with basic mathematical calculations? V2: I could still write down the given information, but I had more trouble formulating a strategy because if I didn't know the method, the basic calculations would automatically be wrong. Therefore, V2's error stemmed from difficulties in concept recall and strategy selection rather than basic calculation skills.

V2	Question 2 	Interview results confirmed that V2's difficulty stemmed from the challenge of determining a solution strategy for an unfamiliar problem. P: Do you find it difficult when you're given complex problems that are different from the examples provided by the teacher? V2: Yes, because they're not the same as the examples, so it's hard. It would be easier if the problems were identical to the examples provided, with only the numbers changed. V1. P: Do you also find it difficult to draw conclusions from the problems you're given? V2: Yes, it's quite difficult, because from the very beginning, it's hard to figure out a strategy for solving the problems, let alone when I'm asked to draw conclusions. V2 depends on the example; when the problem is different, V2 has difficulty determining strategies, procedures, and even drawing conclusions

indicators. To test the validity of the data, students' written answers were compared with their statements during interviews, and it was found that the doubts and limitations in their strategies, as expressed verbally, were also evident in their answer sheets. This validation of the data through comparison of test results and interview data is consistent with the triangulation of data collection techniques used in the study to obtain valid and consistent findings (Muhtarom et al., 2020; Agustin et al., 2024). Following the validation process, the results of the mathematical critical thinking tests and the interview data were synthesized to provide a comprehensive picture of each subject's performance. To facilitate comparisons across mathematical critical thinking indicators, the findings were presented using spider diagrams, which illustrate each subject's performance profile on each indicator.

Figure 5 presents a spider chart comparing mathematical critical thinking skills between subjects V1 and V2 across four indicators:

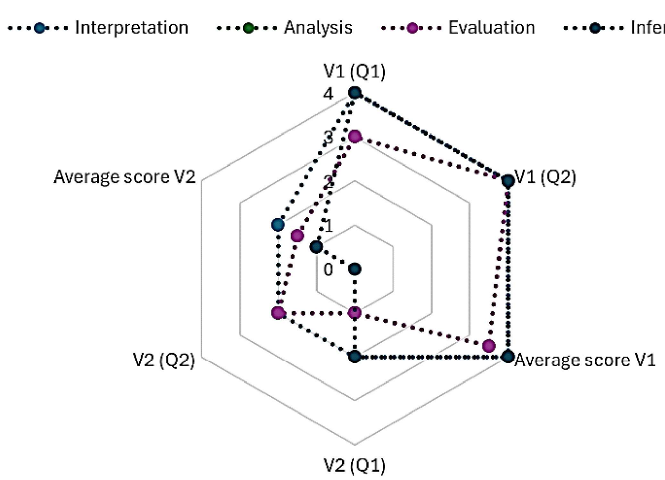


Figure 5. Spider chart comparing mathematical critical thinking indicators of subjects

interpretation, analysis, evaluation, and inference. Overall, the chart shows that subject V1 scored higher than subject V2 on each indicator for both the first question (Q1) and the second question (Q2). This is also reflected in V1's higher average score compared to V2's average score. These results indicate that V1 is better able to understand the information provided, analyze relationships between concepts, evaluate solution strategies, and draw accurate conclusions. In contrast, V2 still demonstrated relatively low performance across all four indicators, suggesting difficulty applying mathematical critical-thinking processes comprehensively. The differences in performance between the two subjects, visualized through a spider chart, provide an initial overview of the variations in mathematical critical thinking abilities, which will be discussed in greater depth for each indicator.

