Journal on Mathematics Education
Volume 16, No. 2, 2025, pp. 529-558

Mathematical reasoning and communication word problems with mathematical problem-solving orientation: A relation between the skills

Ahmad Naufal Aljura[1,*], Heri Retnawati[1], Septinda Rima Dewanti[2], Gulzhaina Kuralbayevna Kassymova[3,4], Rimajon Sotlikova[5], Atut Reni Septiana[6]

[1]Department of Mathematics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
[2]School of Early Childhood and Inclusive Education, Queensland University of Technology, Australia
[3]Abai Kazakh National Pedagogical University, Alamaty, Kazakhstan
[4]Institute of Metallurgy and Ore Beneficiation JSC, Satbayev University, Alamaty, Kazakhstan
[5]TESOL Program, Webster University, Tashkent, Uzbekistan
[6]Department of Materials, University of Manchester, Manchester, United Kingdom
*Correspondence: ahmadnaufal.2023@student.uny.ac.id

Received: 22 February 2025 | Revised: 23 March 2025 | Accepted: 24 April 2025 | Published Online: 19 June 2025

Abstract

Developing students' mathematical reasoning skills (MRS) and mathematical communication skills (MCS) is crucial, as both are foundational to effective mathematical problem-solving (MPS). Despite their theoretical interconnectedness, limited empirical evidence exists on how MRS and MCS relate to MPS, particularly in problem-based contexts. This study investigates the relationship between MRS and MCS within an MPS-oriented framework using a quantitative, descriptive correlational design. A modified mathematical word problem (MWP) essay test was administered to 117 students across two pilot classes. The test items were designed to elicit reasoning and communication processes associated with MPS. Psychometric analyses—including evidence of content validity (Aiken's V), consequential validity, reliability indices (α and ω), and item-level metrics (discrimination and difficulty)—confirmed the instrument's robustness. Factor analysis supported a unidimensional structure aligned with MPS. Correlational analyses revealed significant positive associations between MRS and MCS, meeting bivariate normality assumptions. Pearson's r was 0.529 (95% CI: 0.261–0.722), Spearman's ρ was 0.493 (95% CI: 0.215–0.697), and Kendall's τ was 0.400 (95% CI: 0.101–0.632), indicating a strong relationship. These findings underscore the interdependence of reasoning and communication skills in the context of MPS. The study also offers a detailed analysis of student obstacles in solving MWPs, offering a nuanced understanding of cognitive and linguistic dimensions in MPS. Implications are discussed for researchers, policymakers, and educators, particularly in designing instructional interventions that strengthen MRS and MCS in support of MPS.

Keywords: Correlational Study, Mathematical Communication Skill, Mathematical Problem-Solving Orientation, Mathematical Reasoning Skill, Mathematical Word Problem

How to Cite: Aljura, A. N., Retnawati, H., Dewanti, S. R., Kassymova, G. K., Sotlikova, R., & Septiana, A. R. (2025). Mathematical reasoning and communication word problems with mathematical problem-solving orientation: A relation between the skills. *Journal on Mathematics Education, 16*(2), 529–558. https://doi.org/10.22342/jme.v16i2.pp529-558

One of the primary goals of the National Curriculum of Education in Indonesia, as well as those of various international educational organizations, is to develop students' mathematical proficiency. This objective is intended to foster not only cognitive growth but also psychological, emotional, and practical skill development.

Supporting this view, Hoyles et al. (2002) reported that mathematical competence plays a vital role across a wide range of professional fields, from engineering to tourism.

The National Council of Teachers of Mathematics (NCTM, 2000) has identified five key strands of mathematical proficiency: problem-solving, reasoning, communication, connections, and representation. Similarly, the Indonesian Education Standards, Curriculum, and Assessment Agency (BSKAP, 2022, p. 133) has outlined five essential mathematical competencies: conceptual understanding; reasoning and proof; problem-solving; communication and representation; and connections. Among these, mathematical reasoning (MRS) and communication (MCS) are particularly critical and are often interdependent (Palinussa et al., 2021). Reasoning can be communicated through diverse forms of representation, including dynamic visualizations and verbal explanations (Herbert et al., 2022). When students develop robust reasoning abilities alongside effective mathematical communication skills, they are better equipped to construct meaning from mathematical symbols and concepts (Sumpter & Hedefalk, 2015).

Several studies have examined MRS and MCS across various pedagogical frameworks, including Realistic Mathematics Education (RME) (Habsah, 2017; Palinussa et al., 2021), reflective learning environments (Junsay, 2016), the profiling of MRS based on students' MCS orientation (Sumarsih et al., 2018), and the profiling of MCS based on mathematical problem-solving (MPS) orientation (Puspa et al., 2019). Further, research has also investigated the correlation between MRS and MPS (Anggoro et al., 2022). However, to date, no empirical study has explicitly examined the relationship between MRS and MCS in the context of students' orientation toward MPS. This gap highlights the need for further investigation into how reasoning and communication interact in the development of mathematical problem-solving competencies.

Mathematical Problem-Solving (MPS) Orientation

This study investigates MRS and MCS, with a particular emphasis on their roles within the context of MPS. The research is grounded in the Singapore Pentagon Framework (Leong et al., 2011), which has been credited with contributing to Singaporean students' consistent high performance in international assessments such as PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study). Within this framework, MPS is situated at the center of five interrelated components—attitudes, metacognition, processes, concepts, and skills—signifying its centrality in mathematics education. Similarly, the National Council of Teachers of Mathematics (NCTM, 2000) identifies MPS as a fundamental strand in the development of mathematical proficiency.

Empirical research supports the interconnections among MRS, MCS, and MPS. For instance, Anggoro et al. (2022), using structural equation modeling, found a significant alignment between the theoretical and empirical models of MRS and MPS, suggesting that reasoning skills can be effectively directed toward enhancing students' problem-solving abilities. Furthermore, Sumarsih et al. (2018) argued that students' MRS is often manifested through their MCS, indicating that communication practices can serve as indicators of reasoning processes. Puspa et al. (2019) further demonstrated that variations in MCS may correspond to differences in MPS capabilities, implying that communication can be considered a focal element in students' problem-solving performance. Collectively, these findings underscore the interconnectedness of reasoning and communication skills when oriented toward mathematical problem-solving.

The emphasis on MPS is supported by recent theoretical and empirical work that views it as instrumental in fostering mathematical thinking (Rott et al., 2021). MPS enables students to acquire new mathematical knowledge, apply diverse strategies across varying contexts, and critically reflect on their problem-solving processes (NCTM, 2000). As a result, mastering MPS is essential, given the complex cognitive demands it places on learners (Kingsdorf & Krawec, 2014).

The literature offers several conceptualizations of MPS. Pólya (2004), a foundational figure in this domain, proposed a heuristic approach comprising four phases: understanding the problem, devising a plan, carrying out the plan, and looking back to reflect on the solution. Schoenfeld (1985) expanded this perspective by defining MPS as the ability to confront non-routine problems and introduced a framework based on four interrelated components: resources, heuristics, control, and belief systems. According to Schoenfeld, problem-solving entails more than procedural execution; it involves analytical reasoning, exploration, implementation, and verification. Similarly, Hayes (in Solso et al., 2013) described MPS as a sequence of cognitive processes, including problem identification, representation, solution planning, execution, and evaluation. Eggen and Kauchak (1996) proposed a comparable model that emphasizes problem formulation, strategy selection, implementation, and result assessment.

While these models provide foundational insight, they are largely normative and linear in nature (Aljura et al., 2025), potentially limiting their applicability in real classroom contexts where students may not adhere to fixed procedural stages. In response to this limitation, Rott et al. (2021) developed a descriptive model of MPS that offers greater flexibility. Their model consists of five dynamic and possibly non-sequential phases: understanding (analysis), exploration, planning, implementation (often merged with planning), and verification. This approach acknowledges the non-linear and iterative nature of students' actual problem-solving processes.

Building upon earlier studies, such as Bjuland's (2007) investigation of reasoning within three problem-solving models including Pólya's framework, the current study seeks to extend this work by integrating indicators of both mathematical reasoning and communication. These enhancements are guided by the Rott-Specht-Knipping descriptive model of MPS (Rott et al., 2021), which departs from traditional linear paradigms by allowing for greater flexibility in student approaches. Although this model was originally applied to geometry contexts at the university level, recent work by Aljura et al. (2025) demonstrates its applicability among secondary school students, supporting its broader relevance across educational levels.

Mathematical Reasoning and Communication Skills (MRS and MCS)

MRS and MCS skills have been defined in various ways in curriculum documents, highlighting their importance in achieving mathematics learning objectives (Herbert et al., 2022). BSKAP (2022, p. 136) wrote that:

> *"Reasoning is related to the process of using relationship patterns in analyzing situations to formulate and investigate assumptions."*

BSKAP also wrote that:

> *"Mathematical communication is the formation of a flow of understanding of mathematics learning materials by communicating mathematical thinking using appropriate mathematical language. Mathematical communication also includes the process of analyzing and evaluating the mathematical thinking of others."*

The main definition for MRS is synthesized from Bjuland (2007), who stated that there are three categories for a student to express their MRS oriented towards MPS: visualising, monitoring, and questioning. This heuristic strategy aligns with NCTM (2000), which states that MRS enables students to make and investigate mathematical conjectures, develop and evaluate arguments, and select and use various types of reasoning. However, no one has provided a definition related to MCS oriented towards MPS.

A definition for MPS-oriented MCS can be synthesized through NCTM's (2000) perspective, which states that MCS enables students to organize and consolidate their mathematical thinking and reasoning through communication, articulate their mathematical thinking clearly, analyze and evaluate the mathematical thinking and strategies of others, and use mathematical language to express mathematical ideas accurately. Key aspects include communicating ideas and thoughts through mathematical language, such as graphs, tables, formulas, equations, and mathematical terms. These key elements align with the views of BSKAP (2022, p. 136), Kamid et al. (2020), Rohid et al. (2019), and Tong et al. (2021). Therefore, the conceptual definition of MRS and MCS can be established.

MRS is crucial for students as it plays a vital role in the mathematics learning process. MRS is involved in logical thinking, drawing conclusions while solving mathematical problems, understanding mathematical concepts, utilizing mathematical ideas and procedures effectively, and retaining mathematical knowledge. Similarly, MCS is essential for students as it helps organize and articulate mathematical thinking through communication using mathematical language. It enables students to express, represent, and explain their mathematical ideas clearly to others. In addition to these two conceptual definitions, studying these combined abilities may be considered important and valuable if the focus is on MPS.

Mathematical Word Problem (MWP)

MWP requires students to interpret and analyze information presented in verbal form (Kingsdorf & Krawec, 2014). MWPs may involve single or multiple solution steps, demand the identification of relevant mathematical operations, and sometimes include extraneous or misleading information. These problems are often embedded in real-world contexts, making contextual understanding a critical component of their solution (Böswald & Schukajlow, 2023; Milazoni et al., 2022; Verschaffel et al., 2000, 2020). As Novak and Tassell (2017) note, solving MWPs involves interpreting information, associating numerical values with words, and executing appropriate calculations.