Regarding the first indicator, interpretation V2 had difficulty understanding the given problems. These difficulties indicate that the process of understanding the information in the questions is quite challenging for dyscalculic students (Pongsakdi et al., 2020; Salisa & Meiliasari, 2023). Therefore, it causes students' inability to identify what is known and what is asked in the questions completely (Agusfianuddin et al., 2024; Ario et al., 2025; Pongsakdi et al., 2020; Strohmaier et al., 2025). In addition, the initial process in mathematical CT is the ability to understand and interpret the information contained in the problem, so that if students are unable to interpret, they will have difficulty continuing the thinking process in the next stage (Schukajlow et al., 2023). Difficulties experienced by V2 during the interpretation stage appear to be caused by a combination of factors characteristic of dyscalculia, a high cognitive load involved in reading and processing complex word problems, difficulty in extracting numerical information embedded within sentences, and limited working memory capacity, which hinders the ability to recall the information provided whilst identifying what is being asked. These cognitive constraints collectively prevented V2 from constructing a complete representation of the problem and attempting a solution.

Second indicator: analysis and evaluation V2 showed efforts to construct mathematical models; however, the problem-solving steps remained unsystematic. For the first question, V2 was unable to answer correctly because the strategy employed was inappropriate. For the second question, V2 demonstrated some ability to respond even though the problem-solving strategy was incomplete; the final answer was nonetheless correct. This indicates that V2 is not yet able to connect mathematical concepts with appropriate solution strategies. V2 stated, "When I am given difficult questions, I can never find the answers. The questions given are much more difficult than the examples given by the teacher". Therefore, students who tend to rely heavily on teacher-provided examples will experience difficulties when faced with HOTS mathematics questions (Chirove & Ogbonnaya, 2021; Roorda et al., 2024). This condition indicates that V2's analytical abilities were still limited, which in turn affected the inference stage and resulted in less accurate conclusions (Ricco et al., 2020; van Dijke-Droogers et al., 2022).

Regarding the third indicator, inference, V2 was only able to formulate a conclusion for the second question, and even that conclusion was not fully accurate. Students' inability to draw conclusions is often due to inaccuracies in identifying and applying problem-solving strategies (Bara et al., 2025; Powell et al., 2020). This was further confirmed during the interview, in which V2 expressed a general dislike of mathematics learning. V2 also stated that when working on mathematics problems, V2 found it difficult to determine the appropriate strategy for answering and tended to rely solely on example problems provided by the teacher. When faced with more complex problems requiring different solution steps, V2 felt unable to provide an answer. This indicates that V2's difficulties during the inference stage stem not only from limited strategic flexibility but also from low motivation and negative affect toward mathematics, both of which are known to impair higher-order thinking processes. Meanwhile, V1 was able to identify the information contained in the questions and organize the solution steps more systematically. This ability helped V1 develop effective problem-solving strategies (Agusman et al., 2025; Faradillah & Humaira, 2021; Rubenstein et al., 2020). This is proven through female students' answers that demonstrate accurate analytical skills, enabling them to evaluate the problem-solving process (Akpur, 2020; Reinhold et al., 2020). In addition, V1 stated that "mathematics is like a puzzle for me, even though I sometimes have difficulty solving problems." Therefore, it can be concluded that the involvement of students' CT skills will help them understand and solve complex problems (Bara et al., 2025; Sumarna et al., 2024). However, some studies have found that male students demonstrate strengths in identifying logical patterns when solving problems (Onoshakpokaiye & Eyetan, 2026; Ramírez-Uclés & Ramírez-Uclés, 2020)

A difference in mathematical CT skills was observed between V1 and V2, the two dyscalculic students who participated in this study, in solving HOTS questions. V1 demonstrated better performance than V2 across all CT indicators. However, it is important to note that this study involved only two participants, and caution is warranted when interpreting these differences. The observed gap may reflect not only gender differences but also other confounding variables, such as differences in the severity of dyscalculia, individual cognitive profiles, or motivation levels. Therefore, the performance difference between V1 and V2 should not be attributed solely to gender. This indicates that mathematics learning needs to be designed by taking into account the characteristics of students, especially dyscalculic students, as well as providing exercises that emphasize the process of mathematical CT skills and providing guidance in understanding problems and developing appropriate solution strategies (Azhari et al., 2024; Benavides-Varela et al., 2020; Fadilla et al., 2025). Further research is recommended to examine the mathematical CT skills of dyscalculic students using a larger sample size when solving HOTS questions.