MWPs have a long-standing role in mathematics education, with origins tracing back to ancient civilizations such as Egypt, China, and India (Verschaffel et al., 2020). They are taught at all levels of education and remain an essential component of instructional practice (Daroczy et al., 2015). Solving MWPs requires students to reason mathematically, interpret narrative information, analyze problems, and derive solutions (Aziza et al., 2023; Ningrum et al., 2019). In doing so, students must evaluate and synthesize information to make informed decisions (Lombasari et al., 2022). Moreover, when well-designed, MWPs can foster student engagement by presenting mathematics through meaningful and relatable scenarios (Boaler, 1993; Kingsdorf & Krawec, 2014).

The effectiveness of MWPs in assessment depends on how well the problems are aligned with both the characteristics of the task and the students' abilities. Problem features—such as semantic complexity, syntactic structure, and the inclusion of irrelevant information—can significantly influence students' performance (Kameenui & Griffin, 1989). Challenges arise when problems contain ambiguous language (Sandberg & De Ruiter, 1985), require unconventional strategies (Larsen et al., 1978), consist of multiple solution steps (Quintero, 1983), or lack supportive visual representations (Moyer, Moyer, et al., 1984; Moyer, Sowder, et al., 1984). These factors can result in "reading-related memory overload" (Moyer, Moyer, et al., 1984), where students expend cognitive resources on decoding the problem statement, leaving insufficient capacity for actual problem-solving (Kameenui & Griffin, 1989).

MRS and MCS with MPS Orientation

Reasoning encompasses intuitive, inductive, deductive reasoning (Baroody, 1993; Bozkuş & Ayvaz, 2018), and adaptive reasoning (Bozkuş & Ayvaz, 2018). Intuitive reasoning relies on spontaneous assumptions without formal analysis, inductive reasoning involves identifying patterns in mathematical problems by observing and analyzing relationships between patterns to infer rules for creating a pattern, and deductive reasoning entails predicting, reasoning, and evaluating the logic or argument's appropriateness and consistency (Ansari et al., 2020; Baroody, 1993). Adaptive reasoning requires adapting mathematical facts, procedures, concepts, and methods to various situations, demonstrated when individuals, particularly students, can reflect, explain, and justify their solutions (Syukriani et al., 2017).

MRS is essential for students to engage in cognitively demanding activities and draw new conclusions based on evidence (Herbert et al., 2022). MRS encompasses patterns and problems, conjectures and proofs, representations, and numerical abilities (Nickerson, 2011). NCETM (The National Center for Excellence in the Teaching of Mathematics) defines MRS as the process of logically thinking through mathematical problems to reach solutions (Öztürk & Sarikaya, 2021).

MRS is a process where arguments are exchanged to reach a convincing conclusion (Viholainen, 2011). According to Herbert et al. (2022), students demonstrate MRS when they can explain their thinking, make deductions, provide reasons for their strategies, and transfer knowledge across contexts to reach conclusions. The process of MRS involves identifying a problem, selecting a strategy, applying the strategy, and drawing a conclusion (Lithner, 2000). Bjuland (2007) identified five key MRS processes: sense-making, conjecturing, convincing, reflecting, and generalizing. Next, the explanations of MCS are provided. In classroom learning, mathematical communication is defined as planned interaction in a classroom setting, involving strategies such as questioning, discussion, and group activities to encourage students to express, share, and reflect on their ideas (Kaya & Aydın, 2016). Tong et al. (2021) categorized mathematical communication into verbal communication (speaking and listening), listening (understanding verbal communication through reading), and written communication (completing assignments). Schoen, Bean, and Ziebarth further explained that mathematical communication involves students' ability to clearly and uniquely explain algorithms and methods for solving mathematical problems, as well as their skills in representing real-world phenomena through graphs, equations, sentences, and tables (Qohar, 2011).

The Common Core State Standards (CCSS) for mathematics define mathematical communication skill (MCS) as a crucial aspect of a student's ability to think, where they assess the validity of conclusions, convey them to others, and engage with arguments from others (Cartwright, 2020). It involves the capacity to articulate and elucidate ideas in a manner that is comprehensible to others (Tong et al., 2021). This skill encompasses expressing, comprehending, interpreting, evaluating, and responding to mathematical concepts using terminology, notations, and symbols to convey mathematical ideas effectively (Rohid et al., 2019). MCS also includes the ability of students to utilize mathematical language to articulate mathematical concepts, as well as to organize and integrate their thoughts in communication (Kamid et al., 2020). MRS can be directed towards MPS (Anggoro et al., 2022). Students' MCS reflects their MRS (Sumarsih et al., 2018), and vice versa. MPS is the central focus of MCS (Puspa et al., 2019). These results from the assumption of the relationship between MRS and MCS oriented towards MPS.

Building on the conceptual definitions and explanations of MRS and MCS provided earlier, operational definitions of MRS and MCS can be established in the context of MPS orientation. MRS involves students using their cognitive functions to solve mathematical problems by (1) analyzing the problem (sense-making), (2) formulating conjectures (conjecturing), (3) applying investigation results (convincing), and (4) verifying and generalizing conclusions (reflecting-generalizing). MCS is the skill to solve math problems using various mathematical language components such as terms, symbols, graphs, and tables through (1) selecting relevant components (expressing), (2) composing and understanding their use, (3) interpreting results, and (4) evaluating and presenting the final answer. Table 1 describes the processes and indicators for MRS and MCS involved with solving MWPs.

Table 1. MRS and MCS processes and indicators in solving MWPs

Skill	Code	Process	Indicator
Mathematical Reasoning (MRS)	Sm	Sense-making	a. Students can accurately list relevant information in response to questions in the MWP. b. Students can establish a logical connection between known information and the questions in the MWP.
	Cj	Conjecturing	a. Students can make informed conjectures based on prior findings. b. Students can provide rationale for their conjectures.
	Cv	Convincing	a. Students can correctly apply findings and conjectures. b. Students can comprehensively apply findings and conjectures.
	RG	Reflecting-Generalizing	a. Students can verify answers using alternative methods. b. Students can draw logical conclusions.
Mathematical Communication (MCS)	Ex	Expressing	a. Students can identify pertinent information for MWP questions. b. Students can select appropriate mathematical language based on known information.
	Un	Understanding	a. Students can accurately organize, and link chosen mathematical languages. b. Students can thoroughly organize, and link chosen mathematical languages.
	In	Interpreting	a. Students can correctly apply the organized relationships of mathematical languages in solving MWP questions. b. Students can fully apply the organized relationships of mathematical languages in solving MWP questions.
	Ev-Pr	Evaluating and Presenting	a. Students can verify answers using alternative mathematical languages. b. Students can present the final answer using multiple mathematical languages.

Table 1 presents the operational definitions for each stage/process and indicator for MRS and MCS utilized in developing the test instruments. These definitions were also used to synthesize operational definitions for mathematical reasoning word problems (MRWP) and mathematical communication word problems (MCWP).

Mathematical Reasoning and Communication Word Problem (MRWP and MCWP) Test

Mathematical word problem (MWP) tests can be adapted to evaluate specific mathematical skills, such as mathematical reasoning and communication skills (MRS and MCS). This adaptation leads to the development of mathematical reasoning word problems (MRWP) and mathematical communication word problems (MCWP). MRWP assesses MRS, and MCWP assesses MCS, targeting the indicators in Table 1. Table 2 to Table 5 provide examples of questions from the MWP tests given to students.

Table 2. MRWP Part 1

The "Bike Chain" Problem
Andi and Budi were riding BMX bikes together during a school holiday. A few moments after passing a steep road, Andi's bike chain suddenly broke, and some parts were lost and destroyed. Andi and Budi decided to stop at the side of the road. They tried to find the broken bike chain pieces and salvaged a remaining piece measuring 9 cm. Budi had anticipated that Andi's bike chain might break because it appeared rusty. Therefore, Budi had purchased a spare chain that had not been cut. Imagine you are Budi, and you want to help your friend Andi repair his bike chain by adjusting the size of the spare chain to match the size of Andi's broken bike chain. Analyze the information provided and answer the following questions to determine the appropriate size for the replacement chain.

No	Question	MRS Process
1a	It is evident that on Andi's bicycle, the circumference of the front gear is three times that of the rear gear, and the diameter of the rear gear is 6 cm. What information is crucial for determining the size of the replacement chain?	Sm
1b	Assuming you are using calculations to find a solution, it is important to note that the broken chain is the bottom chain where the two ends, when connected, will form a tangent line to the two gears. Based on the previously obtained information, make a conjecture and provide the reasons for your choice.	Cj
1c	If the distance between the gears is 33 cm, calculate the size of the chain required to repair Andi's bicycle chain.	Cv
1d	Given that the available chain size is 50 cm, determine the length of the chain that needs to be cut to repair Andi's bicycle chain. Additionally, suggest an alternative solution for repairing Andi's bicycle chain if calculating the size in (c) is deemed unnecessary.	RG

Table 3. MRWP Part 2

The "Sailor" Problem
Enrique was a sailor in the 16th century who navigated the oceans in a large ship to circumnavigate the globe. He served as a navigator stationed at the center mast of the ship to monitor its direction. Enrique was curious about his maximum visibility range. If you were in Enrique's position, study the information provided and answer the following questions to determine your visibility range to the horizon.

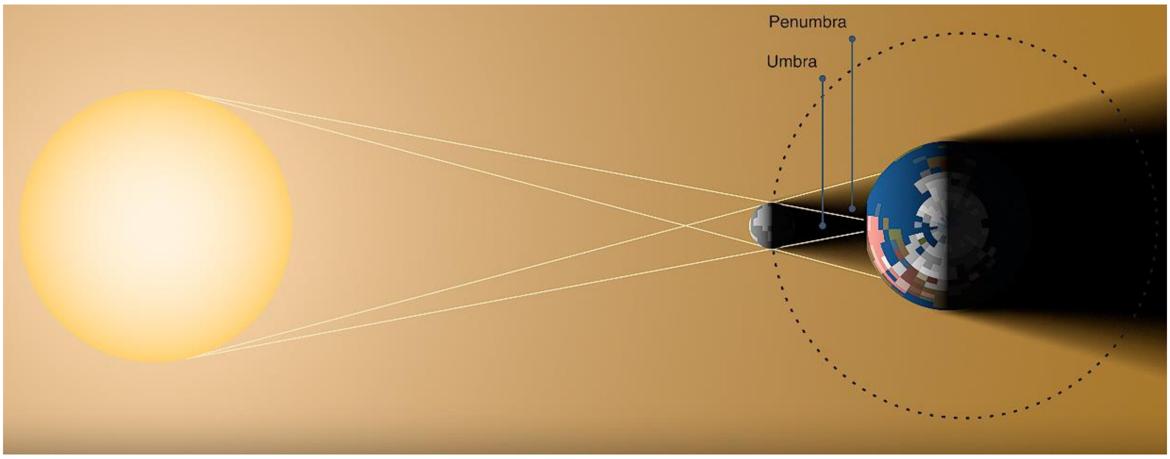
No	Question	MRS Process
2a	It is known that the radius of the Earth is twice the radius of Mars, the diameter of the Moon is half of the diameter of Mars, and the circumference of the Moon is	Sm

No	Question	MRS Process
	10,921 km. What information is necessary for you as a navigator to estimate your visibility to the horizon?	
2b	Based on the given information, make a conjecture and explain how you would determine your visibility to the horizon from the ship's center mast.	Cj
2c	If a ship is visible by the navigator on the center mast with an elevation angle of 45° and an estimated distance of 300 m from your ship's center mast, calculate your visibility to the horizon as seen by Enrique towards the ocean.	Cv
2d	If a ship is located 64.6 km to the right of your ship and the estimated width of your ship is 17.6 m, can you see it if you were at the ship's center mast? Additionally, suggest an alternative method to determine your visibility as Enrique without using the calculation from (c), assuming you can see a small dot in the shape of a ship at the edge of your view.	RG

Table 4. MCWP Part 1

The "Solar Eclipse" Problem

Look at the following picture.