CONCLUSION

This study analyzes the mathematical CT abilities of students with dyscalculia when solving HOTS questions by gender. The findings indicate that the CT abilities of students with dyscalculia differ across the indicators of interpretation, analysis, evaluation, and inference. V1 demonstrated better performance in identifying relevant information, organizing systematic problem-solving steps, evaluating solution strategies, and drawing appropriate conclusions. On the other hand, V2 had difficulty understanding the information in the questions at first, so they were less accurate in determining strategies for solving the problems. These differences indicate that students with dyscalculia still face challenges in processing mathematical information when solving problems that require higher-order thinking skills. Results of this study imply that mathematics learning should be designed with consideration of the characteristics of dyscalculia students and gender, providing learning activities that focus on developing mathematical CT skills through HOTS-based exercises. One recommended approach is Problem-Based Learning (PBL) with visual scaffolding. This model helps students understand problems through visual representations such as diagrams and flowcharts, plan strategies systematically, and evaluate the solutions obtained. This approach is effective in reducing working memory load and improving critical thinking skills of students with dyscalculia, especially in the indicators of analysis, evaluation, and inference. However, this study is limited by the small sample size, so the findings cannot be generalized to a broader population. Further research is recommended to examine dyscalculia across other abilities and to expand the study to other levels.

■ DECLARATION OF GENERATIVE AI USAGE IN THE WRITING PROCESS

While writing this manuscript, the author used DeepL Translate to assist with translation and refine the text. The author has reviewed and edited the content. Therefore, the author takes full responsibility for the content of the published article.

■ REFERENCES

Afifah, N., Pebrianti, T., Faradillah, A., & Ulfah, S. (2026). Development and validation of mathematical critical thinking test using rasch measurement model. *International Journal of Evaluation and Research in Education*. https://doi.org/10.33365/jm.v8i1.875

Agusfianuddin, A., Herman, T., & Turmudi, T. (2024). Difficulties in mathematical language and representation among elementary school students when solving word problems. *Jurnal Elemen*, *10*(3), 567–581.

Agusman, A., Purwanto, P., & Rahardi, R. (2025a). An exploration of critical thinking stages of junior high school students in solving contradictory mathematical problems. *JRAMathEdu (Journal of Research and Advances in Mathematics Education), 10*(3), 200–217.

Agustin, N. D., Sumantri, M. S., & Marini, A. (2024). How can we better comprehend the disposition of elementary school students towards advanced mathematical thinking?. *Journal of Education Research and Evaluation*, *8*(2), 231–239.

Akpur, U. (2020). Critical, reflective, creative thinking and their reflections on academic achievement. *Thinking Skills and Creativity*, *37*, 100683.

Alhazmi, A. A., & Kaufmann, A. (2022). Phenomenological qualitative methods applied to the analysis of cross-cultural experience in novel educational social contexts. *Frontiers in psychology*, *13*, 785134.

Ario, M., Suhendra, Jupri, A., & Nurlaelah, E. (2025). Students' errors and learning obstacles in solving algebraic word problems: hermeneutic phenomenology . *Education Sciences*, *15*(12), 1674.

Azhari, B., Johar, R., Ramadhani, E., Mailizar, M., & Safrina, K. (2024). Mathematics learning model for children with dyscalculia through special intervention. *Jurnal Ilmiah Peuradeun*, *12*(3), 1155-1184.

Bara, E., Berdica, B., & Bara, G. (2025). The role of critical thinking in enhancing student learning and problem solving for 21st century challenges. *International Journal of Social Sciences & Humanities (IJSSH)*, *10*(1), 37-48.

Bartholomew, T. T., Joy, E. E., Kang, E., & Brown, J. (2021). A choir or cacophony? Sample sizes and quality of conveying participants' voices in phenomenological research. *Methodological Innovations* , *14*(2), 20597991211040063.

Benavides-Varela, S., Callegher, C. Z., Fagiolini, B., Leo, I., Altoè, G., & Lucangeli, D. (2020). Effectiveness of digital-based interventions for children with mathematical learning difficulties: A meta-analysis. *Computers & Education*, *157*, 103953.