Source: https://science.nasa.gov/eclipses/geometry/

The event depicted in the picture above is a Total Solar Eclipse. Your teacher has assigned you a task related to the tangent line of the circle during the event. Study the information provided and answer the following questions to estimate the length of sunlight reaching the Moon to create a Penumbra during a Total Solar Eclipse!

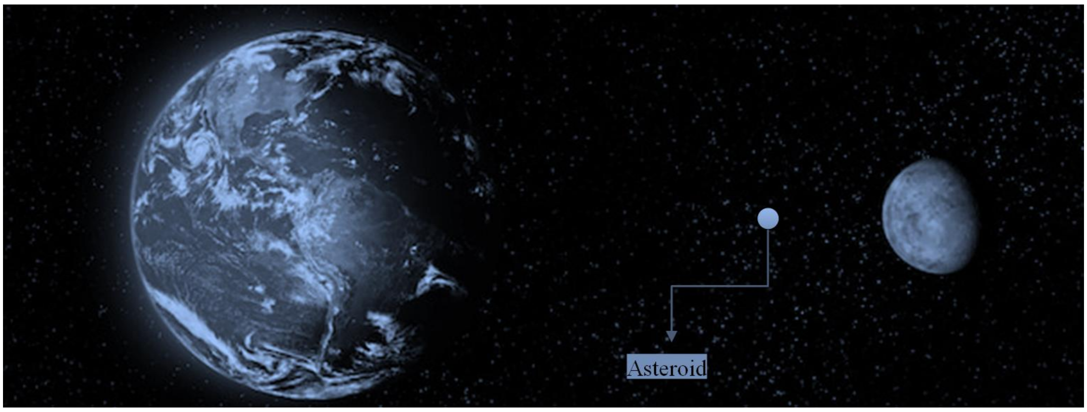
No	Question	MCS Process
3a	The diameter of the Sun is 109 times the diameter of the Earth, and the ratio of the Earth's radius is 3.67 times the radius of the Moon, with the circumference of the Moon being 10,921 km. What information is important and necessary to estimate the length of sunlight to the Moon to form a Penumbra during the event?	Ex
3b	Create a two-dimensional graph illustrating circles and lines that contain the essential information needed to estimate the length of sunlight to the Moon for the formation of a Penumbra in the event.	Un
3c	If the distance between the Sun and the Moon is 149,600,000 km, estimate the length of sunlight required to reach the Moon for the formation of a Penumbra during the event.	In

No	Question	MCS Process
3d	Is the distance between the Sun and the Moon greater than the length of sunlight needed to reach the Moon in a Total Solar Eclipse? Present your answer in the form of a graph sketch and provide a description.	Ev-Pr

Table 5. MCWP Part 2

The “Earth, Asteroid, and Moon” Problem

Look at the following picture.



At the center point between the Earth and the Moon, an asteroid is passing by. The asteroid, considered as a point, is intersected by two-line segments that meet at the asteroid. These two-line segments would act as tangents to the Earth and the Moon if extended. Given these conditions, analyze the information and answer the following questions to compare the distance of the asteroid to the Earth and the distance of the asteroid to the Moon.

No	Question	MCS Process
4a	The diameter of the Moon is 0.27 times the diameter of the Earth, with a circumference of the Moon being 10,921 km. What information is important and necessary to determine the ratio between the distance of the asteroid to the Earth and the distance of the asteroid to the Moon?	Ex
4b	Describe the event using a two-dimensional graph that includes circles and line segments with important information to determine the ratio between the distance of the asteroid to the Earth and the distance of the asteroid to the Moon.	Un
4c	If the ratio between the tangent line segment from the asteroid to the Earth and the length of the tangent line segment from the asteroid to the Moon is 3:1, and the length of the tangent line segment from the asteroid to the Earth is 2,250 km, determine the ratio between the distance of the asteroid to the Earth and the distance of the asteroid to the Moon.	In
4d	Is the ratio between the distance of the asteroid to the Earth and the distance of the asteroid to the Moon the same as the ratio between the length of the tangent line segment from the asteroid to the Earth and the length of the tangent line segment from the asteroid to the Moon? Present your answer in the form of a graph sketch and provide a description.	Ev-Pr

Tangent Lines to Circles in the Indonesian Curriculum

Geometry is a branch of mathematics that deals with points, lines, planes, and spatial objects, and the relationships between them. The term "Geometry" originates from the ancient Greek words *geometría*, "geo" (earth) and "metron" (measure) (Hershkowitz, 2020). Geometry has a long history, dating back to ancient Egypt and the Babylonian civilization, where it was applied in agriculture, animal husbandry, and notably in the construction of the Egyptian pyramids (Silmi Juman et al., 2022). In modern times, geometry has evolved and been applied in diverse fields such as astronomy, architecture, engineering, and physics.

Geometry is a mathematical field that deals with the space around us, focusing on shapes, their properties, and various patterns and thinking processes they inspire and form the foundation for (Hershkowitz, 2020). In line with this, Mammarella et al. (2017) stated that Geometry is focusing on the study of the size, shape, length, relative position of figures, and their spatial properties. Three main aspects have been identified in the development of Geometry: (a) interacting with shapes in a space, shapes, (b) their attributes, and their changes in space as fundamental ingredients for constructing a theory, and (c) shapes as basis for reflecting on visual information by representing, describing, generalizing, communicating, and documenting such information (Hershkowitz, 2020). From these three aspects, we can conclude that mathematical reasoning, communication, and problem-solving skills (MRS, MCS, and MPS) are essential for all aspects of Geometry.

Learning Geometry can enhance students' spatial intelligence (Mammarella et al., 2017). If their spatial intelligence increases, then automatically their MRS and MCS will also increase. Spatial intelligence plays a crucial role in various non-spatial domains, like interpreting graphs and diagrams (Newcombe & Frick, 2010). Moreover, through learning Geometry, students can improve their MPS (Pamungkas & Nugroho, 2020), which of course can be combined with strategies, approaches, and learning tools, including technology and educational media.

In this study, one of the topics in Geometry discussed and focused on is tangent lines to circles. The BSKAP (2022) states the following learning outcome (CP) for the geometry subject in phase F (Grades XI and XII) of the Indonesian curriculum:

> *"At the end of Phase F, students can apply theorems about circles, determine the length of arcs, and calculate the area of a sector of a circle to solve problems, including determining positions on the Earth's surface and calculating distances between two places on Earth."*

Research Purpose and Questions

This study aims to investigate the relationship between MRS and MCS in orientation towards MPS through the provision of MWP (MRWP and MCWP). The research will address the following questions:

1. How is the description of students' MRS and MCS with MPS orientation?
2. Is there a significant correlation between MRS and MCS with MPS orientation?
3. What is the effect size category based on the correlation between MRS and MCS with MPS orientation?
4. What is the confidence interval of the correlation between MRS and MCS with MPS orientation?

METHODS

This study employed a non-experimental quantitative approach using descriptive correlational methods to investigate the relationship between MRS and MCS in orientation towards MPS without giving any treatment or manipulation to the variables (Creswell & Creswell, 2023; Creswell & Guetterman, 2019; Johnson & Christensen, 2019). The study utilized homogeneous purposive sampling, a deliberate technique used to recruit research participants based on their quality and similarities in nature or characteristics (Etikan, 2016). This technique involves identifying and choosing individuals who are willing, capable (able to behave and communicate effectively), and possess sufficient information.

Research Participants

Data were gathered from 117, aged 16 to 18 years, consisting of 43 males and 74 females, grade XI high school students in East Bintan, Indonesia, who had been taught the topic of tangent lines to circles. The participants were purposefully selected from three grade XI classes, with two classes having 40 (20 males and 20 females) and 39 (9 males and 30 females) students, respectively, assigned as the first trial class and one class with 38 (14 males and 24 females) students designated as the second trial class. One mathematics teacher was recruited as a participant for triangulation purposes and to gather more comprehensive data.

Data Collection

Data was collected by administering the MWP test, which included a written essay test of two similar parts. For the first trial class, each test part was administered to each class after the students studied the topic of tangent lines to circles. For the second trial class, the first part was administered after students had studied the tangent lines to circles topic, and the second part was a remedial test. This test is a formative assessment as it is designed to provide feedback and help develop students' mathematical skills (Herbert et al., 2022). Each part of the MWP included the MRWP and MCWP segments, each containing two questions. Students were required to complete one segment within 40 minutes. The test was accompanied by a table of specification and scoring guidelines. Additional data collections, such as teacher interviews and classroom observations, were conducted to complement the primary data on students' MRS and MCS descriptions.

Research Procedure

The research began by developing MWP test instruments (MRWP and MCWP), which were evaluated for quality. Students' responses from the first trial class were analyzed using the classical test theory (CTT) approach, incorporating expert judgments for content validity and item analysis to estimate reliability, discriminatory power, and difficulty levels. Subsequently, the tested test and its quality were administered to students in the second trial class. The students' responses in the selected trial class were then analyzed using correlation and effect size with Fisher's (1921) Z transformation formula. The calculations in this study were facilitated by JASP software (https://jasp-stats.org/), an open and free software tool.

Description of MWP (MRWP and MCWP) Test Quality

In the mathematics learning process, NCTM (2000) emphasized that the ideal assessment principle should be supportive and informative for educators. Thus, certain criteria are needed for evaluation. One of the key evaluation criteria is a high-quality test instrument (Rizbudiani et al., 2021).

The higher the test quality, the more accurately student learning outcomes can be evaluated (Wijayanti, 2020). This evaluation is not solely due to students' ability to answer the test. Although determining the criteria for quality test items is challenging (Osterlind, 1989), item analysis can help assess the quality of each test item. Through item analysis, criteria can be established to determine the quality of each test item, and educators can decide whether an item can be used, needs revision, or should be discarded (Haladyna & Rodriguez, 2021). Evaluating the quality of an assessment tool, especially a test instrument, requires demonstrating validity and reliability (Hancock & An, 2020).

Before conducting the trial, the developed test was assessed by four experts, including one professor and three lecturers, to obtain evidence of content validity. The expert assessment results were calculated using Aiken's V (1985), based on Kania et al.'s (2024) research, to prove the instrument's content validity for measuring MPS. Additionally, the validity category was determined based on Retnawati (2016), where $0.0 < V \leq 0.4$ is classified as low validity, $0.4 < V \leq 0.8$ is classified as moderate validity, and $0.8 < V \leq 1.0$ is classified as high validity.

The estimated Aiken's V results for MWP test parts 1 and 2 ranges from 0.69 to 0.88 for the MRS and MCS indicators. The estimated values for material and construction aspects ranges from 0.75 to 0.88, while the language aspect falls between 0.63 and 0.88. These values indicate moderate to high validity. Expert advice was also sought to enhance and refine the test instrument.

On the other hand, reliability estimation using coefficients α and ω will also be demonstrated. The most widespread and common method for reliability estimation is the α coefficient (Malkewitz et al., 2023; McNeish, 2018; Viladrich et al., 2017). In addition to this study, the coefficient has been utilized in various studies on mathematical problem development (Amalina & Vidákovich, 2022; Novikasari & Dede, 2023).

Several studies have recommended using the ω coefficient due to the strict statistical conditions or assumptions required for estimating the α coefficient (Hancock & An, 2020; McNeish, 2018). Furthermore, the ω coefficient is considered a more robust and practical measure than the α coefficient as it is not constrained by the statistical assumptions of α (Kalkbrenner, 2023). If the data meets the requirements for estimating the α coefficient, then both ω and α coefficients will yield the same value (McDonald, 1999). In essence, the α coefficient is a special case of the ω coefficient. Researchers have the flexibility to calculate α, ω, or both (Kalkbrenner, 2023). Therefore, the choice of estimating reliability through these two coefficients is fundamental.

In addition to establishing content validity and estimating reliability, it is essential to analyze test instrument items to determine the level of difficulty and discriminatory power (Fauzie et al., 2021; Haladyna & Rodriguez, 2021). These factors are crucial psychometric properties that can impact the overall quality of the items (Penfield, 2013). The item-total correlation was calculated to determine the correlation between the item and the total score to estimate the discriminatory power of each item (Guilford, 1950; Hwang, 1970; Marianti et al., 2023; Retnawati, 2016; Reynolds et al., 2008). Although there are alternative methods to estimate discriminatory power (Reynolds et al., 2008), the item-total correlation offers advantages, such as being unaffected by the difficulty level and offering a significance value (Hwang, 1970). Moreover, following the test results with trial class 1, item analysis was conducted to estimate reliability, discriminatory power, and difficulty level. The item analysis results are presented in Table 6.

In Table 6, the reliability of MWP parts 1 and 2 is deemed acceptable as the coefficients α and ω exceed 0.6 (Mohamad et al., 2015). Moreover, the discriminatory power of each item is statistically significant at the 0.1% level with a p-value <0.001. Additionally, all items have been observed to have a high difficulty level, with values below 0.3.

In this study, all four items in both parts of the MSP demonstrated significant discriminatory power in distinguishing test-takers' (students) abilities, further supporting the notion that higher discriminatory power leads to increased test reliability (Haladyna & Rodriguez, 2021). All items in this study were classified as difficult (Johari et al., 2011; Reynolds et al., 2008). Some studies (Fauzie et al., 2021; Johari et al., 2011) recommend revising items with difficulty levels below 0.3 or above 0.7. However, these items may be retained without revision, as their difficulty levels indicate areas where students may need additional instruction (Johari et al., 2011). Overall, based on the analysis presented in Table 6, the MSP test parts 1 and 2 in this study can be considered of good quality, with each item deemed acceptable for use.

Table 6. Reliability, Discriminatory Power, Difficulty Level, and Decisions on MWP Test Parts 1 and 2

MWP	Item	Reliability	Discriminatory Power	Difficulty Level	Decision
Part 1	1	$\alpha = 0.636$	0.641	0.103	Accepted
$(n = 40)$	2	$\omega = 0.665$	0.683	0.134	Accepted
	3		0.639	0.113	Accepted
	4		0.807	0.094	Accepted
Part 2	1	$\alpha = 0.761$	0.872	0.221	Accepted
$(n = 39)$	2	$\omega = 0.732$	0.810	0.151	Accepted
	3		0.617	0.071	Accepted
	4		0.684	0.141	Accepted

Quantitative Proof of MWP (MRWP and MCWP) Test Orientation

In this study, content validity has been established. Although construct validity could also be demonstrated, the focus was on proving that the mathematical word problem (MWP) used measures mathematical reasoning and communication skills (MRS and MCS) based on one dimension of mathematical problem solving (MPS) orientation. As mentioned, the MWP test (MRWP and MCWP) has been designed to be MPS-oriented. Thus, factor analysis was conducted through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) following the steps and inputs from Rogers (2024). The data from MWP parts 1 and 2 from trial class 1 were combined for analysis. Prior to EFA and CFA, assumption tests were conducted using the Kaiser-Mayer-Olkin (KMO) and Bartlett tests. The measure of sampling adequacy (MSA) was determined in the assumption test before EFA, with an overall KMO test result of 0.66. The MSA was deemed acceptable as it exceeded 0.5. Similarly, in the assumption test before the CFA, the overall KMO test yielded a value of 0.67, indicating an acceptable MSA above 0.5. The Bartlett test results further supported these results, showing a p-value of <0.001.

After confirming sample adequacy, EFA and CFA were conducted using the maximum likelihood (ML) method. The EFA revealed from the scree plot that only one factor was present in both parts of the MWP test. The scree plot is depicted in Figure 1.

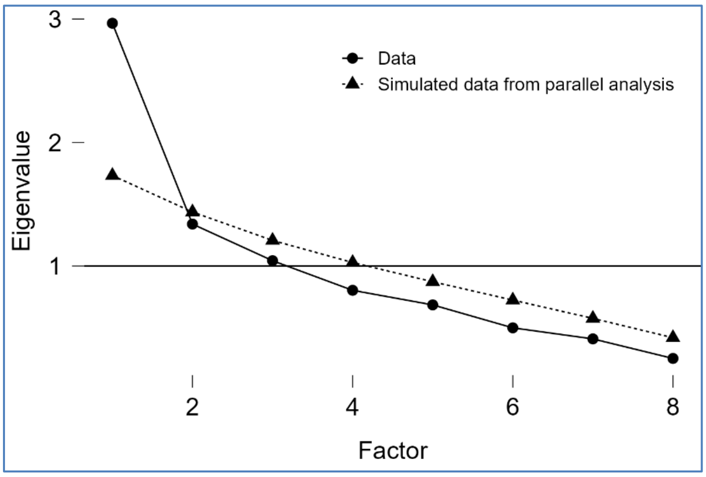


Figure 1. Scree plot

Figure 1 shows that the graph starts to slope in the second factor, indicating the inclusion of only one factor in the MWP test. The results of the parallel analysis (based on principal components) indicate that factor 1 with an eigenvalue of 2.965 and accounting for 28.8% of the variance in the cumulative rotated solution. Furthermore, factor 1 exceeds the simulated data, indicating that its eigenvalue is higher than the mean eigenvalue of the simulated data. Therefore, only factor 1 should be retained.

The CFA results (using the mimic lavaan package) support the EFA findings by confirming that the chi-square test yielded a value of $\chi^2 = 29.2$ with a degree of freedom (df) of 20 and a p-value of 0.084 > 0.050, suggesting a fit for the proposed model of one factor, MPS. These results are further supported by other model fit indices, such as the goodness of fit index (GFI) of 0.916, indicating a model fit with a GFI value exceeding 0.9 (Rosnawati et al., 2015). Additionally, when MWP parts 1 and 2 are separated, the results show a fit for the proposed model of one factor from chi-square test and some of model fit indices. The benchmark values for some of the model fit indices are according to Anggoro et al. (2022).

For MWP part 1, $\chi^2 = 0.746$ and a p-value of 0.689 (> 0.050), GFI of 0.996 (> 0.9), RMSEA (root mean square error of approximation) of 0.000 (< 0.080), CFI (comparative fit index) of 1.000 (> 0.900), and SRMR (standard root mean square residual) of 0.03 (< 0.05). These values indicate a good fit for the model. For MWP part 2, $\chi^2 = 5.014$ and a p-value of 0.082 (> 0.050), GFI of 0.972 (> 0.900), CFI of 0.917 (> 0.900), and IFI (Bollen's incremental fit index) of 0.925 (> 0.900). These results also suggest a good fit for the proposed model of one factor. Therefore, both MWPs demonstrate a good fit for the proposed model of one factor.

RESULTS AND DISCUSSION

Students' MRS and MCS Description

After obtaining a description of the MWP quality (MRWP and MCWP), the next step was to conduct a test for trial class 2. When all students had completed the test, their answers were analyzed and scored based on the scoring guidelines. The descriptive statistics for trial classes 1 and 2 and the combined data are presented in Table 7.

Table 7. Descriptive statistics

Class	Estimation	Part 1			Part 2		
		MRWP	MCWP	MWP	MRWP	MCWP	MWP
Trial 1	n	40	40	40	39	39	39
	Average	11.88	10.31	11.09	18.59	10.58	14.58
	Std. Deviation	7.98	8.44	7.18	13.10	7.87	9.47
	Minimum	0	0	0	0	0	0
	Maximum	25.00	31.25	25.00	43.75	25.00	34.38
Trial 2	n	38	38	38	38	38	38
	Average	7.24	4.29	5.75	20.26	12.03	16.13
	Std. Deviation	3.99	5.66	4.02	19.14	11.85	14.36
	Minimum	0	0	0	0	0	0
	Maximum	18.75	18.80	15.60	68.80	43.80	56.30
Trial 1 + Trial 2 (Combined)	n	78	78	78	77	77	77
	Average	9.62	7.38	8.49	19.41	11.29	15.35
	Std. Deviation	6.73	7.79	6.41	16.27	9.99	12.08
	Minimum	0	0	0	0	0	0
	Maximum	25.00	31.25	25.00	68.80	43.80	56.30

As shown in Table 7, the second trial class achieved the highest average score on the MRWP section of Part 2, yet it also recorded the lowest average score on the MCWP section of Part 1. A consistent pattern emerges in which students' MRWP scores tend to be higher than their MCWP scores, suggesting a potential disparity or directional relationship between MRS and MCS. Moreover, students' performance improved from Part 1 to Part 2 of the MWP test in both MRS and MCS, indicating an overall positive trajectory. This trend is particularly pronounced in Trial Class 2, suggesting that students may have made a concerted effort to improve these skills between the two parts of the assessment.