Bilhaki, R. R., & Faradillah, A. (2025). Mathematical anxiety and resilience as an influence of mathematical literacy skills. *Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, *13*(3), 430–448.

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative research in sport, exercise and health*, *13*(2), 201–216.

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and practice*, *2*(1), 14.

Campbell, R., Goodman-Williams, R., Feeney, H., & Fehler-Cabral, G. (2020). Assessing triangulation across methodologies, methods, and stakeholder groups: The joys, woes, and politics of interpreting convergent and divergent data. *American Journal of Evaluation*, *41*(1), 125–144.

Cárdenas, S. Y., Silva-Pereyra, J., Prieto-Corona, B., Castro-Chavira, S. A., & Fernández, T. (2021). Arithmetic processing in children with dyscalculia: an event-related potential study. *PeerJ*, *9*, e10489.

Cheetham, L. J. (2024). “Because people don’t know what it is, they don’t really know it exists”: a qualitative study of postgraduate medical educators’ perceptions of dyscalculia. *BMC Medical Education*, *24*(1), 896.

Chirove, M., & Ogbonnaya, U. I. (2021). The relationship between grade 11 learners’ procedural and conceptual knowledge of algebra.

Comstock, G., & Grühn, D. (2026). Assessing critical thinking about argumentative essays: the Critical Reasoning and Inference Test (CRIT). *Assessment & Evaluation in Higher Education*, 1–14.

Criºan, D. R., Tendeiro, J. N., & Meijer, R. R. (2020). On the practical consequences of misfit in Mokken scaling. *Applied Psychological Measurement*, *44*(6), 482–496.

Dabaghi, S., Esmaielzadeh, F., & Rohani, C. (2020). Application of Rasch analysis for development and psychometric properties of adolescents’ quality of life instruments: A systematic review. *Adolescent health, medicine and therapeutics*, 173–197.

Deonovic, B., Bolsinova, M., Bechger, T., & Maris, G. (2020). A Rasch model and rating system for continuous responses collected in large-scale learning systems. *Frontiers in psychology*, *11*, 500039.

Diana, N., Agustinasari, A., Khatimah, H., Ibnusaputra, M., Adiansha, A. A., & Muslim, M. (2025). HOTS geometry analysis on critical thinking and numeracy skills reviewed learning style and gender. *Edumatica: Jurnal Pendidikan Matematika*, *15*(2), 245–264.

Dotan, D., & Zviran-Ginat, S. (2022). Elementary math in elementary school: The effect of interference on learning the multiplication table. *Cognitive Research: Principles and Implications*, *7*(1), 101.

Draucker, C. B., Rawl, S. M., Vode, E., & Carter-Harris, L. (2020). Integration through connecting in explanatory sequential mixed method studies. *Western Journal of Nursing Research*, *42*(12), 1137–1147.

Fadilla, A. R., Kamid, K., & Rohati, R. (2025). Analyzing the thinking trajectory of students with dyscalculia in solving spatial mathematical problems. *Journal of General Education and Humanities* , *4*(2), 589–604.

Faiziyah, N., & Priyambodho, B. L. (2022). *Analisis kemampuan berpikir kritis dalam menyelesaikan soal HOTS ditinjau dari metakognisi siswa* [Analysis of critical thinking skills in solving HOTS questions viewed from the perspective of student metacognition]. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(4), 2823-2835.

Faradillah, A., & Fadhilah, Y. H. R. (2021). Mathematical problem-solving on slow learners based on their mathematical resilience. *Jurnal Elemen*, *7*(2), 351–365.

Faradillah, A., & Humaira, T. (2021). Students' Critical thinking skills on algebraic problems through mathematical resilience. *JPI (Jurnal Pendidikan Indonesia)*, *11*(2), 317–327.

Faradillah, A., & Septiana, C. (2022). Mathematical resilience: validity and reliability with rasch model and SPSS in senior high school. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(4), 3545–3555.

Farquhar, J., Michels, N., & Robson, J. (2020). Triangulation in industrial qualitative case study research: Widening the scope. *Industrial Marketing Management*, *87*, 160–170.