Despite this progress, students in both Trial Classes 1 and 2 demonstrated persistent difficulties in solving MWPs. According to teacher interview data, some students struggled with performing operations involving large numbers, understanding geometric concepts such as tangent lines to circles, and appropriately modeling the MWP scenarios. Although these students were actively engaged and displayed curiosity during instructional activities—especially when learning about tangent lines—they encountered significant obstacles when required to apply their understanding to contextual word problems.

Correlation and Effect Size Between MRS and MCS

Next, the researchers estimated the correlation between MRS 1 and MCS 1, and/or MRS 2 and MCS 2. However, before this analysis, it is crucial to conduct a normality assumption test to ensure that the data distribution within each class is normal. The Shapiro-Wilk test was conducted to check for normality (Table 8). This check is essential to prevent an increase in the Type I error rate (rejecting the null hypothesis when it should be accepted because it is true) and a reduction in statistical power (Bishara & Hittner, 2012). In practice, the transformation from r to Fisher's Z is sensitive to violations of bivariate normality (two variables have a joint normal distribution), impacting hypothesis testing, Confidence Intervals (CI), and averaging correlations (Zimmerman et al., 2003).

Table 8. Normality check

Class	Pairwise	Shapiro-Wilk	p
Trial 1	MRS 1-MCS 1	0.965	0.255
	MRS 2-MCS 2	0.894	0.002
Trial 2	MRS 1-MCS 1	0.836	< .001
	MRS 2-MCS 2	0.896	0.002
Combined	MRS 1-MCS 1	0.928	< .001
	MRS 2-MCS 2	0.939	0.002

Table 8 exhibits that only trial class 1 with pairwise MRS1-MCS 1 (n = 40) follows a normal distribution as 0.255 > 0.05 (5% significance level). Thus, the null hypothesis is accepted. Therefore, only this class was selected to estimate the correlation and effect size. The correlation was calculated using Pearson's r, Spearman's ρ, Kendall's τ, and effect size estimation. The estimations of Pearson's r correlation, Spearman's ρ, and Kendall's τ to avoid assuming a linear and monotonic association (refer to linearity and monotonous in a regression approach) among the three correlation coefficients (van den Heuvel & Zhan, 2022). The complete summary is detailed in Table 9.

Table 9 indicates that the relationship between MRS and MCS with MPS orientation is significant and positive, as the p-value is < 0.05 (significance level of 5%). Thus, the null hypothesis is rejected. Meanwhile, Pearson's r value shows a significant relationship between the two at the 0.1% level. Spearman's ρ and Kendall's τ also demonstrated a significant relationship between the two at the 1% level. The effect size results indicate a strong relationship between MRS and MCS with MPS orientation, based on Pearson's r and Spearman's ρ, and a medium relationship based on Kendall's τ.

Table 9. Correlation and effect size between MRS and MCS

Variable	Estimation	MRS 1 Trial 1	MCS 1 Trial 2
MRS 1 Trial 1	Pearson's r	—	
	p-value	—	
	Effect size (Fisher's z)	—	
	SE Effect size	—	
	Spearman's ρ	—	
	p-value	—	
	Effect size (Fisher's z)	—	
	SE Effect size	—	
	Kendall's τ	—	
	p-value	—	
	Effect size (Fisher's z)	—	
	SE Effect size	—	
MCS 1 Trial 1	Pearson's r	0.529***	—
	p-value	< .001	—
	Effect size (Fisher's z)	0.589	—
	SE Effect size	0.164	—
	Spearman's ρ	0.493**	—
	p-value	0.001	—
	Effect size (Fisher's z)	0.54	—
	SE Effect size	0.169	—
	Kendall's τ	0.4**	—
	p-value	0.003	—
	Effect size (Fisher's z)	0.423	—
	SE Effect size	0.092	—

Note: * p < .05, ** p < .01, *** p < .001

Discussion

This study presents the results of MWP test using descriptive statistics. Although there was an observable improvement in mean scores from Part 1 to Part 2 of the MWP test, students' average scores remained below 21 out of 100. This finding highlights the generally low levels of MRS and MCS among Indonesian high school students, particularly in the domain of geometry. These results are consistent with the 2022 PISA findings, in which Indonesian students scored an average of 367 in the "shape and space" category, well below the OECD average (OECD, 2023). Similar patterns have been documented in other regional studies, which also report suboptimal MRS performance among students (Sari et al., 2019).

Interviews with mathematics teachers further support the test results. Teachers reported that students experience difficulties when solving MWPs due to challenges in performing complex calculations, understanding geometric concepts—such as tangent lines to circles—and engaging in mathematical modeling. These difficulties align with prior findings that solving MWPs demands proficiency in mathematical computation, conceptual understanding, and the ability to translate real-world scenarios into mathematical representations (Novak & Tassell, 2017). The inherent complexity of MWPs often makes them cognitively demanding and difficult for students to solve (Daroczy et al., 2015; Verschaffel et al., 2020). Furthermore, limited MCS has been identified as a contributing factor that impedes students' ability to effectively approach and solve such problems (Nurjanah & Jusra, 2022; Yusoff et al., 2022).

Students frequently struggle to identify relevant information, interpret the problem statement, and determine an appropriate strategy when solving MWPs. These challenges not only affect their problem-solving performance but also limit their ability to communicate their reasoning and solutions effectively (Utami et al., 2018). One visible manifestation of these difficulties is the frequent occurrence of conceptual and procedural errors (Dwita & Retnawati, 2022).

Research by Hasan et al. (2019) and Lestari et al. (2019) confirms that errors in MWP solving stem from a variety of factors. Aljura et al. (2025) identified five domains that contribute to students' errors in geometry problems: (1) MPS, (2) cognition, (3) affect, (4) motivation, and (5) self-awareness. For example, students may exhibit limited exploration in the MPS domain, struggle with conceptual understanding in the cognitive domain, experience anxiety in the affective domain, demonstrate low motivation or even engage in dishonest behavior, and display poor preparation or time management in the domain of self-awareness.

To address these challenges, mathematics educators can adopt a range of strategies. First, teachers need to strengthen their diagnostic skills to identify specific difficulties within students' MPS processes (Wijaya et al., 2019). Second, educators can apply structured error analysis techniques, including Newman's Error Analysis (White, 2005), its modified versions (e.g., Chiphambo & Mtsi, 2021; Kotze, 2018; Makamure & Jojo, 2022; Pomalato et al., 2020; Rosli et al., 2020; Wijaya et al., 2014), and analytical frameworks such as the Rott-Specht-Knipping MPS model (Aljura et al., 2025). Third, teachers may incorporate self-reflective learning activities to help students identify and correct errors, increase motivation, and deepen engagement (Karaali, 2015; Kwon & Jonassen, 2011). Fourth, evidence-based instructional strategies, such as Anghileri's (2006) scaffolding framework, can be implemented to address students' cognitive and metacognitive needs (Askell-Williams et al., 2012; Larasati & Mampouw, 2018; Milazoni et al., 2022; Susilowati & Ratu, 2018).

This study employed homogeneous purposive sampling, a non-probability sampling technique. This method was chosen to obtain deep insights by selecting participants with shared characteristics until data saturation was reached, ensuring that each participant contributed meaningful and contextually rich information (Etikan, 2016). Although effective for qualitative depth, this sampling strategy may still be prone to bias (Tongco, 2007).

To enhance the validity and interpretability of findings, triangulation techniques were applied (Jick, 1979; Thurmond, 2001). Between-method triangulation involved conducting interviews with the mathematics teacher who instructed students in the trial classes. Within-method triangulation was achieved by administering two parts of the MWP test, both designed to assess similar MPS constructs. According to Jick (1979), within-method triangulation ensures internal consistency, while between-method triangulation strengthens external validity. As Thurmond (2001) explains, triangulation can increase confidence in results, generate novel insights, confirm or refute existing theories, and lead to a more nuanced understanding of the phenomenon under study.

The findings revealed a statistically significant relationship between MRS and MCS within the context of MPS. This confirms previous theoretical assumptions that MRS and MCS are interconnected, and both contribute to effective mathematical problem-solving (Anggoro et al., 2022; Puspa et al., 2019; Sumarsih et al., 2018). The correlation results are in line with those reported by Kustiawati and Siregar (2022), who explored the relationship between MRS and MCS in the context of GeoGebra-assisted MPS instruction. Their quasi-experimental study, employing a post-test only design, reported a significant

Pearson correlation coefficient of 0.491 at the 1% level. In contrast, the current study demonstrates an even stronger relationship, supported by Pearson's r, Spearman's ρ, and Kendall's τ—each significant at the 0.1% level. Moreover, unlike Kustiawati and Siregar's study, the present research includes rigorous validation of the test instruments, detailed descriptions of student performance in MRS and MCS, and the use of multiple correlation coefficients to strengthen reliability.

Other quasi-experimental studies also offer relevant comparisons. For instance, Palinussa et al. (2021) examined the impact of the RME approach on the MRS and MCS of junior high school students in a rural context. While their study included a qualitative exploration of MPS in geometry, it did not provide quantitative evidence of test validity or explicitly demonstrate orientation toward MPS. The current study addresses this gap by offering statistically supported evidence of both test validity and MPS orientation.

Similarly, Primadani et al. (2020) explored MRS and MCS concurrently by combining the Teams Games Tournament (TGT) learning model with the PQ4R (Preview, Question, Read, Reflect, Recite, Review) strategy. While their quasi-experimental approach highlights an innovative instructional design, their study did not explicitly align with MPS orientation. Across these three studies, none reported effect size values in relation to the correlation between MRS and MCS within an MPS framework. The present study fills this gap by incorporating effect size estimates to provide a more comprehensive interpretation of the magnitude of relationships observed. Following Kelley and Preacher's (2012) definition, effect size serves as a quantitative measure of the strength of a phenomenon and supplements statistical significance testing by offering a practical interpretation of the results. As Fritz et al. (2012) suggest, effect sizes add depth to statistical analysis by quantifying how meaningful a result is in real-world terms. This study calculates effect sizes using Pearson's r, Spearman's ρ, and Kendall's τ—all widely recognized metrics for estimating the strength of association (Berben et al., 2012).

In detail, the effect size Z Fisher (Z_r) can be estimated in the form of r Pearson with the formula $r = \frac{e^{2Z_r - 1}}{e^{2Z_r + 1}}$ (Berben et al., 2012). According to Cohen (2013), a correlation value of r around 0.5 falls into the category of a large effect size, equivalent to a Cohen's d value of around 0.8. It means there is a large quantitative measure of the magnitude of correlation or a strong relation between MRS and MCS with MPS orientation. Additionally, Berben et al. (2012) stated that several studies suggest including the reporting of r with a confidence interval (CI). Since JASP 0.18.3.0 version does not provide CI estimation yet, there is a need to calculate it manually.

The CI on r can be determined through the CI estimate on Z_r. By choosing CI 95%, the variance of Z_r or $V_{Z_r} = \frac{1}{n-3} = \frac{1}{40-3} = \frac{1}{37} = 0.027$ is obtained. Then, the standard error or $SE_{Z_r} = \sqrt{V_{Z_r}} = 0.164$. Finally, for the 95% CI on Z_r, $Z_r \pm 1.96(SE_{Z_r}) = 0.589 \pm 1.96\ (0.164) = 0.589 \pm 0.322$.