Fischer, L., Rohm, T., Carstensen, C. H., & Gnambs, T. (2021). Linking of Rasch-scaled tests: Consequences of limited item pools and model misfit. *Frontiers in psychology*, *12*, 633896.

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in methods and practices in psychological science*, *3*(4), 456–465.

Fridberg, H., Wallin, L., Wallengren, C., Kottorp, A., Forsman, H., & Tistad, M. (2020). Development and evaluation of the measurement properties of a generic questionnaire measuring patient perceptions of person-centred care. *BMC Health Services Research*, *20*(1), 960.

Gökçe, S., & Güner, P. (2024). Pathways from cognitive flexibility to academic achievement: mediating roles of critical thinking disposition and mathematics anxiety. *Current Psychology*, *43*(20), 18192–18206.

Grant, J. G., Siegel, L. S., & D'Angiulli, A. (2020). From schools to scans: A neuroeducational approach to comorbid math and reading disabilities. *Frontiers in Public Health*, *8*, 469.

Harahap, P. A., Dewi, R. A., & Purniati, T. (2025). Analyzing the mathematical critical thinking skills of high school students in arithmetic: a gender differences. *Jurnal Pendidikan Progresif*, *15*(2), 1142–1156.

Hartawan, I. G. N. Y., Pujawan, I. G. N., & Wibawa, N. A. (2026). An exploration of teachers' perspectives on computational thinking in mathematics learning . *Indonesian Journal of Educational Development (IJED)*, *6*(4), 1173–1188.

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social science & medicine*, *292*, 114523.

Hope, D., Kluth, D., Homer, M., Dewar, A., Goddard-Fuller, R., Jaap, A., & Cameron, H. (2025). Exploring the use of Rasch modelling in "common content" items for multi-site and multi-year assessment. *Advances in Health Sciences Education*, *30*(2), 427–438.

Indu, P. V., Vidhukumar, K., Chacko, D., Menon, V., Grover, S., & Gupta, S. (2025). Criterion validity, construct validity, and factor analysis: An introductory overview. *Indian Journal of Psychiatry*, *67*(9), 916–921.

Jacques, Q., Donnan, J., Bishop, L., Howells, R., Gao, Z., & Najafizada, M. (2025). Development of a cannabis health literacy questionnaire: preliminary validation using the Rasch model. *BMC Public Health*, *25*(1), 2539.

Janis, I., Alias, M., Zulkipli, M., & Muhammad-Sukki, F. (2020). Using illustrations to make decisions on the most appropriate qualitative research methodology: The Industry 4.0 scenario. *International journal of qualitative methods*, *19*, 1609406920907247.

Karim, K., & Normaya, N. (2015). *Kemampuan berpikir kritis siswa dalam pembelajaran dalam pembelajaran matematika dengan menggunakan model Jucama di sekolah menengah pertama* [Students' critical thinking skills in learning mathematics using the Jucama model in junior high schools]. *EDU-MAT: Jurnal Pendidikan Matematika*, *3*(1), 92–104.

Khotimah, S. N., & Wulandari, T. S. H. (2025). Implementation of Problem Based Learning (PBL) Learning Model to Improve Mathematical Critical Thinking Skills and Learning Motivation of Grade 5 Students at UPT SD Negeri Kutorejo 3 Tuban. *Indonesian Journal of Education and Mathematical Science*, *6*(2), 85–90.

Kim, H. (2020). Effects of rating criteria order on the halo effect in L2 writing assessment: a many-facet Rasch measurement analysis. *Language Testing in Asia*, *10*(1), 16.

Klein, E., & Knops, A. (2023). The two-network framework of number processing: a step towards a better understanding of the neural origins of developmental dyscalculia. *Journal of Neural Transmission*, *130*(3), 253–268.

Kohar, A. W., Rahaju, E. B., Fardah, D. K., & Hanifah, U. (2024). Assessing student teachers' ability in posing mathematical reasoning problems. *Center for Educational Policy Studies Journal, 14*(2), 9–37.

Köhler, C., Robitzsch, A., & Hartig, J. (2020). A bias-corrected RMSD item fit statistic: An evaluation and comparison to alternatives. *Journal of Educational and Behavioral Statistics*, *45*(3), 251–273.

Kohn, J., Rauscher, L., Kucian, K., Käser, T., Wyschkon, A., Esser, G., & von Aster, M. (2020). Efficacy of a computer-based learning program in children with developmental dyscalculia. What influences individual responsiveness?. *Frontiers in psychology*, *11*, 505087.

Kusmaharti, D., Yustitia, V., Faizah, H., Rohimah, S. M., & Faridah, L. (2025). Critical Thinking Processes in Numeracy Tasks among Primary School Students with High Mathematics Anxiety. *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah*, *10*(2), 571-582.

Lena, M. S., & Nikolov, M. (2025). The roles of teachers and contextual and motivational factors in young learners' motivation: a structural equation modelling (SEM) approach. *Education Sciences*, *15*(10), 1388.

Lim, W. M. (2025). What is qualitative research? An overview and guidelines. *Australasian Marketing Journal*, *33*(2), 199–229.

Love, B., Vetere, A., & Davis, P. (2020). Should interpretative phenomenological analysis (IPA) be used with focus groups? Navigating the bumpy road of "iterative loops," idiographic journeys, and "phenomenological bridges" .. *International Journal of Qualitative*

Methods, *19*, 1609406920921600.

Marný, S., Alýman, M., & Harsýatý, T. (2020). Students' critical thinking skills based on gender and knowledge group. *Journal of Turkish Science Education*, *17*(4), 544–560.

Martins, L. G., & Martinho, M. H. (2021). Estratégias, dificuldades e comunicação escrita na resolução de um problema matemático. *Bolema: Boletim de Educação Matemática* (70), 903–936.

May, J., Redding, E., Whatley, S., £ucznik, K., Clements, L., Weber, R., ... & Reed, S. (2020). Enhancing creativity by training metacognitive skills in mental imagery. *Thinking Skills and creativity*, *38*, 100739.

Mitani, H. (2021). Test score gaps in higher order thinking skills: Exploring instructional practices to improve the skills and narrow the gaps. *AERA Open*, *7*, 23328584211 016470.

Mohammed, S. G. A., Kadah, S. M. S., Khattab, A. N., & ElNeshwey, H. M. (2024). Exploring the relationship between dyscalculia and working memory in Egyptian children. *The Egyptian Journal of Otolaryngology*, *40*(1), 163.

Mokkink, L. B., Boers, M., Van Der Vleuten, C. P. M., Bouter, L. M., Alonso, J., Patrick, D. L., ... & Terwee, C. B. (2020). COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Medical Research Methodology*, *20*(1), 293.

Mokkink, L. B., Herbelet, S., Tuinman, P. R., & Terwee, C. B. (2025). Content validity: judging the relevance, comprehensiveness, and comprehensibility of an outcome measurement instrument–a COSMIN perspective. *Journal of Clinical Epidemiology*, *185*, 111879.

Muharani, D. R., & Sucipto, L. (2023). The effect of self-efficacy and math anxiety on students' non-routine mathematics problem solving ability. *IJECA International Journal of Education & Curriculum Application*.

Muhtarom, M., Shodiqin, A., & Astriani, N. (2020). Exploring senior high school student's abilities in mathematical problem posing. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, 69–79.

Nurmaliza, N., Dahlia, A., Nufus, H., & Nurdin, E. (2022). Analysis of mathematical critical thinking ability viewed from habits of mind. *Jurnal Pendidikan MIPA*, *23*(3), 904–917.

Onoshakpokaiye, O. E., & Eyetan, V. (2026). Exploring anxiety levels in mathematics and gender dispositions among students in secondary school. *Indonesian Journal of Education and Mathematical Science*, *7*(1), 17–28.

Padgett, R. N., & Morgan, G. B. (2020). Using the eRm package for Rasch modeling. *Measurement: Interdisciplinary Research and Perspectives*, *18*(3), 163–176.