Thus, the lower and upper limits for the 95% CI on Z_r are obtained, namely $L_{Z_r} = 0.267$ and $U_{Z_r} = 0.911$, respectively. From here, the lower and upper limits for the 95% CI of r can be determined. For the lower limit of r or $L_r = \frac{e^{2L_{Z_r} - 1}}{e^{2L_{Z_r} + 1}} = \frac{e^{2(0.267)} - 1}{e^{2(0.267)} + 1} = \frac{0.706}{2.706} = 0.261$, and for the upper limit of r or $U_r = \frac{e^{2U_{Z_r} - 1}}{e^{2U_{Z_r} + 1}} = \frac{e^{2(0.911)} - 1}{e^{2(0.911)} + 1} = \frac{5.184}{7.184} = 0.722$. Finally, the obtained values for r and the 95% CI are 0.529 and (0.261, 0.722).

The same procedure was employed to calculate the 95% confidence intervals (CIs) for Spearman's ρ and Kendall's τ. The 95% CI for Spearman's ρ was 0.493 with a range of (0.215, 0.697), while the 95% CI for Kendall's τ was 0.400 with a range of (0.101, 0.632). These calculations are consistent with the methodological example provided by Berben et al. (2012), further reinforcing the reliability of the statistical analysis.

Despite these contributions, this study is subject to several limitations. First, the scope of the investigation is confined to a specific topic within high school geometry, rather than extending to broader mathematical domains such as algebra, statistics, or arithmetic. This narrow topical focus may limit the applicability of findings to other areas of mathematical learning. Second, the study's sample is restricted to high school students, and the observed relationships between MRS and MCS—within an MPS-oriented context—may be influenced by developmental and cognitive characteristics specific to this educational level. Third, potential measurement bias may exist due to the uniform difficulty level of all MWP items. Although this may reflect students' genuine struggle with MWPs, it could also skew the results toward underestimating variability in students' problem-solving capacities. Finally, the exclusive use of MWP-based assessments constrains the exploration of MRS and MCS relationships, as alternative forms of assessment may offer richer or complementary insights.

CONCLUSION

This study investigates the relationship between mathematical reasoning skills (MRS) and mathematical communication skills (MCS) in orientation to mathematical problem-solving (MPS) using mathematical word problems (MWP) essay tests. Building on Bjuland's framework for MRS oriented towards MPS, this study extends the framework to include MCS processes. The test instruments used in this study demonstrate good quality, as evidenced by content validity, reliability estimates, discriminatory power, and difficulty levels. The factor analysis confirms that the MWP provided to students is focused on a single factor, MPS. The test instruments effectively assess students' MRS and MCS, and their performance in solving MWPs. Thus, the validity of the results is supported by quantitative evidence. Next, the correlation results between MRS and MCS show significant relationships, with Pearson's r coefficient at 0.529 (95% CI: 0.261, 0.722), Spearman's ρ at 0.493 (95% CI: 0.215, 0.697), and Kendall's τ at 0.4 (95% CI: 0.101, 0.632). Additionally, based on Cohen's r coefficient, the effect size falls within the large category. There are important limitations to consider when interpreting these results.

Despite these limitations, the study has important implications. Firstly, it provides statistical descriptions for meta-analysis studies, a comprehensive discussion of the factors that hinder students in solving MWP, and detailed alternatives for educators, especially mathematics teachers, to enhance students' MRS and MCS with MPS orientation. Secondly, it provides evidence of the interconnectedness of MRS and MCS in orientation MPS that was assumed from previous studies and addresses gaps in previous studies. Thirdly, it suggests the potential influence of these abilities on students' performance in solving mathematical problems, indicating that MRS and MCS could be determining factors in students' success or failure. It can be assumed that, with MPS orientation, improving MRS may also enhance MCS and vice versa. Fourthly, it offers opportunities for further research into the relationships between other mathematical abilities and their orientation towards MPS. Fifthly, future research could investigate the effect of MRS and MCS oriented towards problem-solving in mathematics learning, potentially integrating different approaches, strategies, methods, and learning models to enhance MRS and MCS. Lastly, policymakers, especially the Ministry of Primary and Secondary Education in Indonesia, can guide the mathematics curriculum in schools with an MPS orientation by leveraging insights from the Singaporean national mathematics curriculum and staying updated on the latest research on MPS.

Several recommendations can be made for educators, policymakers, and researchers. Firstly, educators, especially mathematics teachers, should focus on training students in solving MWPs with MPS because MRS and MCS are interconnected in MPS orientation. Thus, mathematics teachers can enhance these two abilities by using an MPS-oriented approach. Secondly, mathematics teachers should diligently review the test instruments used, particularly formative assessments, to gather precise and valuable information for developing students' mathematical abilities. Thirdly, mathematics teachers, especially in Indonesia, need to enhance their knowledges and competencies through training programs established by the government and conducted by educational institutions. Fourthly, policymakers, especially the Ministry of Primary and Secondary Education in Indonesia, should organize and facilitate training sessions for mathematics teachers more effectively and efficiently. Lastly, researchers are encouraged to explore the relationship between other mathematical competencies and MPS. They should expand on the findings of this study, replicate the study in different educational levels or mathematical topics, consider utilizing mixed-methods approaches, and consider using both multiple-choice and combined (multiple-choice and essay) instruments to investigate the relationship and to explore the causal mechanism between MRS and MCS with MPS orientation, as well as other mathematical competencies. Finally, the mentioned limitations suggest that caution should be exercised when generalizing the results. Nevertheless, the study provides a valuable foundation for future research aimed at uncovering the interrelationships among multiple mathematical competencies and their alignment with MPS processes.

Acknowledgments

We would like to thank the high school in East Bintan for their participation in this study. We also want to express our gratitude to all the anonymous reviewers who provided comprehensive suggestions to improve this article. Special thanks to the Rector of Universitas Negeri Yogyakarta (UNY), Prof. Dr. Sumaryanto M. Kes. AIFO, and all UNY leaders for their continuous support and encouragement to students and lecturers.

Declarations

Author Contribution	:	ANA: Conceptualization, Data Curation, Investigation, Writing - Original Draft, Writing - Review & Editing, Editing and Visualization, Formal analysis, Resources, and Methodology. HR: Writing - Review & Editing, Investigation, Funding acquisition, Project Administration, Supervision, and Validation. SRD, GKK, RS, and ARS: Writing - Review & Editing, Supervision, and Validation.
Funding Statement	:	We acknowledge the Government of the Republic of Indonesia for the BPPTNBH funds allocated to the World Class University program, which facilitated APC funding for the publication of this article.
Conflict of Interest	:	The authors declare no conflict of interest.
Additional Information	:	Additional information is available for this paper.

REFERENCES

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Aljura, A. N., Retnawati, H., Zulnaidi, H., & Mbazumutima, V. (2025). Understanding High School Students' Errors in solving Mathematics Problems: A Phenomenological Research. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, *7*(1), 154–178. https://doi.org/10.23917/ijolae.v7i1.24005

Amalina, I. K., & Vidákovich, T. (2022). An integrated STEM-based mathematical problem-solving test: Developing and reporting psychometric evidence. *Journal on Mathematics Education*, *13*(4), 587–604. https://doi.org/10.22342/jme.v13i4.pp587-604

Anggoro, A. F. D., Haji, S., & Sumardi, H. (2022). Structural equation fit test of mathematical connection ability, mathematical reasoning, and mathematics problem-solving ability of junior high school students. *International Journal of Trends in Mathematics Education Research*, *5*(1), 82–93. https://doi.org/10.33122/ijtmer.v5i1.117

Anghileri, J. (2006). Scaffolding practices that enhance mathematics learning. *Journal of Mathematics Teacher Education*, *9*(1), 33–52. https://doi.org/10.1007/s10857-006-9005-9

Ansari, B. I., Taufiq, T., & Saminan, S. (2020). The use of creative problem solving model to develop students' adaptive reasoning ability: Inductive, deductive, and intuitive. *International Journal on Teaching and Learning Mathematics*, *3*(1), 23–36. https://doi.org/10.18860/ijtlm.v3i1.9439

Askell-Williams, H., Lawson, M. J., & Skrzypiec, G. (2012). Scaffolding cognitive and metacognitive strategy instruction in regular class lessons. *Instructional Science*, *40*(2), 413–443. https://doi.org/10.1007/s11251-011-9182-5

Aziza, N., Sridana, N., Hikmah, N., & Subarinah, S. (2023). Analisis kesalahan dan scaffolding dalam menyelesaikan soal cerita matematika materi pecahan [Error analysis and scaffolding in solving mathematical story problems on fractions]. *Jurnal Ilmiah Profesi Pendidikan*, *8*(1), 221–231. https://doi.org/10.29303/jipp.v8i1.1119

Baroody, A. J. (1993). *Problem Solving, Reasoning, And Communicating, K-8 Helping Children Think Mathematically*. Macmillan Publishing Company.

Berben, L., Sereika, S. M., & Engberg, S. (2012). Effect size estimation: Methods and examples. *International Journal of Nursing Studies*, *49*(8), 1039–1047. https://doi.org/10.1016/j.ijnurstu.2012.01.015

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, *17*(3), 399–417. https://doi.org/10.1037/a0028087

Bjuland, R. (2007). Adult Students' Reasoning in Geometry: Teaching Mathematics through Collaborative Problem Solving in Teacher Education. *The Mathematics Enthusiast*, *4*(1), 1–30. https://doi.org/10.54870/1551-3440.1056

Boaler, J. (1993). The Role of Contexts in the Mathematics Classroom: Do They Make Mathematics More “Real”? *For the Learning of Mathematics*, *13*(2), 12–17. http://www.jstor.org/stable/40248079

Böswald, V., & Schukajlow, S. (2023). I value the problem, but I don’t think my students will: preservice teachers’ judgments of task value and self-efficacy for modelling, word, and intramathematical problems. *ZDM - Mathematics Education*, *55*(2), 331–344. https://doi.org/10.1007/s11858-022-01412-z

Bozkuş, F., & Ayvaz, Ü. (2018). Middle School Mathematics Teachers’ Knowledge of Mathematical Reasoning. *European Journal of Education Studies*, *4*(9), 16–34.