Papini, N., Kang, M., Ryu, S., Griese, E., Wingert, T., & Herrmann, S. (2021). Rasch calibration of the 25-item Connor-Davidson resilience scale. *Journal of Health Psychology*, *26*(11), 1976-1987.

Park, E. Y. (2019). Rasch analysis of the disability acceptance scale for individuals with cerebral palsy. *Frontiers in Neurology*, *10*, 1260.

Parsazadeh, N., Cheng, P. Y., Wu, T. T., & Huang, Y. M. (2021). Integrating computational thinking concept into digital storytelling to improve learners' motivation and performance. *Journal of Educational Computing Research, 59*(3), 470–495.

Peters, L., de Beeck, H. O., & De Smedt, B. (2020). Cognitive correlates of dyslexia, dyscalculia and comorbid dyslexia/dyscalculia: Effects of numerical magnitude processing and phonological processing. *Research in Developmental Disabilities*, *107*, 103806.

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *Zdm*, *52*(1), 33-44.

Powell, S. R., Berry, K. A., & Benz, S. A. (2020). Analyzing the word-problem performance and strategies of students experiencing mathematics difficulty. *The Journal of Mathematical Behavior*, *58*, 100759.

Putra, B. Y. G., Yaniawati, P., Fitri, N., Sharif, S., Supianti, I. I., & Mudrikah, A. (2025). Analyzing student errors in PISA "Change and Relationship" tasks using the AVAEM framework. *Infinity Journal*, *14*(4), 1105–1128.

Putri, I. I., Rahmat, A., Riandi, R., & Riza, L. S. (2024). Assessing undergraduate cognitive system thinking instruments in genetics lectures: a rasch model analysis. *AL-ISHLAH: Jurnal Pendidikan*, *16*(3), 3924–3937.

Rahayuningsih, S., Sirajuddin, S., & Nasrun, N. (2021). Cognitive flexibility: exploring students' problem-solving in elementary school mathematics learning. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, 59-70.

Ramírez-Uclés, I. M., & Ramírez-Uclés, R. (2020). Gender differences in visuospatial abilities and complex mathematical problem solving. *Frontiers in psychology*, *11*, 191.

Reigosa-Crespo, V., Castro-Cañizares, D., Estévez-Pérez, N., Santos, E., Torres, R., Mosquera, R., ... & Valdés-Sosa, M. (2020). Numerical skills and dyscalculia. From basic research to practice in Cuba (Habilidades numéricas y discalculia: de la investigación básica a la práctica en Cuba). *Studies in Psychology*, *41*(2), 373–403.

Reinhold, F., Hofer, S., Berkowitz, M., Strohmaier, A., Scheuerer, S., Loch, F., ... & Reiss, K. (2020). The role of spatial, verbal, numerical, and general reasoning abilities in complex word problem solving for young female and male adults. *Mathematics Education Research Journal*, *32*(2), 189–211.

Ricco, R. B., Koshino, H., Sierra, A. N., Bonsel, J., Monteza, J. V., & Owens, D. N. (2021). Individual differences in analytical thinking and complexity of inference in conditional reasoning. *Thinking & Reasoning*, *27*(3), 319–349.

Rodríguez-Rojas, M. P., Ramírez-Segura, A. M., Valenzuela-Mora, P., Dampuré, J., & Riveros Munévar, F. (2024). Critical thinking evaluation scale: Design and validation in a Colombian population. *Sage Open*, *14*(4), 21582440241297418.

Roorda, G., de Vries, S., & Smale-Jacobse, A. E. (2024, January). Using lesson study to help mathematics teachers enhance students' problem-solving skills with teaching through problem solving. In *Frontiers in Education* (Vol. 9, p. 1331674). Frontiers Media SA.

Rubenstein, L. D., Callan, G. L., Neumeister, K. S., & Ridgley, L. M. (2020). Finding the problem: How students approach problem identification. *Thinking Skills and Creativity*, *35*, 100635.