BSKAP. (2022). *Salinan Lampiran I Keputusan Kepala Badan Standar, Kurikulum, dan Asesmen Pendidikan Nomor 008/Kr/2022 Tentang Capaian Pembelajaran Pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah Pada Kurikulum Merdeka* [Copy of Attachment I of the Decree of the Head of the Education Standards, Curriculum, and Assessment Agency Number 008/Kr/2022 Concerning Learning Achievements in Early Childhood Education, Elementary Education Level, and Secondary Education Level in the Independent Curriculum] (p. 146). https://guru.kemdikbud.go.id/kurikulum/referensi-penerapan/capaian-pembelajaran/mata-pelajaran/fase/?level=SD-SMA&subject=Matematika&phase=F

Cartwright, K. (2020). Analyzing students’ communication and representation of mathematical fluency during group tasks. *Journal of Mathematical Behavior*, *60*. https://doi.org/10.1016/j.jmathb.2020.100821

Chiphambo, S. M., & Mtsi, N. (2021). Exploring Grade 8 Students’ Errors When Learning About the Surface Area of Prisms. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(8), em1985. https://doi.org/10.29333/ejmste/10994

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (Sixth Edition). Sage publications.

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (Sixth Edition). Pearson.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *06*. https://doi.org/10.3389/fpsyg.2015.00348

Dwita, A., & Retnawati, H. (2022). Students’ errors in solving mathematical problems. *AIP Conference Proceedings*, *2575*. https://doi.org/10.1063/5.0107794

Eggen, P. D., & Kauchak, D. P. (1996). *Strategy for Teacher: Teaching Content and Thinking Skills*. Allyn & Bacon.

Etikan, I. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1. https://doi.org/10.11648/j.ajtas.20160501.11

Fauzie, M., Pada, A. U. T., & Supriatno, S. (2021). Analysis of the difficulty index of item bank according to cognitive aspects during the Covid-19 pandemic. *Jurnal Penelitian dan Evaluasi Pendidikan*, *25*(2), 162–169. https://doi.org/10.21831/pep.v25i2.42603

Fisher, R. A. (1921). On the “Probable Error” of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, *1*, 3–32. http://hdl.handle.net/2440/15169

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Guilford, J. P. (1950). *Fundamental Statistics in Psychology and Education* (C. T. Morgan, Ed.; 2nd ed.). McGraw-Hill Book Company.

Habsah, F. (2017). Developing teaching material based on realistic mathematics and oriented to the mathematical reasoning and mathematical communication. *Jurnal Riset Pendidikan Matematika*, *4*(1), 43–55. https://doi.org/10.21831/jrpm.v4i1.10199

Haladyna, T. M., & Rodriguez, M. C. (2021). Using Full-information Item Analysis to Improve Item Quality. *Educational Assessment*, *26*(3), 198–211. https://doi.org/10.1080/10627197.2021.1946390

Hancock, G. R., & An, J. (2020). A Closed-Form Alternative for Estimating ω Reliability under Unidimensionality. *Measurement*, *18*(1), 1–14. https://doi.org/10.1080/15366367.2019.1656049

Hasan, N., Subanji, S., & Sukorianto, S. (2019). Analisis Kesalahan Siswa Kelas VIII dalam Menyelesaikan Soal Cerita Terkait Teorema Pythagoras [Analysis of Errors of Grade VIII Students in Solving Story Problems Related to the Pythagorean Theorem]. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, *4*(4), 468. https://doi.org/10.17977/jptpp.v4i4.12264

Herbert, S., Vale, C., White, P., & Bragg, L. A. (2022). Engagement with a formative assessment rubric: A case of mathematical reasoning. *International Journal of Educational Research*, *111*, 101899. https://doi.org/10.1016/j.ijer.2021.101899

Hershkowitz, R. (2020). Shape and Space: Geometry Teaching and Learning. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 774–779). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_138

Hoyles, C., Wolf, A., Molyneux-Hodgson, S., & Kent, P. (2002). *Mathematical Skills in the Workplace*. https://discovery.ucl.ac.uk/id/eprint/10001565/1/Hoyles2002MathematicalSkills.pdf

Hwang, I. H. (1970). The Usability of Item-Total Correlation as the Index of Item Discrimination. *Korean Journal of Medical Education*, *12*(1), 45–51. https://doi.org/10.3946/kjme.2000.12.1.45

Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, *24*(4), 602. https://doi.org/10.2307/2392366

Johari, J., Sahari, J., Abd Wahab, D., Abdullah, S., Abdullah, S., Omar, M. Z., & Muhamad, N. (2011). Difficulty index of examinations and their relation to the achievement of programme outcomes. *Procedia - Social and Behavioral Sciences*, *18*, 71–80. https://doi.org/10.1016/j.sbspro.2011.05.011

Johnson, B., & Christensen, L. (2019). *Educational research: Quantitative, qualitative, and mixed approaches* (7th ed.). Sage Publication.

Junsay, M. (2016). Reflective learning and prospective teachers' conceptual understanding, critical thinking, problem solving, and mathematical communication skills. *Research in Pedagogy*, *6*(2), 43–58. https://doi.org/10.17810/2015.34

Kalkbrenner, M. T. (2023). Alpha, Omega, and H Internal Consistency Reliability Estimates: Reviewing These Options and When to Use Them. *Counseling Outcome Research and Evaluation*, *14*(1), 77–88. https://doi.org/10.1080/21501378.2021.1940118

Kameenui, E. J., & Griffin, C. C. (1989). The National Crisis in Verbal Problem Solving in Mathematics: A Proposal for Examining the Role of Basal Mathematics Programs. *The Elementary School Journal*, *89*(5), 575–593. https://doi.org/10.1086/461593

Kamid, Rusdi, M., Fitaloka, O., Basuki, F. R., & Anwar, K. (2020). Mathematical communication skills based on cognitive styles and gender. *International Journal of Evaluation and Research in Education*, *9*(4), 847–856. https://doi.org/10.11591/ijere.v9i4.20497

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills. *REID (Research and Evaluation in Education)*, *10*(1), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Karaali, G. (2015). Metacognition in the Classroom: Motivation and Self-Awareness of Mathematics Learners. *PRIMUS*, *25*(5), 439–452. https://doi.org/10.1080/10511970.2015.1027837

Kaya, D., & Aydın, H. (2016). Elementary Mathematics Teachers' Perceptions and Lived Experiences on Mathematical Communication. *EURASIA Journal of Mathematics, Science and Technology Education*, *12*(6). https://doi.org/10.12973/eurasia.2014.1203a

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152. https://doi.org/10.1037/a0028086

Kingsdorf, S., & Krawec, J. (2014). Error Analysis of Mathematical Word Problem Solving Across Students with and without Learning Disabilities. *Learning Disabilities Research & Practice*, *29*(2), 66–74. https://doi.org/10.1111/ldrp.12029

Kotze, H. (2018). Competencies in Mathematical Modelling Tasks: An Error Analysis. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(8), 91922. https://doi.org/10.29333/ejmste/91922

Kustiawati, D., & Siregar, N. (2022). Problem-solving with Geogebra: How the relationship between reasoning and communication in mathematics. *AIP Conference Proceedings*, 020031. https://doi.org/10.1063/5.0096136

Kwon, K., & Jonassen, D. H. (2011). The Influence of Reflective Self-Explanations on Problem-Solving Performance. *Journal of Educational Computing Research*, *44*(3), 247–263. https://doi.org/10.2190/EC.44.3.a

Larasati, Y., & Mampouw, H. L. (2018). Scaffolding to Solve the Problem of Comparative Matter Worth of Value and Turning Value. *Math Didactic: Jurnal Pendidikan Matematika*, *4*(1), 47–56. http://jurnal.stkipbjm.ac.id/index.php/math

Larsen, S. C., Parker, R. M., & Trenholme, B. (1978). The Effects of Syntactic Complexity upon Arithmetic Performance. *Learning Disability Quarterly*, *1*(4), 80–85. https://doi.org/10.2307/1510980

Leong, Y. H., Dindyal, J., Toh, T. L., Quek, K. S., Tay, E. G., & Lou, S. T. (2011). Teacher preparation for a problem-solving curriculum in Singapore. *ZDM*, *43*(6–7), 819–831. https://doi.org/10.1007/s11858-011-0356-z

Lestari, A. R. A., Minggi, I., & Qadry, I. K. (2019). Analisis kesalahan dalam menyelesaikan soal cerita materi bangun ruang sisi datar berdasarkan prosedur Newman [Analysis of errors in solving story problems on flat-sided geometric shapes based on the Newman procedure]. *SIGMA (Suara Intelektual Gaya Matematika)*, *11*(2), 122–129.

Lithner, J. (2000). Mathematical Reasoning in Task Solving. *Educational Studies in Mathematics*, *41*(2), 165–190. http://www.jstor.org/stable/3483188

Lombasari, B. N., Subarinah, S., Azmi, S., & Kurniati, N. (2022). Analisis Kesulitan dalam Memecahkan Masalah Soal Cerita Matematika dan Bentuk Scaffolding yang Diberikan Pada Peserta Didik Kelas X SMA Al Ma'arif NU Sinah Pengembur Tahun Ajaran 2021/2022. *Jurnal Ilmiah Profesi Pendidikan*, *7*(3c), 2007–2017. https://doi.org/10.29303/jipp.v7i3c.876

Makamure, C., & Jojo, Z. M. (2022). An analysis of errors for pre-service teachers in first order ordinary differential equations. *Eurasia Journal of Mathematics, Science and Technology Education*, *18*(6), em2117. https://doi.org/10.29333/ejmste/12074

Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ωt and the greatest lower bound. *Social Sciences & Humanities Open*, *7*(1), 100368. https://doi.org/10.1016/j.ssaho.2022.100368

Mammarella, I. C., Giofrè, D., & Caviola, S. (2017). Learning geometry: The development of geometrical concepts and the role of cognitive processes. In *Acquisition of Complex Arithmetic Skills and Higher-Order Mathematics Concepts* (pp. 221–246). Elsevier. https://doi.org/10.1016/B978-0-12-805086-6.00010-2

Marianti, S., Rufaida, A., Hasanah, N., & Nuryanti, S. (2023). Comparing item-total correlation and item-theta correlation in test item selection: A simulation and empirical study. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *27*(2), 133–145. https://doi.org/10.21831/pep.v27i2.61477

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.

McNeish, D. (2018). Thanks coefficient alpha, We'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Milazoni, T. R., Maison, M., & Nizlel, N. (2022). Analisis Kesalahan Siswa Dalam Menyelesaikan Soal Cerita Matematika Berdasarkan Teori Pemrosesan Informasi dan Pemberian Scaffolding. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *11*(1), 654–666. https://doi.org/10.24127/ajpm.v11i1.4705

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, *204*, 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

Moyer, J. C., Moyer, M. B., Sowder, L., & Threadgill-Sowder, J. (1984). Story Problem Formats: Verbal versus Telegraphic. *Journal for Research in Mathematics Education*, *15*(1), 64. https://doi.org/10.2307/748989

Moyer, J. C., Sowder, L., Threadgill-Sowder, J., & Moyer, M. B. (1984). Story Problem Formats: Drawn versus Verbal versus Telegraphic. *Journal for Research in Mathematics Education*, *15*(5), 342. https://doi.org/10.2307/748424

NCTM. (2000). *Principles and Standards for School Mathematics*. NCTM.

Newcombe, N. S., & Frick, A. (2010). Early Education for Spatial Intelligence: Why, What, and How. *Mind, Brain, and Education*, *4*(3), 102–111. https://doi.org/10.1111/j.1751-228X.2010.01089.x

Nickerson, R. (2011). *Mathematical Reasoning*. Psychology Press. https://doi.org/10.4324/9780203848029

Ningrum, F. E., Ayu, D., & Wardhani, P. (2019). Analisis Kesalahan Siswa SD Dalam Menyelesaikan Soal Cerita Operasi Hitung Bilangan Cacah dan Pemberian Scaffolding. *INDIKTIKA (Jurnal Inovasi Pendidikan Matematika)*, *2*(1), 44–57.

Novak, E., & Tassell, J. L. (2017). Studying preservice teacher math anxiety and mathematics performance in geometry, word, and non-word problem solving. *Learning and Individual Differences*, *54*, 20–29. https://doi.org/10.1016/j.lindif.2017.01.005

Novikasari, I., & Dede, Y. (2023). Toward proficiency: Developing a multiplication mathematical content knowledge test for pre-service mathematics teachers in Indonesia and Türkiye. *Journal on Mathematics Education*, *15*(1), 115–130. https://doi.org/10.22342/jme.v15i1.pp115-130

Nurjanah, R., & Jusra, H. (2022). An Analysis of Senior High School Students' Mathematical Communication Ability Based on Self-Efficacy and Gender. *Indonesian Journal of Science and Mathematics Education*. https://doi.org/10.24042/ijsme.v5i1.12809

OECD. (2023). *PISA 2022 Results (Volume I)*. OECD. https://doi.org/10.1787/53f23881-en

Osterlind, S. J. (1989). *Constructing Test Items*. Springer Netherlands. https://doi.org/10.1007/978-94-009-1071-3

Öztürk, M., & Sarikaya, İ. (2021). The relationship between the mathematical reasoning skills and video game addiction of Turkish middle schools students: A serial mediator model. *Thinking Skills and Creativity*, *40*, 100843. https://doi.org/10.1016/j.tsc.2021.100843

Palinussa, A. L., Molle, J. S., & Gaspersz, M. (2021). Realistic mathematics education: Mathematical reasoning and communication skills in rural contexts. *International Journal of Evaluation and Research in Education (IJERE)*, *10*(2), 522. https://doi.org/10.11591/ijere.v10i2.20640

Pamungkas, M. D., & Nugroho, H. (2020). Implementation of Space Geometry Learning Using GeoGebra to Improve Problem Solving Skills. *MaPan*, *8*(2), 224–235. https://doi.org/10.24252/mapan.2020v8n2a4

Penfield, R. D. (2013). Item analysis. In *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology.* (pp. 121–138). American Psychological Association. https://doi.org/10.1037/14047-007

Pólya, G. (2004). *How to solve it: A new aspect of mathematical method* (Vol. 5). Princeton university press.

Pomalato, S. W. D., Ili, L., Ningsi, B. A., Fadhilaturrahmi, F., Hasibuan, A. T., & Primayana, K. H. (2020). Student Error Analysis in Solving Mathematical Problems. *Universal Journal of Educational Research*, *8*(11), 5183–5187. https://doi.org/10.13189/ujer.2020.081118

Primadani, A. I., Mardiyana, & Triyanto. (2020). Mathematical Reasoning and Communication in TGT Learning Model with PQ4R Strategy. *Journal of Physics: Conference Series*, *1613*(1), 012022. https://doi.org/10.1088/1742-6596/1613/1/012022

Puspa, S., Riyadi, R., & Subanti, S. (2019). Profile of mathematical communication skills junior high school students in problem solving. *Journal of Physics: Conference Series*, *1157*, 032125. https://doi.org/10.1088/1742-6596/1157/3/032125

Qohar, A. (2011). Mathematical Communication: What And How to Develop It in Mathematics Learning? *International Seminar and the Fourth National Conference on Mathematics Education*, 1–12.

Quintero, A. H. (1983). Conceptual Understanding in Solving Two-Step Word Problems with a Ratio. *Journal for Research in Mathematics Education*, *14*(2), 102. https://doi.org/10.2307/748578

Retnawati, H. (2016). *Analisis Kuantitatif Instrumen Penelitian* (First). Parama Publishing. www.nuhamedika.gu.ma

Reynolds, C. R., Livingston, R. B., & Willson, V. (2008). *Measurement and Assessment in Education*. Pearson.

Rizbudiani, A. D., Jaedun, A., Rahim, A., & Nurrahman, A. (2021). Rasch model item response theory (IRT) to analyze the quality of mathematics final semester exam test on system of linear equations in two variables (SLETV). *Al-Jabar: Jurnal Pendidikan Matematika*, *12*(2), 399–412. https://doi.org/10.24042/ajpm.v12i2.9939

Rogers, P. (2024). Best practices for your confirmatory factor analysis: A JASP and lavaan tutorial. *Behavior Research Methods*, *56*(7), 6634–6654. https://doi.org/10.3758/s13428-024-02375-7

Rohid, N., Suryaman, S., & 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

Rosli, S., Shahrill, M., & Yusof, J. (2020). Applying the hybrid strategy in solving mathematical word problems at the elementary school level. *Journal of Technology and Science Education*, *10*(2), 216. https://doi.org/10.3926/jotse.965

Rosnawati, R., Kartowagiran, B., & Jailani, J. (2015). A formative assessment model of critical thinking in mathematics learning in junior high school. *REID (Research and Evaluation in Education)*, *1*(2), 186–198. https://doi.org/10.21831/reid.v1i2.6472

Rott, B., Specht, B., & Knipping, C. (2021). A descriptive phase model of problem-solving processes. *ZDM – Mathematics Education*, *53*(4), 737–752. https://doi.org/10.1007/s11858-021-01244-3

Sandberg, J. A. C., & De Ruiter, H. (1985). The solving of simple arithmetic story problems. *Instructional Science*, *14*(1), 75–86. https://doi.org/10.1007/BF00052438

Sari, Y. M., Kartowagiran, B., Retnawati, H., & Fiangga, S. (2019). The characteristics of mathematical reasoning and proof test on indonesian high school students. *Journal of Physics: Conference Series*, *1200*(1). https://doi.org/10.1088/1742-6596/1200/1/012007

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Academic Press.

Silmi Juman, Z. A. M., Mathavan, M., Ambegedara, A. S., & Udagedara, I. G. K. (2022). Difficulties in Learning Geometry Component in Mathematics and Active-Based Learning Methods to Overcome the Difficulties. *Shanlax International Journal of Education*, *10*(2), 41–58. https://doi.org/10.34293/education.v10i2.4299

Solso, R. L., Maclin, O. H., & Maclin, M. K. (2013). *Cognitive Psychology: Pearson New International Edition* (Eight Editon). Pearson.

Sumarsih, Budiyono, & Indriati, D. (2018). Profile of mathematical reasoning ability of 8th grade students seen from communicational ability, basic skills, connection, and logical thinking. *Journal of Physics: Conference Series*, *1008*, 012078. https://doi.org/10.1088/1742-6596/1008/1/012078

Sumpter, L., & Hedefalk, M. (2015). Preschool children's collective mathematical reasoning during free outdoor play. *The Journal of Mathematical Behavior*, *39*, 1–10. https://doi.org/10.1016/j.jmathb.2015.03.006

Susilowati, P. L., & Ratu, N. (2018). Analisis Kesalahan Siswa Berdasarkan Tahapan Newman dan Scaffolding Pada Materi Aritmatika Sosial. *Mosharafa: Jurnal Pendidikan Matematika*, *7*(1), 13–24. https://doi.org/10.31980/mosharafa.v7i1.337

Syukriani, A., Juniati, D., & Siswono, T. Y. E. (2017). *Investigating adaptive reasoning and strategic competence: Difference male and female*. 020033. https://doi.org/10.1063/1.4994436

Thurmond, V. A. (2001). The Point of Triangulation. *Journal of Nursing Scholarship*, *33*(3), 253–258. https://doi.org/10.1111/j.1547-5069.2001.00253.x

Tong, D. H., Uyen, B. P., & Quoc, N. V. A. (2021). The improvement of 10th students' mathematical communication skills through learning ellipse topics. *Heliyon*, *7*(11). https://doi.org/10.1016/j.heliyon.2021.e08282

Tongco, M. D. C. (2007). Purposive Sampling as a Tool for Informant Selection. *Ethnobotany Research & Applications*, *5*, 147–158. http://hdl.handle.net/10125/227

Utami, R. W., Endaryono, B. T., & Djuhartono, T. (2018). Kemampuan Peserta Didik Dalam Menyelesaikan Soal Cerita Matematika. *Faktor: Jurnal Ilmiah Kependidikan*, *5*(3), 187–192. https://journal.lppmunindra.ac.id/index.php/Faktor/article/view/2719

van den Heuvel, E., & Zhan, Z. (2022). Myths About Linear and Monotonic Associations: Pearson's r, Spearman's ρ, and Kendall's τ. *The American Statistician*, *76*(1), 44–52. https://doi.org/10.1080/00031305.2021.2004922

Verschaffel, L., Greer, B., & Corte, E. (2000). *Making sense of word problems*. Swets and Zeitlinger.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4

Viholainen, A. (2011). The view of mathematics and argumentation. *Proceedings of the 7th Congress of the European Society for Research in Mathematics Education*, 243–252.

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicologia*, *33*(3), 755–782. https://doi.org/10.6018/analesps.33.3.268401

White, A. L. (2005). Active Mathematics In Classrooms: Finding Out Why Children Make Mistakes-And Then Doing Something To Help Them. *Square One*, *15*(4).

Wijaya, A., Retnawati, H., Setyaningrum, W., Aoyama, K., & Sugiman. (2019). Diagnosing students' learning difficulties in the eyes of Indonesian mathematics teachers. *Journal on Mathematics Education*, *10*(3), 357–364. https://doi.org/10.22342/jme.10.3.7798.357-364

Wijaya, A., van den Heuvel-Panhuizen, M., Doorman, M., & Robitzsch, A. (2014). Difficulties in solving context-based PISA mathematics tasks: An analysis of students' errors. *The Mathematics Enthusiast*, *11*(3), 555–584. https://doi.org/10.54870/1551-3440.1317

Wijayanti, P. S. (2020). ITEM QUALITY ANALYSIS FOR MEASURING MATHEMATICAL PROBLEM-SOLVING SKILLS. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, *9*(4), 1223. https://doi.org/10.24127/ajpm.v9i4.3036

Yusoff, A. S. M., Durairaj, K., Mohamed, R., Peng, F. S., Abdullah, N., Ismail, H., Salleh, S. M., & Wahid, Z. (2022). Communication in Mathematics Among School Children: A Systematic Review. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *28*(2), 275–290. https://doi.org/10.37934/araset.28.2.275290

Zimmerman, D. W., Zumbo, B. D., & Williams, R. H. (2003). Bias in Estimation and Hypothesis Testing of Correlation. *Psicológica*, *24*(1), 133–158. https://www.redalyc.org/articulo.oa?id=16924109