Saga, M., Rkhaila, A., Ounine, K., & Oubaha, D. (2022). Developmental dyscalculia: The progress of cognitive modeling in the field of numerical cognition deficits for children. *Applied Neuropsychology: Child*, *11*(4), 904–914.

Salisa, R. D., & Meiliasari, M. (2023). A literature review on dyscalculia: What dyscalculia is, its characteristics, and difficulties students face in mathematics class. *Alifmatika: Jurnal Pendidikan dan Pembelajaran Matematika*, *5*(1), 82–94.

Schreiber, I., & Ashkenazi, H. (2024). The impact of various teaching methods on the knowledge of students of different genders: The case of mathematics word problems abstract. *Heliyon*, *10*(16).

Schukajlow, S., Rakoczy, K., & Pekrun, R. (2023). Emotions and motivation in mathematics education: Where we are today and where we need to go. *ZDM–Mathematics Education*, *55*(2), 249–267.

Shreeves, M., Gugerty, L., & Moore, D. (2020). Individual differences in strategy use and performance during fault diagnosis. *Cognitive Research: Principles and Implications*, *5*(1), 49.

Simanjuntak, D. A., Makmuri, M., & Rahayu, W. (2021). Problem-based learning: Its effect on higher-order mathematical thinking skills in terms of student's initial abilities and mathematical beliefs. *Al-Jabar: Jurnal Pendidikan Matematika*, *12*(2), 457–466.

Stoet, G., & Geary, D. C. (2020). Gender differences in the pathways to higher education. *Proceedings of the National Academy of Sciences*, *117*(25), 14073–14076.

Strohmaier, A. R., Schons, C., Knabbe, A., Vogel, M., Saado, K., & Obersteiner, A. (2025). The relation between text comprehension and computation in mathematical word-problem solving. *Learning and Instruction*, *101*, 102235.

Suherman, S., & Vidákovich, T. (2022). Assessment of mathematical creative thinking: A systematic review. *Thinking Skills and Creativity*, *44*, 101019.

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi, J. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan dan Pengajaran*, *55*(1), 127–141.

Sumarna, N., & Rahmawati, I. (2024). *Pengembangan kemampuan berpikir kritis dan literasi matematika: penerapan desain instruksional pengintegrasian PBL dan ClassPoint* [Developing critical thinking skills and mathematical literacy: implementing integrated PBL and classpoint instructional design]. *PYTHAGORAS: Jurnal Matematika dan Pendidikan Matematika*, *19*(2), 153–164.

Suseelan, M., Chew, C. M., & Chin, H. (2023). Higher-order thinking word problem-solving errors made by low-performing pupils: Comparative case study across school types in Malaysia. *Current Psychology*, *42*(25), 21449–21461.

Syaidi, A., & Murti, R. C. (2024). Optimizing critical thinking skills of fourth grade students through inquiry-based learning: elementary school mathematics. *Jurnal Pendidikan MIPA*, *25*(2), 752-764.

Tomaszewski, L. E., Zarestky, J., & Gonzalez, E. (2020). Planning qualitative research: Design and decision making for new researchers. *International journal of qualitative methods*, *19*, 160940692 0967174.

Uto, M. (2021). Accuracy of performance-test linking based on a many-facet Rasch model. *Behavior Research Methods, 53*(4), 1440–1454.

van Dijke-Droogers, M., Drijvers, P., & Bakker,

A. (2022). Introducing statistical inference: Design of a theoretically and empirically based learning trajectory. *International Journal of Science and Mathematics Education*, *20*(8), 1743-1766.

Wahab, A., Kusuma, Y. S., Juandi, D., Turmudi, T., Buhaerah, B., & Syaiful, S. (2024). Understanding students' struggles in solving mathematical problems: a systematic literature review using polya's framework. *Jurnal Pendidikan Progresif* *14*(3), 1728-1753.

Ye, J., Lai, X., & Wong, G. K. W. (2022). A multigroup structural equation modeling analysis of students' perception, motivation, and performance in computational thinking. *Frontiers in Psychology*, *13*, 989066.